



HAL
open science

Acoustic Cues to Speech Segmentation in Spoken French: Native and Nonnative Strategies

Ellenor Shoemaker

► **To cite this version:**

Ellenor Shoemaker. Acoustic Cues to Speech Segmentation in Spoken French: Native and Nonnative Strategies. Linguistics. University of Texas at Austin, 2009. English. NNT: . tel-00676964

HAL Id: tel-00676964

<https://theses.hal.science/tel-00676964>

Submitted on 6 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

by

Ellenor Marguerite Shoemaker

2009

**The Dissertation Committee for Ellenor Marguerite Shoemaker
certifies that this is the approved version of the following dissertation:**

**Acoustic Cues to Speech Segmentation in Spoken French:
Native and Non-native Strategies**

Committee:

David Birdsong, Supervisor

Carl Blyth

Randy Diehl

Jean-Pierre Montreuil

Marta Ortega-Llebaria

**Acoustic Cues to Speech Segmentation in Spoken French:
Native and Non-native Strategies**

by

Ellenor M. Shoemaker, B.A., M.A.

Dissertation

**Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of**

Doctor of Philosophy

**The University of Texas at Austin
August 2009**

**For Marguerite and Roberta,
with gratitude for blazing the trail**

Acknowledgements

I would first like to express my gratitude to committee members Carl Blyth, Jean-Pierre Montreuil and Marta Ortega-Llebaria, both for their invaluable commentary and suggestions at various (and numerous) stages of this project, and also for the time I spent in each of their classrooms throughout my time at the University of Texas. I would especially like to thank Randy Diehl for his ongoing support. Randy has been a part of this project since the beginning stages of my Masters research. His courses in speech perception in the Department of Psychology sparked my interest in speech segmentation and were instrumental in laying the foundation for my research. Each of these professors has contributed in a unique and indispensable way to my development as a researcher.

I owe the deepest of gratitude to my dissertation director David Birdsong, not only for introducing me to the body of research upon which this study is based, but for eight years of unwavering support and confidence. His seemingly unbounded knowledge in the fields of second language acquisition and psycholinguistics has been an inspiration, as has his limitless patience and accessibility. I feel extremely privileged to have had the chance to work so closely with such a respected scholar.

This research would also not have been possible without the support of the University of Texas. This support included many semesters of teaching assistantships in the Department of French and Italian, for which I am very grateful. The department also generously funded travel to several conferences, as well as payments to experiment participants. I would particularly like to thank to Kevin Hendryx, the department Graduate Coordinator. Without Kevin, the filling out of endless forms and the organizing of a dissertation defense from 5,000 miles away would have been unbearable. I am grateful not only for his extensive help with all matters administrative, but in particular for the *Jeanne* mementos that have kept watch over me throughout the dissertation process.

Gratitude is further owed to the Elsa Spinelli at the Université Pierre Mendes France in Grenoble, France, whose cooperation and generosity during the initial stages of my Masters research made my doctoral research possible. I thank as well Emilie Destruel for the testing of beginning participants. I would also like to thank all the participants who took part in this study for generously giving their time.

Acoustic Cues to Speech Segmentation in Spoken French:

Native and Non-native Strategies

Ellenor Marguerite Shoemaker, Ph.D.

The University of Texas at Austin, 2009

Supervisor: David Birdsong

In spoken French, the phonological processes of liaison and resyllabification can render word and syllable boundaries ambiguous. In the case of liaison, for example, the final /n/ of the masculine indefinite article *un* [œ̃] is latent in isolation or before word beginning with a consonant (*un stylo* [œ̃.sti.lo] ‘a pen’); however, when followed by a vowel-initial word the /n/ surfaces and is resyllabified as the onset of that word (*un ami* [œ̃.na.mi] ‘a pen’). Thus, the phrases *un air* ‘a melody’ and *un nerf* ‘a nerve’ are produced with identical phonemic content and syllable boundaries [œ̃.nɛʁ]. Some research has suggested that speakers of French give listeners acoustic cues to word boundaries by varying the duration of consonants that surface in liaison environments relative to consonant produced word-initially. Production studies (e.g. Wauquier-Gravelines 1996; Spinelli et al. 2003) have demonstrated that liaison consonants (e.g. /n/ in *un air*) are significantly shorter than the same consonant in initial position (e.g. /n/ in *un nerf*). Studies on the perception of spoken French have suggested that listeners exploit these durational differences in the segmentation of running speech (e.g. Gaskell et al. 2002; Spinelli et al. 2003), though no study to date has tested this hypothesis directly.

The current study employs a direct test of the exploitation of duration as a segmentation cue by manipulating this single acoustic factor while holding all other factors in the signal constant. Thirty-six native speakers of French and 54 adult learners of French as a second language (L2) were tested on both an AX discrimination task and a forced-choice identification task which employed stimuli in which the durations of pivotal consonants (e.g. /n/ in [œ̃.nɛʁ]) were instrumentally shortened and lengthened. The results suggest that duration alone can indeed modulate the lexical interpretation of ambiguous sequences in spoken French. Shortened stimuli elicited a significantly larger proportion of vowel-initial (liaison) responses, while lengthened stimuli elicited a

significantly larger proportion of consonant-initial responses, indicating that both native and (advanced) non-native speakers are indeed sensitive to this acoustic cue.

These results add to a growing body of work demonstrating that listeners use extremely fined-grained acoustic detail to modulate lexical access (e.g. Salverda et al. 2003; Shatzman & McQueen 2006). In addition, the current results have manifest ramifications for study of the upper limits of L2 acquisition and the plasticity of the adult perceptual system in that several advanced learners of French showed evidence nativelylike perceptual sensitivity to *non-contrastive* phonological variation.

TABLE OF CONTENTS

INTRODUCTION	13
0.0 Introduction.....	13
0.1 Motivation for Research.....	14
0.2 Overview of Dissertation	16
CHAPTER ONE: Spoken Word Recognition.....	19
1.0 Introduction.....	19
1.1 Stages of Spoken Word Recognition.....	21
1.2 The Role of Context: Autonomy versus Interactivity.....	23
1.3 Models of Spoken Word Recognition	25
1.3.1 Cohort	25
1.3.2 TRACE	28
1.3.3 Shortlist.....	30
1.3.4 Neighborhood Activation Model	30
1.4 The Segmentation Problem.....	32
1.4.1 Lexically-based Speech Segmentation	32
1.4.2 Acoustic and Phonological Cues to Word Boundaries.....	33
1.4.3 The Use of Multiple Cues in Speech Segmentation	41
CHAPTER TWO: Spoken Word Recognition in French	44
2.0 Introduction.....	44
2.1 Syllable-Based Segmentation in French	46
2.2 The Perception of Liaison	47
2.3 Theoretical Accounts of Liaison	51
2.4 Acoustic-Phonetic Cues to Segmentation in Spoken French	53
2.4.1 Cues to Resyllabification and Enchaînement.....	55
2.4.2 Cues to Liaison	56
2.4.3 The Exploitation of Segmental Duration in the Perception of Liaison.....	58
2.5 Summary	59
CHAPTER THREE: Processing Speech in a Second Language.....	60
3.0 Introduction.....	60
3.1 Age Effects in Phonological Processing	60
3.1.1 Becoming ‘Adult Listeners’	62
3.2 Modeling the Acquisition and Perception of Second Language Phonology	63
3.2.1 The Perceptual Assimilation Model (Best 1995).....	64
3.2.2 The Speech Learning Model (Flege 1995)	66
3.2.3 Re-modeling Phonological Acquisition	67
3.3 Spoken Word Recognition in a Second Language.....	70
3.3.1 Lexical Competition.....	70
3.3.2 Prosodic Structure	72
3.3.3 Probabilistic Phonotactics.....	74
3.3.4 Non-Contrastive Allophonic and Phonological Variation	75
3.4 Summary	77

CHAPTER FOUR: Processing French as a Second Language	79
4.0 Introduction	79
4.1 The Acquisition of L2 French Phonology	79
4.2 Production of L2 French	80
4.3 The Perception of Liaison in L2 French	81
PART TWO: The Current Study	86
CHAPTER FIVE: Materials Production	88
5.0 Introduction	88
5.1 Creation of Production Sample	88
5.1.1 Production Participants	88
5.1.2 Materials	89
5.1.3 Recording procedure	94
5.2 Acoustic Analyses	95
5.2.1. Acoustic Analyses by Individual Consonant	100
5.2.2 Discussion	105
5.3 Post-Hoc Acoustic Analyses	107
5.3.1 Vowel Duration	107
5.3.2 Voice Onset Time	119
5.3.3 Discussion	124
5.4 Instrumental Manipulation of Stimuli	125
CHAPTER SIX: AX Discrimination task	129
6.0 Introduction	129
6.1 Predictions	129
6.2 Experiment 1: AX Discrimination Task (real-word stimuli)	133
6.2.1 Participants	133
6.2.2 Stimuli	134
6.2.3 Procedure	134
6.2.4 Results: Experiment 1	135
6.2.5 Discussion: Experiment 1	149
6.3 Experiment 1.1: AX Discrimination Task employing Non-Word Stimuli	152
6.3.1 Predictions	152
6.3.2 Participants	153
6.3.3 Stimuli	153
6.3.4 Procedure	153
6.3.5 Results: Experiment 1.1	153
6.3.6 Discussion: Experiment 1.1	165
6.4 Discussion: Experiments 1 and 1.1 (real versus non-word stimuli)	166
6.4.1 Reaction Time	169
6.4.2 NS versus NNS performance	173
6.5 Experiment 1.2: AX Discrimination (beginning learners)	174
6.5.1 Participants	174
6.5.2 Stimuli	174
6.5.3 Procedure	174
6.5.4 Results: Experiment 1.2	174

6.5.5 Discussion: Experiment 1.2	184
6.6 General Discussion: AX Discrimination Task	185
CHAPTER SEVEN: Forced-Choice Identification Task	188
7.0 Introduction	188
7.1 Predictions	188
7.2 Experiment 2: Forced-choice Identification Task (real-word stimuli)	189
7.2.1 Participants	189
7.2.2 Stimuli	190
7.2.3 Procedure	190
7.2.4 Experiment 2: Results	190
7.2.5 Discussion: Experiment 2	201
7.3 Experiment 2.1: Forced-choice Identification Task (non-word stimuli)	203
7.3.1 Predictions	203
7.3.2 Participants	203
7.3.3 Stimuli	203
7.3.4 Procedure	203
7.3.5 Experiment 2.1: Results	204
7.3.6 Discussion: Experiment 2.1	214
7.4 Discussion: Experiments 2 and 2.1 (real versus non-word stimuli)	215
7.5 Post-hoc Analyses	218
7.5.1 Response Bias	218
7.5.2 Additional Acoustic Factors	230
7.5.3 Discussion: Post-hoc Analyses	236
7.6 General Discussion: Forced-Choice Identification Task	236
CHAPTER EIGHT: NNS Biographical Data	238
8.0 Introduction	238
8.1 L2 Experiential Factors in L2 Attainment	238
8.2 NNS Biographical and Experiential Data	239
8.2.1 Experiments 1 and 1.1: AX Discrimination Task	241
8.2.2 Experiments 2 and 2.1: Forced-choice Identification Task	242
8.2.3 Discussion	243
8.3 Nativelike Performance	244
8.3.1 Experiments 1 and 1.1: AX Discrimination Task	244
8.3.2 Experiments 2 and 2.1: Forced-choice Identification Task	246
8.3.3 Discussion	248
CHAPTER NINE: Conclusion	250
9.0 General Discussion	250
9.1 Result Summary	250
9.2 The Exploitation of Fine-phonetic Detail within Current SWR Models	251
9.2.1 Exemplar Models of Speech Processing	253
9.2.2 Probabilistic/Distributional Model of SWR	254
9.2.3 Parallel Segmental and Suprasegmental Analysis	255
9.3 The Exploitation of Fine-phonetic Detail in a L2	256
9.4 Limitations	257

9.4.1 Performance Variation among Participants	257
9.4.2 Regional Accent	259
9.5 Further Research Directions	259
9.5.1 Eye-tracking Employing Manipulated Stimuli	259
9.5.2 Perceptual Training in L2 Phonology	260
9.5.3 L2 Production	261
9.6 Conclusion	262
Appendix I: Consent Form (Production of Materials)	264
Appendix II: Non-Native Participant Biographical Questionnaire	266
Appendix III: Consent Form (Behavioral Tasks: French)	269
Appendix IV: Consent Form (Behavioral Tasks: English)	271
Appendix V: Native Speaker Participant Biographical Questionnaire	273
Appendix VI: Experimental Instructions: Experiment 1	274
Appendix VII: Experimental Instructions: Experiment 1.1	276
Appendix VIII: Experimental Instructions: Experiment 1.2	278
Appendix IX: Experimental Instructions: Experiment 2	279
Appendix X: Sample Presentation Screen: Experiment 2	281
Appendix XI: Experimental Instructions: Experiment 2.1	282
Appendix XII: Sample Presentation Screen: Experiment 2.1	284
References	285

De quelle couleur est un tiroir quand il n'est pas fermé ?

[i.lɛ.tu.vɛʁ]

INTRODUCTION

Apprendre une langue nouvelle, c'est apprendre à parler dans une langue différente de la sienne, mais c'est également apprendre à écouter et à comprendre dans cette nouvelle langue.
(Lhote 1995:26)

[To learn a new language is to learn to speak in a language different from one's own, but it is also to learn to listen and understand in this new language.]

0.0 Introduction

The domain of spoken word recognition (SWR) is expansive, ranging from the processing of the raw acoustic signal to the semantic analysis of sentences. This dissertation investigates one step in the comprehension of running speech — the localization of word boundaries and the segmentation of the acoustic signal into discrete and meaningful processing units.

The objectives of the present study are twofold. First is the investigation of segmental duration as a cue to segmentation in spoken French by native speakers, which will add to a growing body of research on the exploitation of fine-grained acoustic detail in lexical access and word recognition. Currently accepted psycholinguistic models of SWR assume that the phoneme is the smallest pre-lexical unit that directly accesses the mental lexicon (e.g. TRACE, McClelland & Elman 1986; Shortlist, Norris 1994), however recent work has demonstrated that sub-phonemic variation can modulate lexical access (see for example McMurray et al. 2002; Salverda et al. 2003; Shatzman & McQueen 2006) revealing a significant limitation of such a phoneme-based recognition system. The present study will bring new data to bear on this controversy.

Second, this dissertation will examine the perceptual capacities of highly advanced adult learners in the use of fine-grained acoustic detail in the aural comprehension of French as a second language (L2). The bulk of the literature on the

acquisition of L2 phonology to date has largely focused on the acquisition of L2 phonemic contrasts; the use of non-contrastive phonetic detail by highly advanced L2 learners has not been sufficiently explored (for an exception see Darcy et al. 2007). Work presented in this dissertation thus expands the current body of knowledge on the plasticity of the adult perceptual system and the upper limits of L2 phonological acquisition.

0.1 Motivation for Research

A fundamental aspect of human language is the listener's ability to recognize discrete words in a continuous stream of speech. There is no reliable acoustic equivalent of gaps between words as exist in written texts (Lehiste 1972; Nakatani & Dukes 1977). Speech sounds attach to one another without pause in a continuous acoustic signal. Add to this the fact that, unlike the printed word, which can be focused on for any period of time and re-read as desired, the sounds of spoken language are transient and allow the listener to attend to only a small portion of the acoustic stream at a time. Nevertheless, native speech perception is automatic and effortless.

The ease of speech processing in one's native language (L1) stands in sharp contrast to the conscious effort that can be required in the aural comprehension of a L2. Research on the notion of a *critical period* for language learning has attributed this discrepancy between native and non-native language processing to a post-pubescent pruning of perceptual sensitivity that leads to perceptual deficiencies later in life for those who undertake the study of a L2. Several researchers hold that this decline in sensitivity leads to a *perceptual foreign accent* (Strange 1995) and leaves late learners with possibly insurmountable deficits in the perception of L2 phonology (for a review of research on 'non-native listening' see Cutler 2001, 2002).

Recent models of the acquisition of L2 phonology however, have proposed a different view of this heightened period of sensitivity. Kuhl (2000, 2005), for example, has suggested a model that privileges experience over a finite window of time. In this model learning continues until stability in the system is achieved. Put differently, acquisition slows only once there has been sufficient input to establish a distribution that can reliably and systematically predict the categorization of further acoustic input. This goes against the traditional notion of a critical period where acquisition slows as a function of chronological age, but rather puts forward the idea that learning ends when the influx of significantly novel input ends, suggesting that plasticity is maintained through continual exposure to novel input. This model also rejects the notion of the *suppression* of a new (L2) phonological system and instead offers that the acquisition of a new phonological system may be hindered by the degree of entrenchment of the native system.

Specifically in the domain of SWR, a large body of research has established that acoustic and phonological cues to speech segmentation are not exploited to the same extent and in the same manner cross-linguistically (Cutler & Norris 1988; Cutler et al. 1989; Pallier et al. 1993; Sebastián-Gallés et al. 1992; Tabossi et al. 2000). Paradoxically, the very segmentation strategies that render the comprehension of our native language so efficient can hinder the aural comprehension of a L2. Research dealing with specific cues

to segmentation such as phonotactics (Weber 2001) and prosody (Cutler et al. 1989; Dupoux et al. 1997) has suggested that L2 learners are constrained by L1 segmentation routines.

However, more recent research has suggested that learners can not only suppress native segmentation strategies in the processing of an L2 (Cutler, McQueen & Suomi 1997), but can acquire and implement non-native segmentation routines as well (Golato 2002). These results challenge strong claims of limitations on the plasticity of phonological learning and perceptual processing.

French is a language that poses particular challenges for the learner in the comprehension of running speech. As Grammont (1938) observed,

Quelqu'un qui ne sait pas où commencent et où finissent les mots français ne pourrait jamais le deviner en entendant parler. [...] D'ordinaire les mots se disent par groupes, par séries, sans aucun arrêt, et si étroitement unis l'un à l'autre qu'il n'est pas rare qu'une syllabe soit constituée par la fin d'un mot et le commencement d'un autre. (p. 102)

(Someone who does not know where French words begin and end would never be able to guess by listening to running speech.[...] Ordinarily words are spoken in groups or sets, without stopping, and so tightly connected to one another that it is not rare that a single syllable includes the end of one word and the beginning of another.)

The resyllabification phenomena to which Grammont refers result from two phonological processes in spoken French that reflect a strong penchant for a consonant-vowel (CV) syllable structure: 'enchaînement', or concatenation, and liaison¹. These processes often render syllable and word boundaries ambiguous (e.g. *un air* 'a melody' and *un nerf* 'a nerve', both transcribed and syllabified [œ.nɛʁ]²).

The effects of resyllabification and the misalignment of syllable and word boundaries on the perception of spoken French have generated extensive research (Yersin-Besson & Grosjean 1996; Gaskell et al. 2002; Spinelli et al. 2002, 2003; Nguyen et al. 2007 among others), mainly due to a body of work that has proposed the syllable as the basic perceptual unit for speech processing in French (Mehler et al. 1981; Cutler et al. 1989). Given the prominent role of the syllable in French, the prevalence of resyllabification would presumably incur severe processing costs and impede speech segmentation processes.

However, Spinelli, McQueen, and Cutler (2003) found that in the case of liaison in spoken French, perceptual efficacy and processing are not hindered by resyllabification. Using cross-modal priming and a lexical-decision task (i.e. a task in

¹ These processes will be further elaborated in Chapter Two.

² Throughout this dissertation, a period (.) is used to denote a syllable boundary, while the symbol # is used to denote a word boundary.

which participants decide whether a sequence is a real word or not), these authors tested whether phrases rendered ambiguous by the possibility of liaison hindered lexical access. In this study, reaction time was measured and recognition of vowel-initial words was not delayed by resyllabification subsequent to liaison. For example, the words *oignon* ‘onion’ and *rognon* ‘kidney’ were both recognized with equal speed in the phrases *le dernier oignon* ‘the last onion’ and *le dernier rognon* ‘the last kidney’, respectively, though the phrases are putatively homophonous, [lə.dɛʁ.nje.ɔ̃.jɔ̃]. Spinelli et al. hypothesize that listeners exploit “subtle but reliable” variations in segmental duration to locate word boundaries and that access to mental representations is facilitated by these cues, though they did not test this hypothesis directly (2003: 248). In line with this hypothesis is research on the production of French which has revealed significant differences in duration between liaison consonants (e.g. /n/ in *un air*) and initial consonants (e.g. /n/ in *un nerf*). Several studies have shown that consonants that surface in liaison environments are consistently shorter than the same consonant in initial position (Wauquier-Gravelines 1996; Spinelli et al. 2003; Shoemaker 2006). However, no research to date has directly demonstrated that these durational differences can influence lexical interpretation in cases of global ambiguity in French.

0.2 Overview of Dissertation

The current study addresses this research gap with a direct test of native and non-native speakers’ perceptual abilities in the exploitation of allophonic variation in cases of lexical ambiguity in French by isolating and exaggerating durational differences in pivotal consonants (e.g. the /n/ in [œ.nɛʁ]) through instrumental manipulation. Manipulated pivotal consonants are employed in two behavioral tasks in order to investigate perceptual sensitivity to this single segmentation cue.

Part One of the dissertation presents the theoretical and empirical foundations of SWR that have motivated the present study. Chapter One offers a detailed comparison of SWR models in the literature that seek to elucidate the cognitive mechanisms at work in the comprehension and segmentation of continuous speech. Chapter Two examines how these processes pertain specifically to the comprehension of spoken French, which, as noted above, is a language whose phonological particularities pose very specific challenges to SWR processes. Chapter Three then looks at the comprehension and segmentation of continuous speech in a L2, including a discussion of the influence of L1 segmentation routines on L2 processing. Chapter Three also examines L2 SWR within the framework of current models of the acquisition of L2 phonology. Chapter Four specifically explores the comprehension and segmentation of spoken French by non-native speakers, a domain which has received little attention thus far.

Part Two of the dissertation presents an empirical investigation into the processing of lexical ambiguities in spoken French by both native speakers and adult learners of L2 French. Chapter Five details the production, measurement, and instrumental manipulation of stimuli in the perception portion of the study. The chapter also provides additional acoustic data on durational differences between liaison and initial consonants from the utterances used to make the experimental materials. Chapters

Six and Seven present two perceptual tasks employing these manipulated stimuli. Outlined in Chapter Six is an AX discrimination task which taps the perceptual capacities of native and non-native speakers of French in differentiating stimuli in which the duration of the pivotal consonants has been manipulated. Chapter Seven reports the results of a forced-choice identification task which investigates the extent to which native and non-native listeners exploit durational differences in the lexical interpretation of ambiguous phonemic sequences. Chapter Eight specifically examines behavioral data gathered from non-native participants as a function of biographical and experiential data in an attempt to shed light on determining factors in the ultimate attainment of L2 phonology. Finally, Chapter Nine consists of a discussion of the findings and implications of these experiments, including an analysis of results within currently accepted models of SWR and L2 phonological acquisition. Further directions of study in these domains are also explored.

PART ONE:

THEORETICAL AND EMPIRICAL

FOUNDATIONS OF SPOKEN WORD RECOGNITION

CHAPTER ONE: Spoken Word Recognition

We speak in order to be heard and need to be heard in order to be understood.
(Jakobson & Waugh 1979: 96-7)

1.0 Introduction

From a psycholinguistic point of view, the study of spoken word recognition consists in essence in the investigation of the transformation of a continuous acoustic signal into discrete linguistic elements from which meaning is extracted. It is generally assumed that listeners possess a stock of abstract representations of both individual sounds and words in a mental lexicon. This lexicon is thought to contain the entirety of what we know about a word, including its acoustic realization, meaning, orthography (in literate populations), and rules governing its syntactic roles. As the incoming acoustic signal is processed, these mental representations are matched against what the listener hears. Miller and Jusczyk (1989) among others point out two specific areas that would seem to make the mapping of the acoustic input onto these mental representations almost insurmountably complicated: the lack of invariance in speech sounds and the absence of explicit word boundaries in the speech signal.

To begin, there is no one-to-one mapping of the physical signal onto what we perceive as individual phonemes. This stands in contrast to written language, which is represented by a sequence of individually differentiated letters or characters. Indeed, the sounds that make up speech were originally assumed to be analogous to moveable type in a printing press, in that individual sounds could be mixed and matched in limitless patterns, and early accounts of SWR were in fact derived from models of written word recognition (Morton 1969; Forster 1976). However the notion that the speech signal can be divided into distinct phonemes is greatly removed from the physical reality of speech. The acoustic properties of segments are far from fixed. Phonemes are highly susceptible not only to their physical context in the speech signal, but also to external factors stemming from intra- and inter-speaker variability.

As noted by Liberman (1996) phonemes are not like “beads on a string” as there are no clear dividing lines as to where one phoneme ends and another begins. The production of a segment is greatly affected by the physical properties of the segments that bound it. As the vocal tract moves and changes shape to form sounds, segments are *coarticulated* as the gestures involved in the production of adjacent segments combine and overlap in the most efficient way possible, resulting in one sound blending into the next. Vowels for example are not strictly ‘vocalic’ in that they contain information which gives cues to the identity of preceding and following consonants (e.g. formant transitions). What is perceived as the same vowel sound by listeners in, for example, *cat* /kæt/ and *bag* /bæg/ actually have different physical realizations. Conversely, acoustic cues to the vowel segment /æ/ in these two words are spread across both consonants. As a result, there is a great deal of temporal overlap among the physical manifestations of what we perceive as individual sounds, rendering the identification of boundaries between segments problematic.

The specific phonological rules of a particular language also introduce variation into the speech signal. *Assimilation*, like coarticulation explained above, involves the modification of individual segments as they take on features of surrounding segments; however unlike coarticulation, assimilation tends to be language-specific. For example, English speakers assimilate sounds more based on place of articulation, whereas French speakers assimilate more based on voicing. In English, the final /t/ in *night* [naɪt] becomes physically comparable to a /p/ in *night* [paɪt] *bus*, where the coronal /t/ is assimilated to the following labial /b/. The same segment /t/, however, becomes physically similar to the velar stop /k/ in *night* [naɪk] *game*, where the /t/ is assimilated to the following /g/. This is contrasted with French, however, where the segment /t/, for example, becomes voiced in *botte* [dɔt] *grise* ‘grey boot’ as it assimilates to the following voiced /d/.

In addition to variability stemming from the physical parameters of production, variation also arises between individual speakers due to age, gender, and/or dialect, among other factors. Furthermore, any individual speaker can produce the same string of speech in myriad ways depending on changes in speech rate, register, or emotional state. Speakers hyperarticulate speech when they choose to speak slowly and clearly, precisely articulating each individual sound. However, when speakers speak quickly, privileging speed over substance, they can hypoarticulate, running speech sounds together even more and producing centralized and reduced vowels. Take for example the phrase *I do not know* in English which can be produced ranging from the clearly articulated *I do not know* to *I dunno* [aɪdʌnəʊ] to a mere differentiation in pitch, where not one segment is clearly articulated (Hawkins 2003).

All of these phenomena would seem to pose considerable challenges for the matching of what we hear onto mental representations of language. However, despite all of the variability in the speech signal, speech perception is extraordinarily accurate and efficient, even in degraded conditions. Somewhat paradoxically, the above phenomena

actually render the SWR system more efficient. Coarticulation and assimilation are advantageous to perception in that information pertaining to each segment is spread over time; consequently any given point in the physical signal will carry information about more than one segment. Research has demonstrated that coarticulation effects actually facilitate the perception of segments. Consonants are more readily identified when bounded by vowels (Liberman, Delattre, Cooper & Gerstman 1954) and similarly vowels are identified more easily when bounded by consonants (Strange, Verbrugge, Shanweiler & Edman 1976). The result is that information encoded in speech is comprehended far more rapidly than non-speech sounds. Humans can comprehend over 20 phonemes per second, whereas more than 10 non-speech sounds per second are perceived as a continuous buzz as the human hearing capacity is not able to differentiate the individual sounds (Harley 2001). Indeed some research has shown that listeners can identify words in context about 125 milliseconds after their onset (Marslen-Wilson & Welsh 1978) and before the offset is heard (Marslen-Wilson 1985) attesting to the speed with which the SWR system operates.

We now turn to a discussion of the stages assumed to underlie the SWR system, as well as several specific models of SWR in the literature that seek to explain the cognitive mechanisms that process the physical signal into what we comprehend as language.

1.1 Stages of Spoken Word Recognition

The general framework of SWR assumed in the literature is one of progressive stages of abstraction. We begin outlining these stages with an overview of the time course of the recognition process. Frauenfelder and Tyler (1987), in a comprehensive review of SWR models, divide the speech recognition process into three principal stages: initial contact, lexical selection (also activation), and recognition.

Obviously, the first contact a listener has with speech is the acoustic signal. *Initial contact* refers to the reception of the physical acoustic signal by the listener and entails the extraction of an abstract representation in the speech signal that is used to access representations stored in the mental lexicon. Processing at this stage is generally referred to as *pre-* (or *sub-*) *lexical* in that the processing takes place before actual lexical units are accessed. Consequently, by definition, a pre-lexical representation is made up of units smaller than the word itself. The study of SWR has seen much lively debate among researchers regarding what constitutes the pre-lexical processing unit. However, the nature of the pre-lexical code has yet to be definitively identified. Candidates for a pre-lexical form have included temporally defined spectral templates (see for example Klatt 1980, 1989); distinctive features of phonemes (McClelland & Elman 1986); phonemes themselves (Pisoni & Luce 1987); syllables (Mehler et al. 1981; Segui, Dupoux & Mehler 1990); morae (Cutler & Otake 1994) and prosodic units (Grosjean & Gee 1987). Though this dissertation assumes the existence of a pre-lexical representation, it should be noted that some researchers reject the notion altogether, instead assuming that the physical signal maps directly onto mental representations without any intervening abstraction (see for example Marslen-Wilson & Tyler 1980; Luce & Pisoni 1998).

The next phase is *lexical selection*, also referred to as *activation*. As sensory input accumulates, lexical entries go from being potentially accessible in the mental lexicon to being activated and available for retrieval. Processing at (and above) this level is referred to as lexical (or sometimes *post-lexical*) in that it occurs after access to the lexicon has been achieved and entire words have been activated. The goal of processing at this level is the selection from among these lexical hypotheses of the actual words intended by the speaker. It is generally agreed that the lexical stage of processing is characterized by competition among these activated mental entries as the acoustic signal unfolds (e.g. McClelland & Elman 1986; Norris 1994; Gaskell & Marslen-Wilson 1997; Luce & Pisoni 1998).

Evidence supporting simultaneous activation and subsequent competition comes from a large and growing body of work (Zwitserslood 1989; Shillcock 1990; McQueen et al. 1994; Wallace, Stewart & Malone 1995; Gow & Gordon 1995; Norris et al. 1995; Tabossi, Burani & Scott 1995; Vitevitch & Luce 1999; Vroomen & de Gelder 1995, 1997; Zwitserslood & Schriefers 1995; McQueen, Otake & Cutler 2001). One empirical method employed to examine competition processes is the tracking of eye movements. This experimental paradigm exploits the fact that while listening to speech, participants make trackable saccadic eye movements to picturable objects. This method allows researchers to track online lexical access processes in close to real time. (See Tanenhaus & Spivey-Knowlton 1996 for an overview of this methodology.)

Several studies have used this methodology to test the competition hypothesis (McQueen, Norris & Cutler 1994; Norris, McQueen & Cutler 1995; Salverda et al. 2003; Shatzman & McQueen 2006, among others) and have demonstrated that multiple candidates are indeed simultaneously considered before a 'winner' is selected. In one such study (Cutler 2002) participants were presented with four images (e.g. a ladder, a shell, a piece of ham and a hamster) and four shapes (e.g. a circle, a diamond, a triangle and a square) on a computer screen. The participants were instructed to, for example, *put the triangle under the hamster* by manipulating the shapes with a computer mouse. By tracking participants' eye movements, researchers showed that participants fixated as long and as often on the piece of ham (the first syllable of *hamster*) as the hamster before executing the task. Participants did not, however, fixate as long on the shell or the ladder. These results suggest that as the acoustic signal unfolded, participants simultaneously entertained the two candidates starting with the sequence /hæm/ before a final selection was made. This effect has been replicated in French (Dahan, Swingley, Tanenhaus & Magnuson 2000) as well as Dutch (Salverda, Dahan & McQueen 2003).

The third and final stage of SWR is *recognition* itself. As the signal continues to unfold over time, these candidates compete with one another for selection until the acoustic input reaches a 'divergence' or 'uniqueness' point. In this final stage, definitive support for one particular lexical candidate has accumulated and other candidates fall out of competition. As noted above, however, words are often recognized before their offset and the exact recognition point of a word often depends on a complex interplay of bottom-up (acoustic) and top-down (context) factors. These factors are discussed further below in the context of particular models of word recognition.

1.2 The Role of Context: Autonomy versus Interactivity

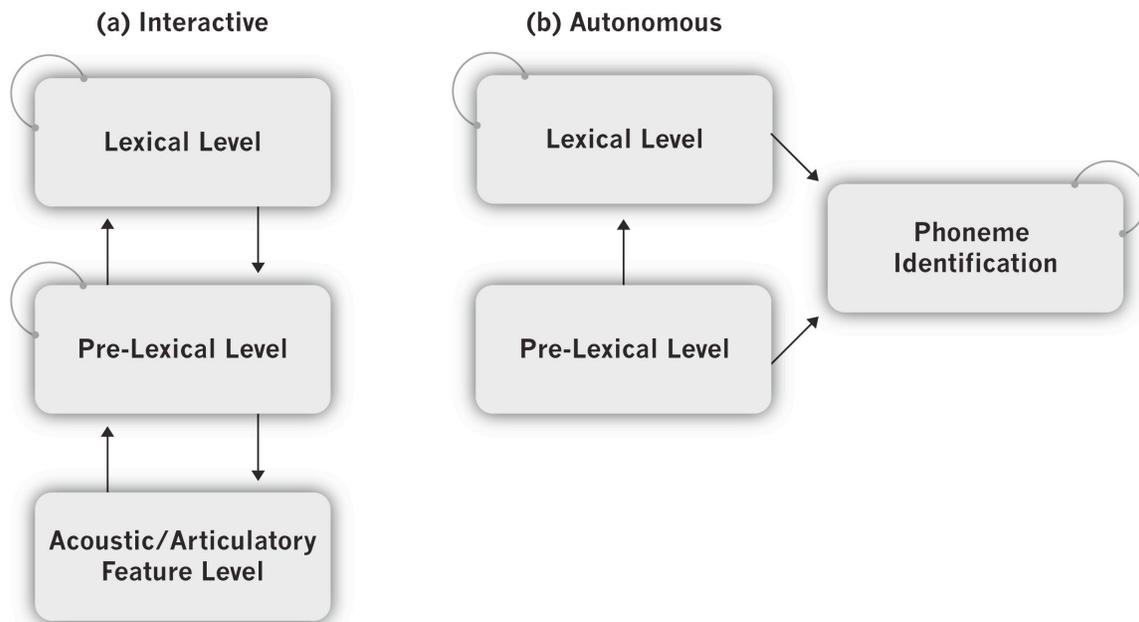
As we will explore below, specific models of SWR diverge considerably on the role of lexical (top-down) information (e.g. context, semantic/syntactic acceptability, lexical frequency, etc.) and the direction of information flow between stages of the recognition process discussed above. In this respect, models are considered to be either *autonomous* or *interactive*. Autonomous models (e.g. Norris 1994) are modular and propose that only information originating from low-level processing (i.e. bottom-up acoustic cues) is employed in the activation of lexical candidates. Top-down information is available solely *after* candidates have been activated in the lexicon. Autonomous models assume that the flow of information is unidirectional, from the bottom toward higher levels only. There is no flow of syntactic or semantic information from higher levels to lower activation levels. However, though there is no exchange of information between levels, information can flow among activated candidates within the same level.

Interactive models (e.g. McClelland & Elman 1986) on the other hand allow for feedback between levels of processing and allow for the use of contextual information (e.g. lexical semantics, pragmatics, plausibility, etc.) in recognition processes. In interactive models, different sources of information from all levels of processing can interact with one another to modulate perception. Put differently, context is thought to modulate the actual sensory analysis of the acoustic signal. As noted above, context also plays a role in autonomous models, but only once lexical candidates have already been activated. Figure 1-1 shows differences in the flow of information in these contrasting models.

Support for interactive models comes from studies suggesting that top-down influence can alter the perception of a phoneme regardless of its physical realization. For example, in a study by Ganong (1980) participants were asked to identify stimuli containing a continuum of six ambiguous phonemes ranging from /k/ to /g/ in which the ambiguous segments had been spliced into sequences such as *-iss /is/* or *-ift /ift/*. Categorical perception predicts the boundary at which listeners would perceive a /k/ or a /g/, however in this study the lexical context in which the ambiguous segment appeared altered participants' perception. (See also Fox 1984 and Pitt 1995.) For proponents of interactivity, these results offer strong evidence that phonemes are categorized not only by their physical realization, but also by lexical information. (See McClelland, Mirman & Holt 2006 for further discussion of evidence supporting interaction in speech processing)

One major criticism of the use of feedback in SWR is the claim that it may increase the risk of misperception. As Norris et al. (2000) note, "the system may perceive events that, although consistent with top-down expectation, are not actually present in the real world." (p. 307) Proponents of autonomous models maintain that phoneme restoration as found by Ganong does not necessarily prove that interactivity is taking place. Fodor (1983) points out that listeners could be guessing the phoneme using lexical knowledge rather than actually perceiving its presence. Similarly, Norris et al. (2000) question whether context actually modulates perception and analysis of the signal or whether the identification of phonemes could be the result of 'feedforward' connections from the lexicon to a higher decision-making level of processing.

Figure 1-1: *The flow of information in (a) an interactive spoken word recognition model and (b) an autonomous spoken word recognition model. Arrows show the direction of information flow. Loops indicate where information flows among lexical units in the same level. (Adapted from McClelland, Mirman & Holt 2006)*



Norris and colleagues (e.g. Norris et al. 2003) have more recently altered their stance on interactivity in light of data from the study of statistical learning by proposing that feedback plays a part in learning processes, but not in online perception. They offer a differentiation between two types of top-down influence: “a lexical bias on phonemic decision-making that does not involve any form of feedback, and lexical feedback for perceptual learning” suggesting in effect that the units involved in forming a speech percept may not be the same units by which the acoustic signal is processed pre-lexically (p. 231). Mirman, McClelland and Holt (2006), however, observe that this proposal would require a separate mechanism for the propagation of feedback that is utilized in learning alone, while it would be more feasible to posit that this mechanism already exists as an inherent component of processing itself.

Researchers are in agreement that context plays an indispensable role in the comprehension of speech, but the question as to when and where context plays its role remains an active debate in psycholinguistics. The evidence to date would seem to favor interactivity in speech comprehension. Support for autonomous models is anchored more in the claim that recourse to interactivity is unnecessary than in empirical evidence in support of autonomy. In addition Tanenhaus et al. (2000) point out that, given “the ubiquitous nature of feedback in the brain,” it is hard to imagine that speech processing would not also benefit from interactivity (p. 348). They further note that much of the

empirical evidence that does exist for autonomous models is based on idealized input that is relatively free of noise and variation, while real speech is highly variable and often produced in less than optimal conditions. As phoneme restoration studies demonstrate, feedback would have obvious advantages in degraded speech conditions.

1.3 Models of Spoken Word Recognition

Despite the dissimilarities noted above, competition models of SWR share many core principles. Divergences among currently accepted models are often subtle. The main differences among models, as we will explore below, arise in the specific time-course of events and the flow of information between processing levels. We now discuss several models of SWR in turn.

1.3.1 Cohort

Among the first and most influential models of SWR was Cohort (Marslen-Wilson & Welsh 1978; Marslen-Wilson & Tyler 1980). In Cohort three successive stages of recognition are identified: access, selection, and integration. In the access stage these authors propose that there is a direct mapping of distinctive features onto units in the mental lexicon with no intervening phonological (pre-lexical) activation. Bottom-up processing of spoken input simultaneously activates a set of viable candidates (the *cohort*) that share identical phonemic content. Unlike subsequent models, which propose that phonemes are activated before entire words, in Cohort “the system does not wait until segmental labels can be assigned before communication to the lexicon. Featural cues start to affect lexical choice as soon as they become available in the speech input” (Lahiri & Marslen-Wilson 1991: 258).

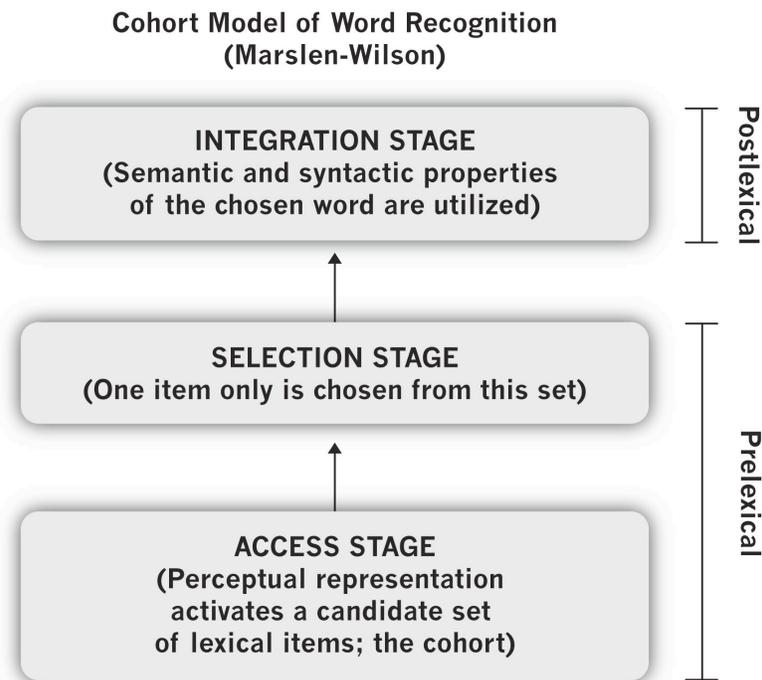
The second stage, or the selection stage, refers to the selection of one particular candidate which best matches the input from among the cohort candidates. Cohort works in a strictly linear fashion; the initial cohort is made up of all lexical entries that share an onset. One notable component of this model is the notion of a *uniqueness* (or *divergence*) *point*, the point in the input at which a candidate is distinguished from its competitors. For example in retrieving the word *spin*, initially every word in the lexicon that begins with /s/ is activated. As the input continues over time, only those words that begin with /sp-/ remain activated, and so on. Unlike later models, Cohort does not posit lateral inhibition between candidates, through which the increased activation of a candidate can reduce the activation levels of its competitors. Deactivation in this model results solely from phonemic mismatch.

In the third and final integration phase of Cohort, semantic and syntactic properties of the chosen word are taken into account. At this stage the word is integrated into its sentential context. The three stages of Cohort are represented graphically below in Figure 1-2.

Empirical evidence for the exploitation of a uniqueness point in speech processing has been demonstrated. Using cross-modal priming in a lexical-decision task in Dutch, Zwitserlood (1989) showed that lexical access can be modulated through manipulation of the time course of the uniqueness point between competitors. Participants in the first portion of this study heard auditory stimuli with aligned onsets (e.g. either *kapitein*

‘captain’ or *kapitaal* ‘capital’) and were presented with a visual probe before the phonemic content of the two words diverged (i.e. before or at /t/). Participants were then asked to decide whether the visual target represented a real word or not. Reaction times suggested that there was facilitation in the recognition of words associated to both items (e.g. facilitation was found for *boot* ‘ship’ associated to *kapitein* and *geld* ‘money’ associated to *kapitaal*) suggesting that both candidates were simultaneously being considered. However, when visual probes were presented after the offsets of words such as *kapitein* or *kapitaal*, priming was found only for an associated word. These results were taken as strong evidence for a divergence point in the recognition process — words with matching input from the onset are simultaneously activated until continuing phonemic input serves to differentiate them.

Figure 1-2: *The three stages of the Cohort model of spoken word recognition: access, selection and integration. Processing levels associated with each stage are shown on the right. (Adapted from Harley 2001.)*



Cohort is considered to be a *constrained activation model* in that it only responds to specific portions of a lexical unit, particularly the onset. In early versions of Cohort, only words which shared onsets were allowed into the cohort. Given the temporal nature of speech, it is logical to assume that a segmental mismatch in word-initial position disrupts lexical access more than in word-medial or word-final position, though this notion is taken to the extreme in Cohort. According to a strict interpretation of the original model, an initial phonemic mismatch would result in an entirely erroneous cohort of candidates. In addition, as pointed out by Harley (2001), one problem with the original

Cohort model is that while it privileges the use of onsets for setting up the original cohort, it fails to propose an explicit mechanism for the identification of where onsets lie. Word boundaries in this model are presumed to emerge after words are recognized, but there is no explanation as to how cohorts can be set up prior to boundary localization.

The privileged exploitation of onsets in recognition has indeed found empirical support. *Shadowing* is a task in which participants listen to running speech which contains distortions and are asked to repeat back what they hear with as little delay as possible. Using shadowing, Marslen-Wilson & Welsh (1978) for example showed that distortions at the beginning of a word disrupted repetition much more than distortions at the end of a word. When words were distorted toward the end of the word, participants repeated them back in their intact form more than 50% of the time, while distortions at the beginning of words were rarely repaired successfully.

Similarly, research in Dutch (Marslen-Wilson & Zwitserlood 1989) showed that when words were mismatched by the initial phoneme, they were not included in the initial cohort. For example, in a cross-modal priming task, lexical access to *honing* ‘honey’ was blocked by *woning* ‘dwelling’, even though all subsequent segments were identical. The word *honing* facilitated access to *bij* ‘bee’, but *woning* did not.

However, further research has suggested that words showing segmental mismatch at the onset are in fact activated in the lexicon, albeit to a lesser extent than words where onsets align. These effects have been shown in English (Alloppenna et al. 1998) as well as in French (Frauenfelder et al. 2001) suggesting that the word recognition system is more tolerant of initial segmental mismatch than originally modeled in Cohort.

The phenomenon of lexical embedding poses further problems for a constrained activation model. Languages have vocabularies of tens of thousands of words made up of a very limited number of phonemes. Consequently, phonemic sequences are often repeated and small words are inevitably embedded within larger ones. The single word *startle* for example contains phonemic support for *star*, *tar*, *art*, *start*, and *tart*. In addition, words can be embedded across word boundaries. Consider the sequence *they may drink rum* [ðeimeɪdrɪŋkrʌm] which contains phonemic content corresponding not only to the four words intended by the speaker but to *aim*, *maid*, *aid*, *ring*, *rink*, and *crumb* as well. (Cutler 2001) This repetition leads to temporary local ambiguities in the speech stream. The original linear version of Cohort does not directly address how these embeddings could be successfully rejected in recognition.

1.3.1.1 Subsequent Revisions to Cohort

Research subsequent to Cohort resulted in substantial revisions to the model (Gaskell & Marslen-Wilson 1997; Marslen-Wilson 1987). In the original instantiation of Cohort, deactivation was achieved solely through a mismatch of acoustic input. However later research showed that a word can be recognized before the uniqueness point is reached by incorporating top-down information such as context and lexical frequency (Grosjean 1980; Tyler & Wessels 1983). Later versions of Cohort took these factors into account and added post-lexical, top-down effects of syntactic and semantic context which could also serve to deactivate competitors.

An additional modification to this model deals with the notion of activation, which was originally presumed to be binary in that a lexical unit was either fully activated by the acoustic input or not. Revised *distributed* versions of Cohort, however, have incorporated the notion of a continuum of activation in which activation levels gradually decay if no further acoustic support is received. This improvement also provided a mechanism capable of identifying and recovering from recognition errors.

The original instantiation of Cohort also proposed that the number of activated candidates in no way affected the recognition process. However, based on research showing that words with many phonemically similar competitors are identified more slowly than words with fewer competitors (Cluff & Luce 1990; Luce & Pisoni 1998), later versions of Cohort took this into account by rendering levels of activation sensitive to context, lexical frequency and neighborhood density.

1.3.2 TRACE

The most influential model of SWR since Cohort is the connectionist simulation model TRACE (McClelland & Elman 1986). Connectionist frameworks are based on models of neuronal organization in the brain and maintain that rules and behavior emerge from the repetitive activation and subsequent strengthening of neural connections. Learning occurs as these connections form networks through the progressive accumulation and strengthening of input-output connections in the neural circuitry.

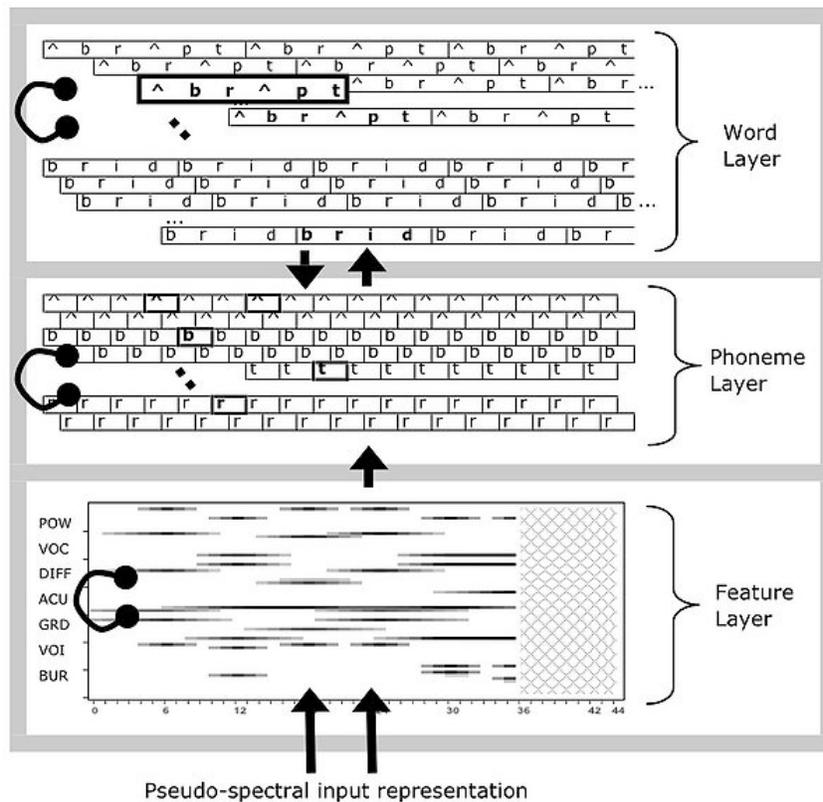
TRACE is a highly interactive competition model that posits three levels of activation: features, phonemes, and words. McClelland and Elman propose that between levels, connections are excitatory, whereas within a particular level, connections are inhibitory. Acoustic input activates distinctive features of phonemes, which then excite phonemes, which then excite all words that contain these phonemes. Within the same level, however, activation of one unit inhibits (i.e. decreases) the activation of its neighbors, which is how TRACE models the competition process.

TRACE features a bidirectional flow of information and feedback between all three levels. Thus TRACE differs from Cohort not only in positing pre-lexical activation of phonemes, but also in the importance given to the role of context in lower-level acoustic processing. Lexical knowledge can aid in the initial activation of candidates in this model, where as in Cohort (and Shortlist as we will see below) context plays a role solely after activation has occurred. Figure 1-3 below demonstrates interactivity between the three levels of TRACE as well as lateral inhibition within levels in the activation of the word *abrupt* /ə'brʌpt/.

Whereas Cohort is a constrained activation model, TRACE is a *continuous activation model*. Recall that Cohort gives preference to activating units in the mental lexicon specifically from word onsets. In contrast, activation in TRACE is characterized by the continual activation of phonemes and words across time. TRACE attempts lexical access at every phoneme, thus the whole network of activated units is in effect reproduced at every access attempt in that each activated phoneme creates a new set of candidates that contain that phoneme. The composition of the competitor group thus continually changes as the input unfolds. Candidates are deleted and added in response to rising and falling activation levels. All units are aligned with each other in discrete time

slices, and units facilitate and inhibit each other to the extent that they overlap within these time slices. In this way TRACE can handle lexical embeddings more effectively than Cohort — embedded words that do not share the same onset can be simultaneously activated. In comparison to Cohort, TRACE better accounts for phenomena such as those revealed by Vroomen and de Gelder (1997). Using a lexical-decision task and cross-modal priming in Dutch, this study showed that for example *framboos* ‘raspberry’ facilitated the recognition of *kwaad* meaning ‘angry’. They attributed this to the fact that *boos*, also meaning ‘angry’, is embedded in *framboos*. A strict sequential model of SWR such as the original version of Cohort would not predict these facilitatory effects.

Figure 1-3: *The three levels of processing in the TRACE model of spoken word recognition showing the activation of the mental representation of the word abrupt. Arrows show communication between levels; semi-circles represent inhibition at a particular level. (Adapted from Strauss, Harris & Magnuson 2007)*



Criticisms of TRACE include the implausibility of its temporal structure. Several researchers have pointed out that the repeated reduplication of the network across time is neither plausible nor efficient. Another possible criticism of TRACE is that it is biased towards longer words. As noted above, a word with the highest degree of activation emerges as the winner. Since longer words contain more segments, they receive more activation. It should be pointed out that, though this bias may seem maladaptive given

that speakers tend to use more shorter words than longer words, it is an efficient strategy in that it attempts to account for the largest portion of the acoustic stream possible at any given point.

1.3.3 Shortlist

The competition model Shortlist (Norris 1988, 1990, 1992, 1993) is also a connectionist model of SWR that attempts to tackle some of the shortcomings of TRACE. Shortlist consists of only two stages: an initial selection of candidates activated from bottom-up information and a competition phase. In the first stage of this model, a group of candidates is activated from both bottom-up activation of phonemes and inhibition of mismatched phonemes. Similar to TRACE, and in an important improvement over Cohort, the initial stage of activation in Shortlist produces multiple candidates which may or may not perfectly align to produce an optimal parse of the acoustic input. Thus, like TRACE, Shortlist also proposes a mechanism for the processing of embeddings. In Shortlist, as in Cohort, the generation of candidate words is separate from the processes of competition. This distinction is thought to be an improvement over TRACE in that the connectionist networks do not need to be continually reproduced as the input progresses over time. In the second stage of Shortlist, candidates compete in an interactive activation network very similar to TRACE. Deactivation is achieved via lateral inhibitory links, though Shortlist also posits bottom-up inhibition of mismatching phonemes.

The most significant dissimilarity from TRACE is that Shortlist is an autonomous model. In TRACE all words in the lexicon are available to be activated at any point in the recognition process, while in Shortlist, only words that are activated in the initial bottom-up activation stage pass on to the second stage of competition. Unlike TRACE, Shortlist does not allow top-down influences on the phoneme level of activation. The flow of information between the level of the phoneme and the level of the word is unidirectional (bottom-up).

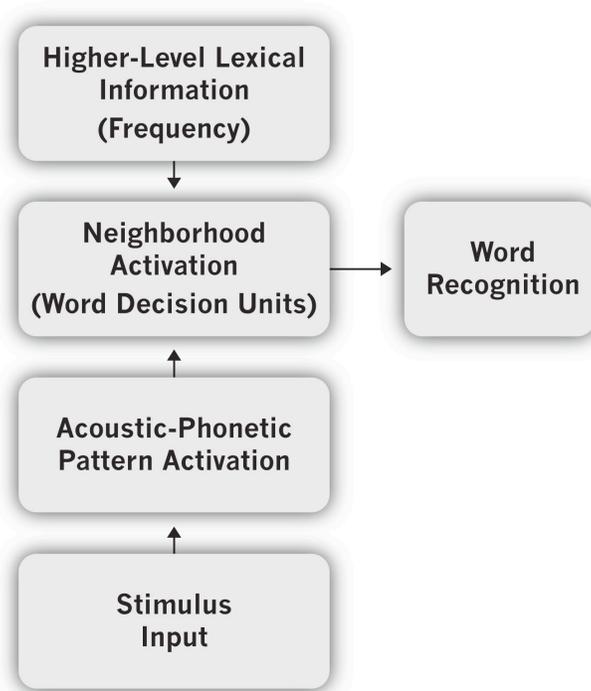
1.3.4 Neighborhood Activation Model

The Neighborhood Activation Model (NAM) (Luce & Pisoni 1998) is a mathematical competition model that offers a quantitative account of SWR. NAM specifically addresses research showing that words with more similar-sounding competitors are recognized more slowly than words with fewer competitors (Luce & Pisoni 1998) and that higher frequency words are recognized more readily than less frequent words (Marslen-Wilson 1990). Furthermore, unlike TRACE and Shortlist, NAM posits competition effects without lateral inhibition between candidates and does not posit the use of a pre-lexical representation.

A crucial point brought up in this model is that a word's prior probabilities (e.g. lexical frequency, semantic plausibility) affect recognition in concert with bottom-up acoustic evidence. Probabilities are computed for each matching phonetic pattern taking into account word frequency, the activation level of the pattern (degree of match to the input pattern), and finally the activation levels and frequencies of all other competitors. Only once all of these factors are taken into account does the highest computed

probability win. This model suggests that for context-free environments where differences in phonemic content are not robust enough to dominate the competition, lexical frequency would play an increased role in the process. The flow of information in NAM is shown in Figure 1-4.

Figure 1-4: *Information flow in the Neighborhood Activation Model. (Adapted from Goldinger, Luce & Pisoni 1989.)*



NAM predicts *neighborhood density effects* in that words with more competitors will be recognized more slowly than words in lower density neighborhoods. This effect has been demonstrated in several word-spotting studies. In one such study (Norris, McQueen & Cutler 1995), participants were presented with nonsense words and were asked to indicate the presence of a particular real word embedded in the nonsense string. These authors found for example that *mint* is easier to detect in *mintowf* [mɪntəʊf] than in *mintayf* [mɪnteɪf]. These authors proposed that this discrepancy is due to the fact that the sequence /teɪ/ is a much more common onset in English than the sequence /təʊ/. There is less lexical competition for *towf* and the word preceding it is recognized faster, offering evidence that lexical competition is modulated by the number of activated competitors in the following syllable and that detection can be retarded by competition.

Vroomen and de Gelder (1995) also investigated the effects of number of competitors on lexical competition using cross-modal priming in a lexical-decision task in Dutch. These authors found evidence that the degree of inhibition in priming is

proportional to the number of competitors in cases of phonetic overlap, concluding that competition can be “quantified in the sense that the level of competition is a function of the number of available competitors” (p. 104).

1.4 The Segmentation Problem

In the current section, we focus on an issue which has generated extensive research in the domain of SWR, namely the question as to how the speech signal is segmented into discrete processing units. The term *segmentation* is used throughout this dissertation to refer to the localization of word and/or syllable boundaries in the speech signal. The prosodic domain as well as the stage of lexical access within which segmentation strategies apply depend largely, as we will see, on the theory of speech segmentation to which one subscribes. Likewise, the classification of the *unit* into which speech is divided has generated a great deal of research, as well as considerable controversy. In reviewing the current literature on speech segmentation, we remain neutral with respect to both the prosodic domain as well as the unit exploited in the segmentation of speech, but we return to these issues below in Chapter Nine.

As noted above, the speech signal lacks explicit and reliable word boundaries as exist in written texts (see Norris, McQueen, Cutler & Butterfield 1997 for a review). Add to this the degree of variability in the speech signal, where phonetic and phonological variation can cross both syllable and word boundaries, and the segmentation of continuous speech into meaningful processing units seems all the more daunting for the listener. The literature is generally divided between two theories as to how the speech signal is segmented into lexical units. First is the theory that segmentation is the by-product of lexical competition, while a second theory proposes that segmentation is based on acoustic and/or phonological cues to boundaries in the signal. We discuss each in turn below.

1.4.1 Lexically-based Speech Segmentation

Lexically-based segmentation centers on the notion that the division of the speech signal emerges as a result of competition between candidates in the lexical access processes explained above. This view proposes that adult listeners already possess a well stocked mental lexicon and consequently recognize where one word begins by identifying where the preceding word ends. Competition models such as Cohort, TRACE, and Shortlist do not posit specialized mechanisms for the identification of word boundaries.³ The signal is segmented into non-overlapping words when the competition process achieves an optimal parse of the signal. The segmentation of the signal therefore occurs solely at the lexical level, subsequent to the identification of words in the lexicon.

In TRACE, for example, which is a *continuous* activation model, words can be activated in the network by a phonemic match at any point in the word. TRACE builds segmentation into the model by increasing lateral inhibition of activated words whose phonemic content overlaps. Once a word is definitively recognized, the activation levels

³ Later versions of Shortlist however do incorporate metrically-based segmentation, which will be discussed below in §1.4.2.1.

of overlapping competitors are significantly decreased. In Shortlist, segmentation occurs in a fashion similar to TRACE, the only noteworthy difference being that, as explained above, competition occurs among only those words that have passed through to the competition phase.

The most obvious benefit of segmentation via competition is that word boundaries emerge independently of explicit acoustic marking. As we will explore further below, acoustic cues to word boundaries in the speech signal are not always present or can be noisy and/or unreliable, making lexical segmentation an attractive strategy.

This type of segmentation strategy however poses problems for theories of development and acquisition. It is not explicitly clear how lexically-based segmentation functions when the perceptual system is confronted with input that does not map onto previously stored representations. For example, an infant has not yet acquired a mental stock of lexical items and accordingly is obliged to segment speech by means other than word recognition. Word boundaries must first be located in order to extract discrete lexical units from the signal to add to the lexicon. Many researchers point out that it is more realistic to assume that the development of word recognition proceeds first via the exploitation of statistical and probabilistic cues, which are discussed in the next section. Evidence for this progression of development comes from research showing that infants as young as three weeks old can discriminate word boundaries, well before lexical entries have been acquired (Christophe, Dupoux, Bertoncini & Mehler 1994). Further developmental evidence has suggested that infants may have specialized segmentation mechanisms that are sensitive to cues to word boundaries such as probabilistic phonotactics and lexical stress (Jusczyk 1997), which will be discussed further in Chapter Three. We now turn to a discussion of the use of acoustic cues in the adult processing system.

1.4.2 Acoustic and Phonological Cues to Word Boundaries

An ever-growing body of research has established that listeners are sensitive to acoustic-phonetic and phonological detail at multiple levels of linguistic organization, ranging from sub-phonemic variation to prosodic structure. A second theory of segmentation is based on the exploitation of this detail in the identification of word and syllable boundaries. A critical difference between lexically-based segmentation and this view is that acoustic-phonetic cues operate at the *pre*-lexical level of processing, i.e. before lexical units have been activated in the lexicon. This section will examine three separate cues to the segmentation of speech: prosodic structure, probabilistic phonotactics and allophonic variation.

1.4.2.1 Prosodic Structure

A substantial body of research suggests that native listeners make use of the rhythmic characteristics of language to identify word boundaries in the speech stream. The syllable was originally assumed to be a natural (and universal) unit for processing both because it is the most natural domain of phonological processes such as coarticulation (see for example Massaro 1974; Liberman & Studdert-Kennedy 1978) and because a line of research showed that mental representations of syllables are more

accessible to naïve speakers and listeners than other units such as the phoneme (see for example Morais, Cary, Alegria & Bertelson 1979).

In their classic study, Mehler, Dommergues, Frauenfelder, and Segui (1981) investigated the processing of syllables in French through the use of a syllable-monitoring task. In this task, participants were asked to respond as quickly as possible when they heard a particular syllable target with either a CV or a CVC structure. These authors demonstrated that targets are accessed more easily when syllable boundaries align than when syllable boundaries are misaligned. The sequence /ba/, for example, was identified more quickly in *ba.lance*, where syllable boundaries match, than in *bal.con*, where the target /ba/ is embedded in /bal/. Conversely, the sequence /bal/ was identified more readily in *bal.con* than in *ba.lance*. Mehler et al. concluded that the syllable “constitutes a unit of speech processing” in language comprehension and that syllables could well serve as accessing units” to the mental lexicon (1981: 303-4), leading researchers to propose that human language was universally represented in syllables⁴.

However, because the Mehler et al. study tested French stimuli on native French speakers, it failed to take into account the fact that different languages could exploit different rhythmic strategies. Further study replicated the Mehler et al. findings to varying degrees for Spanish, Catalan, and Italian (Pallier et al. 1993; Sebastián-Gallés et al. 1992; Tabossi et al. 2000 respectively) but failed to do so for speakers of English. Cutler, Mehler, Norris and Segui (1986) tested both French and English speakers with similar stimuli as used in the Mehler et al. study. In this study both monolingual participant groups were tested on stimuli from both languages. The authors found syllable effects for French-speaking participants when they were listening to both French and English, but found no syllable effects for English speakers in either language – the targets /ba/ and /bal/ were detected with equal speed in both *balance* and *balcony*. These results suggest that the syllable plays a special processing role in French as compared to other languages. Indeed, syllable boundaries tend to be easily identifiable in French as compared to languages with a stress-based rhythmic structure. This mostly stems from the fact that open (vowel-final) syllables represent the dominant syllable pattern in French. Adda-Decker, Mareüil, Adda, and Lamel (2002), for example, examined a corpus of 30 hours of French radio speech and found that 80% of the syllables in the corpus were open and included the syllable types CV, V, CCV, CCCV. The CV structure specifically accounted for 55% of these. Similar proportions are cited in earlier work (Léon 1964: 59).

In English, however, intervocalic consonants are argued to be ambisyllabic in that their syllabic assignment is ambiguous in that, for example, the /l/ in *balance* may be heard as the onset of *-lance* or the coda of *bal-* (see for example Kahn 1980; Gussehoven 1986). Treiman and Davis (1988) showed that speakers of English were inconsistent in the syllabification of intervocalic consonants. In a syllable inversion task, in which participants are asked to reverse the order of syllables in a word such as *lemon*, these

⁴ Cutler (2001) has more recently called for a distinction to be made between *access unit* and *processing unit*, proposing that prosodic structure produces “segmentation effects, not recoding effects” (p. 7). She stresses that, though a unit such as the syllable may be used to access lexical representations, this does not necessarily entail that abstract representations of speech are encoded in this same unit.

authors found that participants produced *on-lem* just as often as *mon-le*, suggesting that mental representations of syllables in English are not fixed. In French, however Content, Kearns and Frauenfelder (2001) used a similar task to test French speakers' syllabic representations and found that in a word such as *palais* /paʁe/ 'palace', intervocalic /l/ was syllabified as the onset of the second syllable 97.6% of the time. As noted by these same authors, the very fact that an intervocalic consonant in English can be ascribed to both syllables that bound it suggests that a processing routine based on syllable boundaries would be extremely inefficient.

Not only does English differ from French in the transparency of syllable boundaries, but English maintains an important metrical distinction between strong and weak syllables. In English, strong syllables contain full, unreduced vowels, whereas vowels in weak syllables are often reduced to schwa. The rhythm of English is characterized by an alternation between these strong and weak syllables, and strong syllables tend to occur at the beginnings of words. In fact, ninety percent of polysyllabic content words in English begin with a strong syllable and seventy-five percent of strong syllables are word-initial (Cutler & Carter 1987).

Cutler and Butterfield (1992) investigated whether English speakers exploit this strong-weak alternation in the localization of word boundaries. They proposed that English listeners are sensitive to the statistical distribution of strong syllables and exploit this regularity in the segmentation of speech. These authors tested this theory by generating 'slips of the ear' in the laboratory environment by presenting stimuli at levels just above hearing thresholds and asking participants to report what they had heard. Participants heard utterances including words that do not adhere to the characteristic strong onset pattern of English. After hearing, for example, *conduct ascents uphill* participants reported hearing sentences such as *the doctor sends a bill*. Similarly, the phrase *by loose analogy* elicited responses such as *by Luce and allergy*. These responses suggest that participants had inserted word boundaries before the strong syllables and had deleted word boundaries before weak syllables. Listeners seemed to assume that a strong syllable was directly preceded by a word boundary and adjusted their segmentation accordingly, suggesting that for English listeners strong syllables are considered to be privileged points in the signal for initiating lexical access. The findings of Cutler and Butterfield (1992) were subsequently replicated in Dutch, a language which also has a tendency toward strong syllable onsets (Vroomen, Van Zon & De Gelder 1996).

These results taken in combination with those of syllable effects in the Romance language family led to the proposal of a more universal account for the role of rhythmic structure in segmentation: The Metrical Segmentation Strategy (MSS; Cutler & Norris 1988). The MSS proposes that segmentation based on rhythmic structure is a language-universal procedure, but that the phonological representations used in segmentation are particular to each language (or family of languages). Numerous studies have supported this account. Segmentation in English and Dutch is stress-based (Cutler & Norris 1988, van Zon & de Gelder 1993) while segmentation is syllable-based in the Romance languages (Mehler et al. 1981; Pallier et al. 1993; Sebastian-Galles et al. 1992; Tabossi et al. 2000). Otake, Hatano, Cutler and Mehler (1993) also demonstrated that Japanese segmentation is based on the sub-syllabic unit of the mora.

1.4.2.2 Probabilistic Phonotactics

Each language is unique concerning which clusters of phonemes are allowed to co-occur at both the onset and offset of a syllable. The inventory of permissible sequences can thus determine where word and syllable boundaries lie, i.e. an illegal string of phonemes would necessarily signal a boundary. For example, /pf/ is an illegal onset cluster in English, but is allowed in German (e.g. *pferd* ‘horse’). Thus /pf/ would necessarily signal a word or syllable boundary in English (e.g. *stepfather*), but in German would not.

Constraints on allowable sequences of phonemes have been shown to influence the segmentation of the speech stream. McQueen (1998) showed for example that for speakers of Dutch the word *rok* ‘skirt’ is easier to spot in *fiemrok* than in *fiedrok*. McQueen proposed that this is because the syllabification of these sequences is phonotactically determined. The sequence /dr/ is allowed in Dutch as a syllable onset, while the sequence /mr/ is an illegal cluster. Consequently, the only possible segmentation for /mr/ is /m.r/ where the sequence straddles a syllable or word boundary, while the recognition of *rok* is slowed by competitors beginning with /dr/. McQueen found similar effects for English. The word *rock* was recognized more readily in *foomrock* than in *fooprock* for the same reasons. The sequence /mr/ is not allowed as an onset in English, while /pr/ is a common onset. Effects of phonotactics on speech segmentation have been further replicated for English (Weber 2000a), as well as for German (Weber 2000b), and Cantonese (Yip 2004).

Phonotactic constraints are not limited to allowable consonant clusters. Restrictions on the number of morae per syllable in Japanese have also been shown to affect segmentation processes (McQueen, Otake & Cutler 2001). Further evidence of phonotactic influence on speech segmentation comes from languages that require vowel harmony, a constraint on vowels that can occur together within a word. All vowels in a single lexical item must belong to the same class (e.g. front vs. back or rounded vs. unrounded). Thus, if two adjacent syllables have incompatible vowels, a word boundary necessarily falls between them. Work undertaken by Suomi, McQueen and Cutler (1997) suggested that speakers of Finnish, a language which utilizes vowel harmony, do indeed make use of this information in the online segmentation of the speech signal. Listeners for example found it easier to detect the nonsense string *hymy* in *puhymy* (where there is mismatch between vowel categories and a boundary is required) than in the string *pyhymy* (where vowels are compatible).

1.4.2.3 Allophonic Segmental Variation⁵

We have already discussed several phenomena which give rise to segmental variation in the speech signal, namely, coarticulation, assimilation, and speaker variability. Here we add to these phenomena the location of a segment within a prosodic hierarchy. There is a long tradition of research concerning the acoustic-phonetic

⁵ Discussion here is limited to allophonic variation in languages other than French, which is discussed in detail below in Chapter Two.

differences that arise at boundaries in the speech signal at multiple levels of prosodic organization (e.g. syllable, word, intonational phrase). As early as 1947, Pike noted a distinction between ‘phonologically recognizable’ and ‘not phonetically perceptible’ variation that occurred at such boundaries (p. 162). He observed that though variation may not be ‘phonetically perceptible’ to the listener, it is nonetheless systematically present in the acoustic signal and observable in laboratory analyses. Troubetzkoy (1949) made a further distinction between ‘positive’ and ‘negative’ boundary signals in that a signal may mark either the presence or absence of a boundary. He also differentiated between phonemic boundary markers, as in the case of phonotactic constraints discussed above, and non-phonemic boundary markers, as in the case of allophonic variation which we turn to now.

A continuously growing body of research has shown that listeners are sensitive to variation at multiple levels of linguistic organization. One of the earliest studies of allophonic variation at word boundaries (Lehiste 1960) examined minimal pairs rendered ambiguous by the lexical assignment of a pivotal consonant (e.g. *grade A/grey day; plate ought/play taught; known ocean/no notion; lawn chair/launch air*). In this study, participants listened to recordings of such ambiguous pairs and identified what they heard. Participants were able to correctly identify over two-thirds of the phrases, which Lehiste took as evidence of the presence of some type of acoustic differentiation between the pairs, even though their phonemic content is identical. What she observed through subsequent acoustic analyses was that those minimal pairs which were most easily identified contained acoustic markers on either side of the word boundary (as opposed to throughout the whole sequence). She also found that these cues depended on the segment or combination of pre-and post-boundary segments. For example, she found that vowel-initial words were marked by glottalization before vowel onsets. She also observed the aspiration of word-initial stops, durational variation in the consonants /l/ and /r/ and lengthening of pre-boundary vowels.

Nakatani and Dukes (1977) endeavored to determine more precisely where the cues to the location of word boundaries lie. Using cross-spliced examples of the minimal pairs used by Lehiste, these authors investigated four possible locations of boundary indicators that had been identified by Lehiste. They divided each two-word sequence into four portions: the beginning of Word 1 up to the onset of the pivotal consonant (/pleɪ/ in *play taught* and *plate ought*); from the onset of the pivotal consonant to the middle of the pivotal consonant (i.e. the first half of /t/ in *play taught* and *plate ought*); from the middle of the pivotal consonant to the offset of the pivotal consonant (i.e. the second half of /t/ in *play taught* and *plate ought*) and finally from the offset of the pivotal consonant to the end of Word 2 (i.e. /ɔ:t/ in *play taught* and *plate ought*). The authors then cross-spliced these four portions resulting in all 16 possible combinations for each minimal pair. Participants listened to each altered stimulus and were then given four choices representing all combinations of the two words (e.g. *play taught, plate ought, play ought* or *plate taught*) in a forced-choice identification task. They found that participants were able to correctly identify the ambiguous two-word sequences most of the time, even up to 100% of the time for certain pairs. Their evidence suggested that the initial and final

portions of the splices (e.g. /pleɪ/ and /ɔ:t/) did not determine participants' responses, even though significant durational differences in these portions were observed. The evidence instead suggested that listeners most consistently attended to the offset of the pivotal consonant in the determination of their responses, including the presence of laryngealization and glottal stops before vowel-initial words. Their results also suggested that the onset of the pivotal consonant is also a reliable cue, but to a lesser extent than the offset.

A great deal of research since these early attempts has confirmed and extended the findings of Lehiste and Nakatani and Dukes. The bulk of this work has demonstrated that the most reliable cues to the location of word boundaries are located at the beginnings of words. Onsets seem cross-linguistically to be the most marked by features such as aspiration (Lehiste 1960; Nakatani & Dukes 1977), laryngealization and glottal stops preceding word-initial vowels (Lehiste 1960; Hoard 1966; Nakatani & Dukes 1977), the ratio of closure duration to duration of entire stop consonant (Boucher 1988), the occurrence of full vowels (Cutler 1990; Cutler & Carter 1987; McQueen & Cutler 1992), more careful articulation of vowels (Gow & Gordon 1993), shortening of consonants in clusters (Lehiste 1960; Christie 1977) as well as the lengthening of both phonemes and syllables at onsets (Klatt 1973; Lehiste 1972; Oller 1973; Umeda 1975; Nakatani & Schaffer 1978; Beckman & Edwards 1990; Gow & Gordon 1993, 1995; Quené 1992) which we discuss in further detail now.

1.4.2.3.1 Duration as a Segmentation Cue

The duration of phones has consistently been shown to vary as a function of the segment's position in a prosodic hierarchy (Klatt 1976; Lehiste 1972; Nakatani & Schaffer 1978; Fougeron 2001). Initial segments are systematically longer than the same segment in medial or final position in English (Lehiste 1961; Klatt 1976; Gow & Gordon 1995), French (Fougeron & Keating 1997; Fougeron 2001) and Dutch (Shatzman & McQueen 2006). These results are all consistent with findings that suggest that speakers strengthen the articulation of segments at the edges of prosodic domains (Cho & Keating 2001; Fougeron 2001; Cho et al. 2007).

Duration as a function of word length was shown by Lehiste (1972), who demonstrated that the duration of an entire syllable decreases as the number of syllables in a word increases. For example the sequence [sli:p] is longer in the monosyllabic *sleep* than in *sleepy*, and longer in *sleepy* than in *sleepiness*. Klatt (1976) also found that a phonemic sequence in a monosyllabic word is an average of 15% longer than the same sequence in a polysyllabic word. Similarly, Turk and Shattuck-Hufnagel (2000) found that, for example, the sequence /tju:n/ was longer in *tune acquire*, where produced as mono-syllabic, than in *tuna choir*. Salverda et al. (2003) further confirmed this showing that mono-syllabic tokens (e.g. *ham*) were longer than the same phonemic sequence in a polysyllabic word (e.g. *hamster*).

As noted by Lehiste (1972), however, identifying the presence of allophonic variation such as duration does not necessarily entail that listeners make use of this variation in the online comprehension of speech. Klatt (1976) also questioned whether

durational differences in segments are actually exploited in word recognition. He was of the opinion that, for English, durational differences encode ‘considerable information in connected speech’ (p. 1220), but he was not convinced that listeners make use of these durational differences in running speech as a comprehension strategy. Indeed he posed the question, “When is a durational rule perceptually motivated and when is it a consequence of constraints on the production mechanism?” (p. 1220) He also speculated that it may be impossible to make use of durational cues until after a listener has heard an entire utterance. Only then, he points out, is it possible to evaluate whether lengthening or shortening is due to a segment’s position in a prosodic hierarchy, stress, or emphasis or other speaker-intentional factors.

Since the work of Lehiste and Klatt, a large body of research has indeed shown that listeners are sensitive to variation in segmental duration. Acoustic correlates of syllable structure showed an effect on fragment priming in Italian (Tabossi et al. 2000). Participants were presented with CVC primes produced either as CV.C or CVC. and made lexical decisions to visual targets containing these primes. For example, participants heard the fragment /sil/ produced as the beginning of either *silenzio* ‘silence’ or *silvestre* ‘woodland’. Results showed that priming effects were stronger when fragments matched the syllabification of the target, a difference these authors attributed to durational variation of the pivotal consonant /l/ (though they did not directly test this hypothesis).

Similarly, Warner, Jongman, Sereno, and Kemps (2004) found that speakers of Dutch use non-contrastive durational differences to differentiate pairs of words said to be ‘homophonous’ due to word-final obstruent devoicing, such as *meet* ‘measures’ and *meed* ‘avoided’, both transcribed [me^ht]. In the first part of this study listeners were asked to identify the presence of final (orthographic) *-t* in minimal pairs recorded by four different native speakers of Dutch. Listeners were able to correctly identify *-t*, differentiating it from orthographic *-d*, significantly more than half the time. Warner et al. attribute these results to the presence of durational differences in the vowels preceding the final consonant in each word. They found that the vowel preceding the (orthographic) voiced consonant was an average of 3.5 milliseconds (ms) longer than the vowel preceding the (orthographic) unvoiced consonant. Even though this durational difference was systematically present in this production sample, it should be noted that it is extremely unlikely that a difference of this length would be perceptually salient (see for example Huggins 1972; Klatt 1976; Lehiste 1976).

However, in the second part of Warner et al. (2004), these authors instrumentally lengthened the vowel in the voiced member of each minimal pair. In exaggerating this durational difference the researchers were hoping to pinpoint segmental duration as the cue to underlying final voicing. Indeed, in a second identification task employing the manipulated stimuli, participants performed significantly better than in the first portion of the study. The researchers note that due to the extremely small difference in these vowels “it seems unlikely that listeners make use of this cue in perceiving natural speech.” However, the results of the second identification task using instrumentally modified stimuli show that “under the best of conditions, they are able to do so” (p. 266), suggesting that vocalic duration in this environment does indeed have cue value.

The studies reviewed thus far in this section have examined duration as a cue to the identification of words in isolation. Research within the domain of speech segmentation has also demonstrated that durational differences can effect the localization of word boundaries in continuous speech. Quené (1992) for example analyzed ambiguous CVCVC sequences in Dutch such as *die pin* ‘that pin’ (CV.CVC) and *diep in* ‘deep in’ (CVC.VC) and observed acoustic variation in the pivotal consonant /p/ as well as in the preceding and following vowels (V₁ and V₂ respectively). The consonants produced as word-initial were an average of 21 ms longer than the same consonant produced in final position. Quené did not find significant durational differences in V₁ and V₂ according to boundary location, though he did find significant differences in decay time for V₁ and rise time for V₂ (rise time refers to the duration from the vowel onset to the point at which amplitude reaches ninety percent of its maximum value; decay time is the duration from ninety percent to offset). Quené found that Dutch listeners were eighty percent correct in differentiating the ambiguous CVCVC sequences in a forced-choice identification task. In the second portion of this study Quené manipulated the acoustic factors for which he had found significant differences and once again conducted a forced-choice identification task. What he found was a significant correlation between the duration of the pivotal consonant and participants’ responses, but no correlation with acoustic variation in vowels. Quené proposed that “...listeners can locate word boundaries in connected speech on the basis of durational cues alone, even if no obvious correlates of word boundaries (viz. boundary segments) are present in the signal” (p. 334).

Another line of research has examined acoustic-phonetic differences that signal word boundaries in embedded words. Davis, Marslen-Wilson, and Gaskell (2002) found that lexical access processes were sensitive to subtle durational differences between a mono-syllabic word and the same phonemic sequence embedded in a longer word (e.g. *cap* and *captain*). This study employed a cross-modal priming task, in which participants heard /kæp/ produced either as mono-syllabic *cap* or as the first syllable of *captain*. The results showed significantly more activation of the shorter word when participants were presented with mono-syllabic productions, and conversely more activation of the longer sequences when participants were presented with a spliced portion of the disyllabic sequence.

More recently, using eye-tracking, Salverda et al. (2003) also demonstrated that durational differences in syllables are sufficient to guide listeners in determining whether a sequence represents a mono-syllabic word or forms part of a longer word. Participants listened to stimuli where, for example, the first syllable of the word *hamster* had been replaced with productions of the mono-syllabic word *ham*. When instructed to *click on the hamster* participants looked more often and fixated longer on a picture of a ham when presented with a stimulus containing the mono-syllabic word spliced into the disyllabic word than when presented with the unaltered recording of disyllabic *hamster*. These results suggested that the longer [hæm] sequence signaled a prosodic boundary and led to the activation of mono-syllabic *ham*.

Interestingly, not all mono-syllabic tokens in the Salverda et al. study were actually realized with longer durations, and the eye-tracking effects were lost when shorter mono-syllabic tokens were spliced into the disyllabic words. In other words,

duration of the sequence and not its lexical origin predicted responses. This offers compelling evidence that it is in fact solely duration that is guiding the participants' responses in this task.

Similarly, Shatzman and McQueen (2006) found that the online interpretation of ambiguous sequences in Dutch can be influenced by segment duration alone. This study investigated the recognition of sequences rendered ambiguous by the lexical assignment of /s/ such as *eens pot* 'once jar' and *een spot* 'a spotlight'. They found variation in several acoustic factors in the ambiguous pairs such as segment duration, closure duration of the stop consonant, the duration of the entire word, root mean square energy of /s/, root mean square energy of the stop consonant following /s/. However, the duration of the segment /s/ alone was significantly predictive of participants' performance on an identification task that also tracked eye movements. In a second experiment in this study, the researchers instrumentally manipulated the duration of /s/ by both shortening and lengthening this segment while holding all other information in the signal constant. Even stronger correlations between responses and the duration of /s/ were found with manipulated stimuli, confirming that duration was a sufficient and reliable cue to the segmentation of these ambiguous sequences.

The exploitation of sub-phonemic phonetic detail in the lexical interpretation of ambiguous input is also in line with research showing that sub-phonemic mismatch can slow lexical access. Marslen-Wilson and Warren (1994) for instance demonstrated that response latencies were longer in the recognition of *job*, when the sequence *jo-* was spliced from a token of *jog*. Similarly, Andruski, Blumstein and Buton (1994) showed that the manipulation of voice onset time (VOT) can affect lexical access. Voiceless-stop initial words in which VOT was shortened were activated less strongly than words with normal VOT values in a cross-modal priming task. These results are consistent with a host of studies demonstrating the effect of mismatching information on lexical access (Connine, Blasko & Wang 1994; Connine, Titone, Deelman & Blasko 1997; Frauenfelder, Scholten & Content 2001; Marslen-Wilson, Moss & van Halen 1996; Soto-Faraco, Sebastián-Gallés & Cutler 2001).

1.4.3 The Use of Multiple Cues in Speech Segmentation

What conclusions can be made from the above body of research? The body of literature discussed above offers a somewhat contradictory picture of acoustic cues available to listeners in the speech stream. Though the bulk of research has supported the notion that there are systematic acoustic cues in the speech stream that correspond to multiple prosodic boundaries, this same body of work has demonstrated that these cues are noisy, unreliable, and often not present in the signal. In addition, the vast majority of works cited here deal with speech produced in the laboratory environment; therefore generalizations to processing in natural speech environments must be made with caution.

Clearly, an ideal of psycholinguistics is to model *universal* SWR processes and not the exploitation of particular cues as they pertain to particular languages. However, the tendency in the literature (this dissertation included) is to isolate and focus on a single cue to segmentation in laboratory environments, deliberately removing and controlling for other cues that might be present in the signal. Much less research has offered any sort

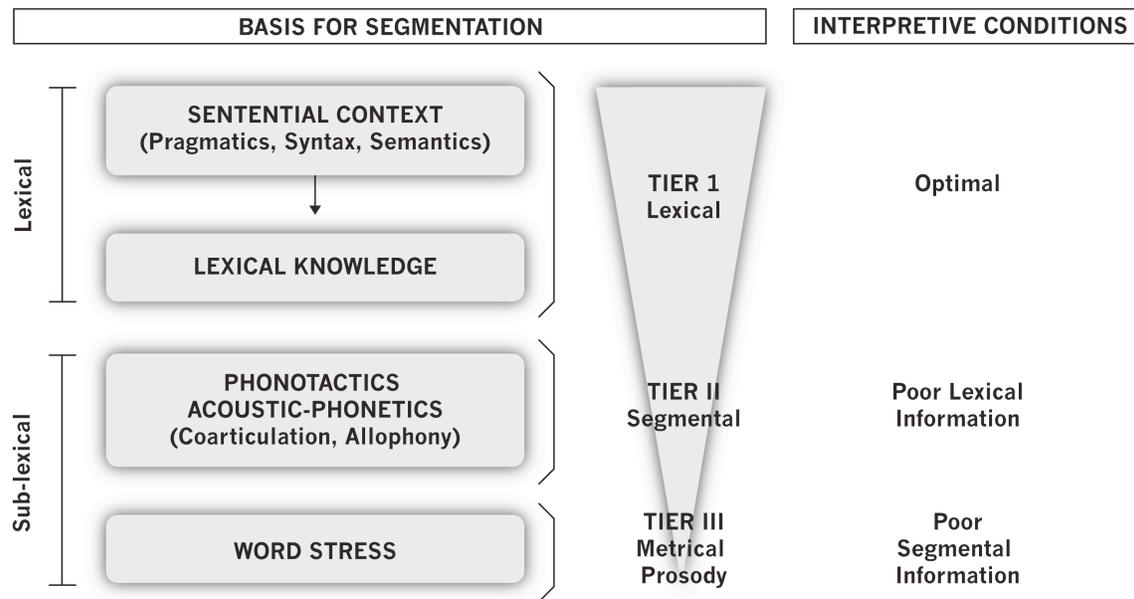
of unified model involving the simultaneous exploitation of multiple cues at different processing levels in the segmentation of connected speech.

Recent work by Mattys and colleagues, however, has begun to explore the weighting of multiple segmentation strategies in an attempt to formulate a hierarchy of cues based upon the saliency of each individual cue in speech processing. For example, Mattys (2003) showed differential sensitivity to stress and phonotactic cues in English listeners when these cues were presented in clear speech as opposed to noise. When the two cues were pitted against one another in clear speech participants showed more sensitivity to phonotactic constraints, but more sensitivity to stress when stimuli were presented in noise. A further study investigating the simultaneous exploitation of stress and coarticulation showed that coarticulation outweighed stress in clear speech, while stress outweighed coarticulation in a signal presented in noise (Mattys 2004).

Further research by Mattys, White and Melhorn (2005) utilizing a cross-modal fragment priming paradigm compared the strength of *knowledge-driven* (i.e. top-down; lexical) cues with *signal driven* (i.e. bottom-up; acoustic) cues in the identification of English words. In this series of studies, participants made lexical decisions about visual targets after listening to nonsense utterances in which the final portion of the utterance serves as an auditory prime for a visual target (e.g. the nonsense sequence /revə'mærə/ would serve as a prime for the visual target *marathon*). By manipulating coarticulation, phonotactics and lexical cues in the sequence preceding the fragment prime, these authors were able to compare the relative priming effects of each cue. Specifically, they found that, in English, lexicality (e.g. whether the sequence preceding the fragment is itself a word or non-word) is given preference over segmental cues (e.g. phonotactic and acoustic-phonetic variation) and that segmental cues are in turn given preference over suprasegmental cues (e.g. whether target fragment appears in a strong or weak syllable). The results of several experiments examining conflicting cues presented in both clear speech and in noise led these authors to suggest that knowledge-driven cues are privileged when all cues are “optimally available”. Signal driven cues, on the other hand, gain prominence when top-down information is “unavailable, impoverished, or ambiguous” (p. 487).

Mattys et al. (2005) present the first empirically driven attempt at integrating previously documented cues to speech segmentation into a comprehensive processing system that mirrors the complexity of natural speech. This hierarchy attempts to capture the fact that, though each cue presumably has an independent effect on the activation of lexical candidates, some cues trump others when multiple cues are available to the listener. This weighting fluctuates depending on the saliency or availability of other cues at any given point in the signal. Mattys et al. (2005) present a schematic of this hierarchy which we have adapted here in Figure 1-5 below.

Figure 1-5: *Hierarchy of cues to speech segmentation. (Adapted from Mattys et al. 2005).*



Given that the speech signal is characterized by considerable amounts of variation and an often irregular distribution of processing cues, the work of Mattys and colleagues attempts to offer a more realistic view of speech processing based on the simultaneous exploitation of multiple cues. This work, combined with existing research on cues to processing and segmentation, would lead us to conclude that no single segmentation cue is either necessary *or* sufficient in the processing of *natural* speech.

CHAPTER TWO: Spoken Word Recognition in French

*De l'oral à l'écrit, il y a un monde. La différence est si grande
que la description du français oral ressemble plus souvent à
celle d'une langue exotique qu'à celle du français écrit.*
(Morel & Danon-Boileau 1998:7)

[From the oral to the written, there is a whole world.
The difference is so large that the description of oral French more
often resembles that of an exotic language than that of written French.]

2.0 Introduction

The effects of resyllabification referred to in the introduction to this dissertation make spoken French a particularly interesting case study for models of lexical access and speech segmentation. Resyllabification in French results in the frequent misalignment of syllable and word boundaries due to three phonological processes that occur at the juncture between two words (henceforth W_1 and W_2): elision, *enchaînement*, and liaison. These processes serve both to avoid hiatus (the occurrence of two consecutive vowel sounds with no intervening consonant) and to preserve an open (vowel-final) syllable structure. Elision refers to the omission of a word-final vowel sound in W_1 when W_2 is vowel-initial. For instance, *le ami* /lə#ami/ 'the friend' (masculine) is produced as *l'ami* [la.mi] where the /ə/ of the definite article *le* is dropped. *Enchaînement* occurs when W_1 is consonant-final and W_2 is vowel-initial. The coda of W_1 is resyllabified across the word boundary to become the onset of W_2 . The phrase *une amie* 'a friend' (feminine) is thus produced as [y.na.mi] where syllable and word boundaries are mismatched, instead of [yn.a.mi] where boundaries would be aligned. Liaison on the other hand concerns consonants in final position that are represented graphically⁶, but are latent when the

⁶ Except in cases of *epenthetic* liaison, as in *quatre* [z] *enfants* 'four children', in which a liaison consonant is spuriously and anomalously introduced in production but is not represented in the orthography of the W_1 .

word is pronounced in isolation or followed by a consonant-initial W_2 . The latent consonant is realized before a vowel-initial W_2 and then resyllabified through *enchaînement*. For example, the determiner *un* (singular masculine indefinite article) is pronounced [œ̃] in isolation or before a consonant (e.g. *un stylo* [œ̃.sti.lo] ‘a pen’). When preceding a vowel onset in W_2 , however, as in *un ami* ‘a friend’ (masculine), the latent /n/ surfaces and is syllabified as the onset of *ami*. Accordingly, the sequence is syllabified [œ̃.na.mi] instead of [œ̃n.a.mi] where word boundaries would be respected.

Enchaînement and liaison are two distinct phonological processes, but are nonetheless closely linked; a consonant realized in liaison is subsequently resyllabified through *enchaînement*⁷. And while *enchaînement* applies to any consonant normally produced in final position, liaison only concerns the consonants /g, n, p, t, ʁ, z/. Moreover, final consonants that are *enchaînées* do not change their relationship to their graphic form⁸. However, consonants that appear in environments of liaison can be represented by different orthographies. Table 2-1 below shows consonants that surface in liaison and their associated orthographies and pronunciation.

Table 2-1: *Consonants that surface in liaison environments in spoken French and their associated orthographies, pronunciation and examples.*

Orthography	Pronunciation	Example
g	/g/ (/k,t/) ⁹	<i>long homage</i> ‘long hommage’
n	/n/	<i>un avion</i> ‘a plane’
p	/p/	<i>trop aimé</i> ‘too loved’
t d	/t/	<i>petit ami</i> ‘boyfriend’ <i>grand ami</i> ‘great friend’
r	/ʁ/	<i>dernier an</i> ‘(the) last year’
s z	/z/	<i>les amis</i> ‘the friends’ <i>chez elle</i> ‘at/to her house’
x		<i>deux enfants</i> ‘two children’

⁷ Liaison without *enchaînement* is attested (Encrevé, 1988), but examples are limited to public discourse and carefully produced speech.

⁸ With the exception of <f> in *neuf* ‘nine’, which is pronounced /f/ when the word appears in isolation or before a consonant, but /v/ before a vowel (e.g. *neuf heures* [nœ.vœʁ] ‘nine o’clock/nine hours’).

⁹ Liaison with /g/ is extremely rare, and its realization is highly variable. This segment is discussed in further detail in Chapter Five.

The realization of consonants in liaison environments is syntactically as well as phonologically conditioned. Consequently, liaison is not realized at every word boundary where the phonological environment discussed above would render it possible. There are generally thought to be three categories of liaison as proposed by Delattre (1951, 1966): *obligatoire* (obligatory), *interdite* (prohibited) and *facultative* (optional). Examples of syntactic environments where liaison is obligatory are between determiner and noun (e.g. *un enfant* ‘the child’ [œ.nã.fã]/*[œ.ã.fã]) or between personal (clitic) pronoun and finite verb (e.g. *vous avez* ‘you (pl) have’ [vu.za.ve]/*[vu.a.ve]). Environments where liaison is prohibited or blocked occur for example between lexical singular noun and verb (e.g. *l’enfant aime* ‘the child loves’ [lã.fã.em]/*[lã.fã.tem]) or between lexical singular noun and following adjective (*l’enfant anglais* ‘the English child’ [lã.fã.ã.glɛ]/*[lã.fã.tã.glɛ]). Finally, liaison can be produced optionally, for example after certain adverbs including *trop* ‘too’ in *trop aimé* ‘too loved’, [tʁo.pe.me] or [tʁo.eme], or after the negative particle *pas* ‘not’ as in *pas encore* ‘not yet’, [pa.zã.kœʁ] or [pa.ã.kœʁ]. In contexts where liaison is optional, the choice to realize a liaison consonant is usually a stylistic one on the part of the speaker. A higher register of speech would call for more frequent realizations of liaison consonants in optional contexts. In fact, the realization of optional liaison has been observed to be the most salient phonetic feature in the marking of register in spoken French (Hannahs 2007).

2.1 Syllable-Based Segmentation in French

As discussed in the previous chapter, the syllable represents an important perceptual unit in the comprehension of spoken French (Mehler et al. 1981; Cutler et al. 1989). Subsequent research further refined the concept of syllable-based lexical access by demonstrating that misalignment incurs greater processing costs at syllable onsets than at offsets. Using a word-spotting task, Dumay, Banel, Frauenfelder and Content (1998) showed that reaction times were significantly faster in identifying the word *lac* embedded in the non-word ZUN.LAC, where *lac* is necessarily aligned with a syllable onset given that /nl/ is an illicit onset in French, than in ZU.GLAC, where /gl/ is an allowed onset and word and syllable boundaries are not necessarily aligned. However, no significant difference in reaction time was observed when misalignment occurred at syllable offsets. The same word, *lac*, was detected with equal speed in LAC.TUF as in LA.CLUF. These authors cite these results as strong evidence that syllable onsets constitute favored points of lexical access in French.

Following these results, Content, Kearns, and Frauenfelder (2001) argue against a *boundary approach* to syllabic segmentation in French, where syllable onsets and offsets are given equal weight in processing, instead favoring a segmentation routine based specifically on the identification of syllable onsets. Content et al. (2001) observe that “by definition, onsets signal the beginning of a new event, which may require the immediate allocation of attentional or memory resources [...] In contrast, decisions about the offset of an event do not have the same urgency and may even benefit from being delayed to resolve local ambiguity” (p. 179). The Syllable Onset Segmentation Heuristic (SOSH)

(Dumay, Content & Frauenfelder, 1999; Content, Dumay & Frauenfelder, 2000) captures this distinction. SOSH is a universal segmentation strategy that presumes that the onsets of syllables tend to coincide with the onsets of words and that listeners exploit this tendency by initiating segmentation attempts at syllable boundaries. This heuristic is consistent with models of SWR that privilege word onsets for lexical access (e.g. Cohort) as well as the Metrical Segmentation Strategy (Cutler & Norris, 1988) which proposes that listeners pay attention to strong onsets in English and Dutch. SOSH is also supported empirically by studies that have shown processing costs due to syllable misalignment at onsets in different languages, even those which do not rely on syllable-based segmentation such as English (Weber 2001) and Dutch (Vroomen & de Gelder 1999). Developmental data have also suggested that infants pay special attention to syllable onsets (Jusczyk 1999), lending further support to SOSH.

However, as Dumay, Content, and Frauenfelder (1999) themselves point out, “one important shortcoming of a syllable-based segmentation strategy is its difficulty in handling potential resyllabification phenomena resulting from phonological processes applying across word boundaries as in the case of French liaison, [in that]... an incorrect lexical alignment would be made on the basis of syllable onsets” (p. 281).

As we have seen, several phonological processes in French can create syllable boundaries that do not coincide with word boundaries. These processes can produce both *local* and *global* lexical ambiguities. Local ambiguity is ambiguity that is resolved once an entire utterance is processed. In the phrase *excellent tableau* [ɛk.se.lã.ta.blo] ‘excellent painting’, for example, the final /t/ of *excellent* is not realized and syllable boundaries are identical to word boundaries, while in the phrase *excellent abri* [ɛk.se.lã.ta.bʁi] ‘excellent shelter’, the final latent /t/ of *excellent* is realized through liaison and resyllabified as the onset of *abri*. The two phrases are phonemically identical through the sequence /ta/ until subsequent input resolves the ambiguity; only through the continuation of the signal can *abri* be identified as vowel-initial. The phonological processes outlined above can also create *global* lexical ambiguities — ambiguity that is not resolved even once the entire utterance is processed — as in the case of *il n’a aucun air* ‘he has no melody’ and *il n’a aucun nerf* ‘he has no nerve’, both [il.na.o.kœ.nɛʁ]. Due to the realization of liaison and subsequent resyllabification, both phrases emerge as viable parses of this sequence.

Local and global lexical ambiguities must be resolved online by the listener. The following questions therefore present themselves: What perceptual mechanism serves to disambiguate the input? More specifically, what perceptual mechanism allows a resyllabified liaison consonant (hereafter LC) to be distinguished from a word-initial consonant (hereafter IC) in spoken French, thus allowing for the activation of vowel-initial lexical candidates in the mental lexicon?

2.2 The Perception of Liaison

The prevalence of resyllabification and subsequent syllable misalignment in French would seem to hinder speech recognition and segmentation in that these processes can be slowed as a function of the number of candidates activated in the lexicon. Words

are recognized more slowly when they have more competitors (see for example Luce & Pisoni 1998; Norris, McQueen & Cutler 1995; Vroomen & de Gelder 1995, among others). The processes of liaison and resyllabification increase the number of lexical hypotheses that must be considered in that they entail the activation of both vowel- and consonant-initial candidates for W_2 . However, resyllabification and liaison have not been shown to incur processing costs.

A study by Gaskell, Spinelli, and Meunier (2002) suggested that resyllabification can in fact *facilitate* the recognition of vowel-initial words that have undergone resyllabification in spoken French. A cross-modal priming study showed significant priming as compared to a control for vowel-initial words in three conditions: liaison (e.g., *un généreux italien* [œ̃.ʒe.ne.ʁø.zi.ta.ljɛ̃] ‘a generous Italian’), *enchaînement* (e.g., *un virtuose italien* [œ̃.viʁ.tuø.zi.ta.ljɛ̃] ‘an Italian virtuoso’) and a syllable-aligned condition (e.g., *un chapeau italien* [œ̃.ʃa.po.i.ta.ljɛ̃] ‘an Italian hat’). Reaction times were measured and participants recognized vowel-initial targets preceded by resyllabified consonants, in both liaison and *enchaînement* conditions, significantly faster than targets that matched syllable boundaries. For example, *italien* was spotted more quickly in *un généreux italien*, a liaison environment, and in *un virtuose italien*, an *enchaînement* environment, than in *un chapeau italien* where syllable boundaries are aligned. Their suggestion was that resyllabification (either through *enchaînement* or liaison) is somehow acoustically marked and that this marking aids in the lexical competition process. However, they did not perform acoustic analyses on the stimuli employed in the study in order to test this hypothesis. This result extended the previous findings of Wauquier-Gravelines (1996), who showed in a word-monitoring task that *éléphant* ‘elephant’ is recognized as easily in *un petit éléphant* [œ̃.pə.ti.te.le.fã] ‘a small elephant’, a liaison environment, as in *un joli éléphant* [œ̃.ʒo.li.e.le.fã], where liaison is not possible.

Spinelli, McQueen, and Cutler (2003) further probed lexical access processes and revealed significant priming effects for both consonant-initial and vowel-initial words in globally ambiguous sentence pairs such as *c’est le dernier rognon*, ‘it’s the last kidney’, and *c’est le dernier oignon*, ‘it’s the last onion’, both [se.lə.dɛʁ.nje.ʁɔ̃.ɲɔ̃]. These researchers employed four priming conditions in a lexical-decision task: an ambiguous liaison condition (*c’est le dernier oignon*), an ambiguous non-liaison condition (*c’est le dernier rognon*), an unambiguous condition where liaison would not be possible (*c’est un demi rognon*, ‘It’s a half kidney’), and finally an unambiguous baseline condition using an unrelated word where liaison would not be possible (*c’est un ancien nitrate*, ‘It’s an old nitrate’).

Significant priming effects were found for both vowel-initial (*oignon*) and consonant-initial (*rognon*) candidates in the ambiguous conditions. In other words, the ambiguity caused by liaison and subsequent resyllabification did not impair the lexical activation of the vowel-initial candidate. Furthermore, priming effects followed the intention of the speaker, i.e. priming effects were stronger for *oignon* than for *rognon* when the speaker intended *oignon*, and vice versa. Their results also suggested that words not intended by the speaker in ambiguous contexts (e.g. *oignon* when *dernier rognon* is

produced) were activated, but not as strongly as in the intended production. Significantly, they did not find priming effects for *oignon* in an unambiguous condition where liaison is not possible (e.g. *demi rognon*), suggesting that solely the liaison environment allows for the activation of both consonant- and vowel-initial lexical candidates.

Despite research that has suggested that liaison does not impede lexical access processes, consonants that surface in liaison environments do seem to retain a unique cognitive status relative to other consonants. The complexity involved in processing LCs is attested by developmental data. Children learning French have difficulty acquiring the usage of liaison, and it is mastered relatively late in the language acquisition process (Basset 2000; Wauquier-Gravelines & Braud 2005). Production errors in child speech demonstrate that the resyllabification of LCs can affect the mental representation of vowel-initial words in development. For example, Wauquier-Gravelines (2003) noted mistakes in child production such as *[le.ne.le.fã] instead of [le.ze.le.fã] for *les éléphants* ‘the elephants’. This error suggests that the child had been exposed to the input *un éléphant* [œ.ne.le.fã] ‘an elephant’ and analyzed the LC /n/ of *un* as the onset of W_2 instead of encoding it as a resyllabified consonant. Similar errors include a misanalysis of *oiseau* /wazo/ ‘bird’ as in *[œzwazo] for *un oiseau* [œnwazo] ‘a bird’ where children parsed previous input *les oiseaux* [le.zwa.zo] ‘the birds’ as W_2 /z/-initial. Dugua (2002) also investigated this phenomenon in a task that elicited productions of vowel-initial words in isolation by asking children to call out to a series of animals (e.g. *Ours! Viens ici!* ‘Bear! Come here!’). She found that 15% of child productions were of the type *Nours!* (from *un ours* [œ.nuks] ‘a bear’) or *Nâne!* (from *un âne* [œ.nan] ‘a donkey’), suggesting that mental representations of these lexical items retained the resyllabified liaison consonant.

Adult data suggest that liaison environments continue to be processed differently than contexts where liaison is not possible. Though lexical access is not delayed by liaison as discussed above, one study showed that SWR can be slowed by *potential* liaison. In a word-monitoring task, Dejean de la Bâtie and Bradley (1995)¹⁰ asked participants to detect /t/-initial words in four conditions: a /t/-initial word preceded by a word that could potentially trigger liaison with /t/ (e.g. *grand théâtre* ‘big theater’); a /t/-initial word preceded by a word where liaison would not be possible (e.g. *vrai théâtre* ‘real theater’); a vowel-initial word preceded by liaison with /t/ (e.g. *grand éléphant* ‘big elephant’); and finally a vowel-initial word preceded by a word where liaison would not be possible (e.g. *vrai éléphant* ‘real elephant’). These phrases were embedded in the semantically neutral phrase *C’est un*¹¹... ‘It’s a ...’. Reaction times were measured and participants were slower to detect a word in a potential liaison context (e.g. *grand théâtre* ‘big theater’) than in a non-liaison context (e.g. *vrai théâtre* ‘real theater’). The authors

¹⁰ This study also tested learners of L2 French. Results for non-native speakers will be discussed below in Chapter Four.

¹¹ The phrase ‘*C’est un*...’ also presents an optional liaison environment with /t/ in *est*. The authors do not address this fact, nor do they specify whether this portion of the stimulus was produced with or without liaison, but it should be noted that this could have affected responses.

suggest that when a word ending in consonant that can trigger liaison, such as *grand*, is encountered in the input, the possibility of liaison occurrence delays the lexical segmentation process. Longer reaction times suggest that listeners wait for more of the signal to unfold before parsing the sequence, demonstrating that the perceptual system processes environments of liaison differently in the SWR process. These results combined with those cited above suggest that, somewhat paradoxically, it is not the presence of liaison that seems to slow processing, but the *potential* presence of liaison.

In the second portion of this study, the four conditions of stimuli were embedded in sentences with biasing context. Instead of appearing in the semantically neutral phrase *C'est un...* 'It's a...', the phrases were preceded by contextual information. For example, *excellent acteur* 'excellent actor' and *vrai acteur* 'real actor' were preceded by *Dans ce film, j'ai découvert un...* 'In this film, I discovered a ...'. While there had been a significant difference in reaction times between conditions of potential liaison and non-liaison in semantically neutral contexts, this difference disappeared when the phrases were placed in a contextualizing frame. These further results point to the fact that, though locally ambiguous acoustic input may delay processing, top-down factors such as semantic context can override ambiguity and eliminate online processing costs.

Further demonstrating that LCs possess a distinctive mental representation in the phonological grammar relative to other consonants, Wauquier-Gravelines (1996) showed that /t/ and /n/ are harder to detect in liaison environments than in word-initial position. This study used two phoneme-monitoring tasks and recorded both response latencies and detection accuracy. Reaction times indicated that participants had more difficulty detecting the presence of /t/ in *un grand éléphant* [œ.gʁɑ̃.te.le.fɑ̃] 'a big elephant', where /t/ surfaces and is resyllabified, than in *un grand téléphone* [œ.gʁɑ̃.te.le.fon] 'a big telephone', where /t/ is fixed as lexical-word initial. Similarly, /n/ was more difficult to detect in *un avion* [œ.na.vjɔ̃] 'a plane' than in *un navire* [œ.na.viʁ] 'a ship'. Wauquier-Gravelines attributes the response latency effect to a sort of 'deafness' toward the LC based on a difference in phonological status between LCs and ICs. This issue is discussed in further detail below.

Interestingly, Wauquier-Gravelines (1996) also found that accuracy rates were significantly lower for /n/ than for /t/. The segment /n/ was detected only 44.6% of the time when realized as a LC, but 87.5 % of the time as an IC. The segment /t/ on the other hand was detected 67.8 % of the time in liaison environments and 92.8 % of the time as word-initial. In accounting for this variation, Wauquier-Gravelines proposes that the syntactic structure of *un avion* as opposed to *grand éléphant* is characterized by greater phonological 'cohesion' as proposed by Tranel (1987) and therefore the /n/ was more difficult to detect (though she does not elaborate on what defines 'cohesion').

Nguyen, Wauquier-Gravelines, Lancia and Tuller (2007) further investigated the detection of LCs and ICs by extending previous methodologies to include comparisons with word-final and word-medial consonants. This study also employed a phoneme-detection task and examined the detection of /n/ and /z/ in four positions: W₂-initial (e.g. *des zéros* [de.ze.ʁo] 'some zeros'), W₁-final *enchaînée* (e.g. *seize élèves* [sɛ.ze.lɛv]

‘sixteen students’), word-medial (e.g. *du raisin* [dy.ʁɛ.zɛ̃] ‘some grapes’) and liaison (e.g. *des écrous* [de.zɛ.kʁu] ‘some nuts’). Nguyen et al. replicated the results of Wauquier-Gravelines (1996). Detection rates were lower and correct responses were significantly slower for LCs as compared to the three other conditions. Noteworthy as well is the fact that detection rates for W₁-final and word-medial targets were intermediate to those of IC and LC targets. For the segment /n/, the detection rates of LCs (56%) were lower than both W₁-final *enchaînée* consonants (76%) and medial consonants (75%), which were in turn lower than W₂-initial consonants (87%). The pattern of detection rates for /z/ was in the same direction, though differences were less robust. This study also further extended the results of Wauquier-Gravelines (1996) with respect to the segment /n/. Mean detection rates across all four conditions were significantly lower for /n/ (70%) than for /z/ (92%) and reaction times were longer for /n/ (average 1024 ms across conditions) than for /z/ (average 726 ms across conditions). These results suggest that LCs, though they undergo *enchaînement* as do final fixed consonants, retain a distinctive representational status from that of both fixed final and medial consonants.

Nguyen et al. (2007) note that differences in perception could be attributable to the lexical status of the word in which these segments appear. LCs tend to appear in short function words (e.g. *un* ‘a’ or *des* ‘the/some’), while ICs appear almost exclusively in content words (e.g. *navire* ‘ship’ or *zéros* ‘zeros’). There is a phenomenon in reading known as the *missing letter effect* by which letters are harder to detect in function words than in content words. For example, the letter *t* is harder to detect in *the* than in *weather* (Healy, 1976). Nguyen et al. propose that there could be an auditory equivalent of this effect, which could account for differences in the detection of LCs relative to ICs. (However, this effect does not explain the difference between /n/ and /z/ given that both segments appear in function words in environments of liaison.) The missing letter effect could also account for the difference in detection rates between /n/ and /t/ observed in Wauquier-Gravelines’ (1996) study. The liaison /t/ in this study appeared in the adjective *grand* ‘big’, a content word, while the liaison /n/ appeared in *un*, ‘a’, a function word, which may have contributed to the lower detection rate.

2.3 Theoretical Accounts of Liaison

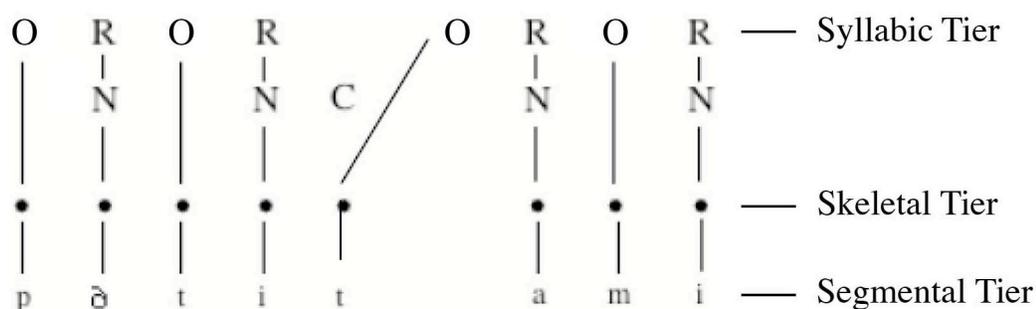
The works cited above present an apparent conflict. The results of Gaskell et al. (2002) and Spinelli et al. (2003) suggest that the recognition of words in liaison environments is not impeded. However, the findings of Wauquier-Gravelines (1996) and Nguyen et al. (2007) have shown that the LCs themselves are not easily accessed. Furthermore, Dejean de la Bâtie (1995) showed that the presence of a *potential* liaison can retard recognition. How then are LCs encoded in a listener’s phonological grammar?

A comprehensive treatment of current theories on the phonological status of liaison as well as the lexical status of LCs is beyond the scope of the present project (see Tranel 1995, Côté 2005 for in-depth reviews), however, we will briefly explore two theories, one phonological and one lexical, that offer differing accounts for the data discussed thus far. As we will see, however, no one theory has yet emerged that can

account for all of the behavioral evidence concerning the perception of liaison discussed above.

Within an autosegmental framework, where phonological representations rely on linking and delinking to segmental slots, LCs are treated as floating segments at both segmental and syllabic tiers (Encrevé 1988; Encrevé & Scheer 2005). LCs are thus thought to stand in contrast to fixed segments, which are assumed to be anchored in a skeletal slot. When realized, LCs are associated to an empty onset position at both tiers, which takes place only under certain conditions. A schematic representation of liaison within an autosegmental framework is shown below in Figure 2-1.

Figure 2-1: *An autosegmental account of liaison showing the phrase petit ami 'boyfriend'. (Adapted from Wauquier-Gravelines 2006.)*



Wauquier-Gravelines (1999) proposes that more cognitive resources are required to process an underlying, floating LC relative to a fixed onset consonant, which could account for the differences in response latencies and detection rates observed in behavioral tasks. If a phoneme-monitoring task operates at the phonological (pre-lexical) level as Wauquier-Gravelines assumes, then mental representations of LCs could be inaccessible due to the fact that the segment has yet not been associated to a slot. Wauquier-Gravelines therefore suggests that the processing of liaison at the lexical level (for example in word-monitoring and priming tasks) is fundamentally different in that access to vowel-initial lexical units is not impeded because the LC has already been associated at the lexical level as W_1 word-final (as assumed by Encrevé 1988) and resyllabified into the empty onset slot.

Within this same theoretical framework, Nguyen et al. (2007) speculate that if the LC belongs lexically to W_1 and is therefore encoded as a word-final consonant, this also could account for lower detection rates. As discussed above and in the previous chapter, many researchers assume that the word recognition system gives special prominence to word onsets as opposed to offsets. A segmentation and processing strategy for French such as SOSH, which privileges syllable onsets to offsets, could predict this effect.

An autosegmental framework, however, has difficulty accounting for differences in the realization of liaison due to what has been referred to as the degree of “syntactic

cohesion” (Tranel 1987: 171) between two words¹². Though he does not explicitly state what makes a syntactic association ‘tighter’ between one group of contiguous words as opposed to another, Tranel evokes this concept to account for obligatory cases of liaison (e.g. between determiner + noun) and also for the fact that, in optional cases of liaison, LCs are realized more often between (monosyllabic) preposition + noun than after an adjective or adverb. An autosegmental approach, which focuses on the phonological properties of segments and assigns no syntactic role to skeletal slots, offers little explanation for this difference. However, ‘cohesion’ has been shown to play a role in the detection of LCs. Indeed, Nguyen et al. (2007) found that detection rates for LCs were significantly lower when the LC appeared in function words such as determiners or prepositions (49%) than in an adjective or adverb (62%), which these authors propose is a function of the “degree of lexicalization” between the two words (p. 21).

Bybee (2001a/b, 2005) has re-analyzed the notions of “syntactic cohesion” and “degree of lexicalization” as an effect of frequency. She has proposed an exemplar model of word recognition in which both lexical frequency and frequency of co-occurrence play a critical role in mental representations of liaison, suggesting that words that occur together frequently in liaison environments (e.g. *un ami, les amis*) are stored in the mental lexicon as chunks and retrieved accordingly. Bybee has proposed a frequency continuum in which stored representations range from invariable fixed phrases such as *c’est-à-dire* [se.ta.diʁ] ‘that is to say’ at one end of the continuum, to constructions such as [NOUN + z- [vowel]-ADJ]_{Plural} at the other, where template slots are filled with open class lexical units such as *enfants intelligents* [ã.fã.zẽ.te.li.ʒã] ‘intelligent children’ in optional cases of liaison. Citing Agren (1973) she notes that in optional liaison there is a trend toward a loss of liaison in modern French. In support of frequency effects, Bybee observes that optional liaison is better preserved by speakers in more frequent word combinations, while liaison is more likely to not be realized in less frequent word combinations.

A frequency-based recognition system can account for the perceptual differences observed between groups of words with varying “degrees of lexicalization” as found in Nguyen et al. (2007). Words that occur together more frequently, such as [determiner + noun], are placed more toward the fixed-phrase end of the continuum and thus more likely to be analyzed as one chunk. LCs are assumed to be more entrenched in these representations and consequently harder for the perceptual system to detect. However, as Nguyen et al. point out, this entrenchment should extend to all consonants in the sequence, not just the LC. Response latencies in highly lexicalized word sequences should show no difference between the detection of a LC and the detection of a IC, which was not shown to be the case.

2.4 Acoustic-Phonetic Cues to Segmentation in Spoken French

One phonetically-based solution to the perception of liaison is to posit that speakers of French give listeners acoustic cues regarding their intended lexical

¹² Also referred to as an *extrême étroitesse* ‘extreme tightness’ between two words by Delattre (1951:26).

assignment of pivotal consonants. In the previous chapter, we discussed the exploitation of acoustic cues to segmentation in numerous languages; acoustic-phonetic variation at multiple prosodic boundaries in French is attested as well.

Welby (2003) demonstrated that French listeners can exploit rises in fundamental frequency (F_0) as a cue to the onset of content words. Welby manipulated this cue in nonsense phrases and found that listeners interpreted sequences such as [me.la.mɔ̃.din] as a single nonce word *mélamondine* when the F_0 rise began at the first syllable, /me/, but as two words *mes lamondines* ‘my lamondines’ when the rise began at the second syllable, /la/.

Spinelli, Welby, and Schaegis (2007) examined possible acoustic cues to elision in spoken French using globally ambiguous minimal pairs such as *l’affiche* ‘the poster’ and *la fiche* ‘the sheet’, both [la.fiʃ]. In an ABX discrimination task, in which participants hear a series of three stimuli (A-B-X) and are asked to identify X as identical to either A or B, participants were 66.3% correct in distinguishing these minimal pairs. Acoustic analyses of the minimal pairs revealed differences in formant and fundamental frequency values, as well as in segmental and syllabic durations. Correlations between acoustic measurements and response patterns suggested that the identification of the two-word utterances (e.g. *la fiche*) was based on F2 values of the first vowel (/a/). F2 was significantly lower in vowel-initial *affiche* than in the definite article *la*. Participants’ responses also marginally correlated with the duration of /a/, which was longer in the definite article (e.g. /a/ in *la fiche*) than in the elided form (*l’affiche*). Consistent with the findings of Welby (2003), identification of the one-word (elided) items (e.g. *l’affiche*) correlated with F_0 values of /a/, which was higher in the content word than in the determiner. Though participants were significantly above chance in their performance on this task, it is worth noting that the rather low accuracy rate suggests that these acoustic cues may not be reliably exploited in natural speech when other more reliable cues are present in the signal.

Segmental duration has also been shown to be an important acoustic cue to comprehension in spoken French¹³. Duration is an extremely robust correlate of word and phrasal stress in French (Delattre 1951, 1966), which is fixed (i.e. never lexical). Lexical stress contrasts as exist in English (e.g. *record*, [rɪ.'kɔrd], verb, versus *record*, [rɛ.kɔrd], noun) and Spanish (e.g. *tomo* [to.mo] ‘I drink’ versus *tomó* [to'mo] ‘he drank’) are not found in French; stress accent falls consistently on the final syllable of a word in isolation or the final syllable of a phrase, which is lengthened and given special prominence (Dell & Vergtonaud 1984). Banel and Bacri (1994) exploited this durational pattern and found that listeners are sensitive to this cue in the segmentation of connected speech. Given ambiguous phonemic sequences such as [ba.gəʒ], which can be interpreted as a single lexical item, *bagage* ‘luggage’, or as a two-word phrase, *bas gage* ‘low pledge’, listeners

¹³ It should be noted as well that French does have instances of contrastive consonant length, but these are limited to verbal forms with a stem ending in /ʁ/ (e.g. *courir* ‘to run’ and *mourir* ‘to die’). Segmental duration here allows for a phonological opposition between the present conditional tense *il courrait* ‘he would run’ and the imperfect tense *il courait* ‘he was running’.

were more likely to identify that sequence as *bagage* when the second syllable was instrumentally lengthened and as *bas gage* when the first syllable was lengthened. The authors attribute this to the listener's expectation that a word-final syllable will be lengthened relative to a word-internal syllable.

2.4.1 Cues to Resyllabification and Enchaînement

Fougeron, Bagou, Stefanuto and Frauenfelder (2002, 2003) found that speakers of Swiss French consistently mark resyllabified (*enchaînées*) consonants by varying segmental duration. These studies measured durational differences among the segments /p, b, t, d, k, n, f, s, ʃ, l/ in three conditions: as word-initial segments (e.g. /l/ in *cas légal* 'legal case'), as resyllabified segments (e.g. *cale égale* 'equal wedge'), as well as in a final condition, as syllable-initial consonants (e.g. **qualégal*). In this final condition, however, the authors were obliged to use non-words due a lack of real French lexical items meeting the phonemic requirements. Significant durational differences were found among all three conditions. Syllable-initial consonants were significantly shorter than word-initial consonants. Furthermore, resyllabified consonants were significantly shorter than both syllable- and word-initial consonants. These results lead Fougeron et al. to propose that speakers produce consonants according to a *comparaison triangulaire*, a 'triangular comparison' among three consonant positions: word-initial, syllable-initial, and *enchaînée*.¹⁴

Fougeron (2007) showed further evidence that resyllabified consonants are shorter than the same segment as it appears at both word and syllable boundaries. However, this same study failed to replicate the finding of significant differences between word-initial and syllable-initial consonants. All of the above results are consistent with previous work by Fougeron (2001) showing that word-initial and accentual-phrase-initial consonants tend to be longer than the same consonant in syllable-initial position.

The study of consonant clusters at word boundaries in French has also suggested that duration can effect segmentation and word recognition (Dumay, Content & Frauenfelder 1999). This study examined two-word utterances in which two types of clusters, obstruent + liquid (OBLI) and /s/ + obstruent (SOB) were produced as either the onset of W_2 (e.g. /kr/ in *demi-croche* [dəmi#kʁɔʃ] 'half eighth note') or straddling the boundary between the two words (e.g. *magique roche* [maʒik#ʁɔʃ] 'magic rock'). Measurements taken from eight native speakers revealed no reliable durational differences in SOB clusters in relation to word boundary location, but did reveal durational differences for the OBLI clusters, suggesting that variation in segmental duration can depend on the nature of the consonant and/or consonant combination. The OBLI clusters showed significant lengthening of both the pre-boundary vowel and the

¹⁴ This three-tiered durational hierarchy could also account for a portion of the results in Nguyen et al. (2007), where detection rates of word-final and word-medial consonants were intermediate to those of ICs and LCs.

liquid in VC#CV sequences as compared to the V#CCV sequences. No differences, however, were found for the obstruent.

This study also investigated the exploitation of these acoustic differences in speech processing by employing the tokens in a word-spotting task. Participants were asked to detect real CVC words (e.g. *roche*) embedded in VCCVC nonce sequences spliced from the two intended segmentations (e.g. [ik#kɔʃ] and [i#kɔʃ]). Reaction times were measured and participants had significantly more difficulty identifying the CVC target in the misaligned condition ([i#kɔʃ]) than in the aligned condition ([ik#kɔʃ]) for the OBLI clusters. Of particular interest is the fact that embedded CVC words produced as part of the SOB sequences, where acoustic analyses had shown no significant durational variation between the two segmentations, showed no difference in reaction times. Diminished response latencies for misaligned OBLI clusters relative to SOB clusters suggests that listeners are sensitive to durational variation and can use this cue in online segmentation.

Content, Bagou, Frauenfelder and Fougeron (2004) further examined acoustic cues to *enchaînement* using exclusively non-words. Thirty-two three-syllable nonce sequences were created of the form CCVCVCV for this production study. The word boundary was manipulated such that the medial consonant appeared either as word-initial (CCV#CVCV) or as word-final and subsequently resyllabified as word-initial through *enchaînement* (CCVC#VCV). The authors measured segmental duration of consonants and vowels as well as fundamental frequency and spectral characteristics of pre- and post-boundary vowels in utterances produced as both questions and declarative statements. Significant lengthening of the pivotal consonant was found in word-initial position. In addition, consistent with the findings of Dumay, Content and Frauenfelder (1999), pre-boundary vowels were significantly longer in non-final position (i.e. in the CCVC#VCV sequence). Systematic differences in fundamental frequency and formant values were not found. These results confirm those of previous studies by Fougeron and colleagues that the acoustic realization of segments is in French affected by the intended lexical segmentation of the speaker.

2.4.2 Cues to Liaison

In the case of liaison, the majority of the classical literature on the acoustic-phonetics of French has maintained that consonants are identical at the acoustic level whether they appear as LCs or ICs (see for example Encrevé 1988; Grammont 1960; Nyrop 1925; Passy 1917), though as early as 1940, Delattre noted that consonants that surface in liaison are *plus faibles* ‘weaker’ than the same segment in initial position. More current research has validated Delattre’s observation and shown that systematic durational differences between consonants that surface in liaison environments and their lexical-word-initial counterparts are consistently produced by speakers. Dejean de la Bâtie (1993) found that the duration of both the closure and following burst are both shorter for liaison /t/ compared with word-initial /t/. Wauquier-Gravelines (1996) found similar results for /t/, which had an average closure duration of 50 ms in liaison position

and 70 ms in initial position, though she did not find significant durational differences between liaison and word-initial /n/ (58 ms versus 61 ms).

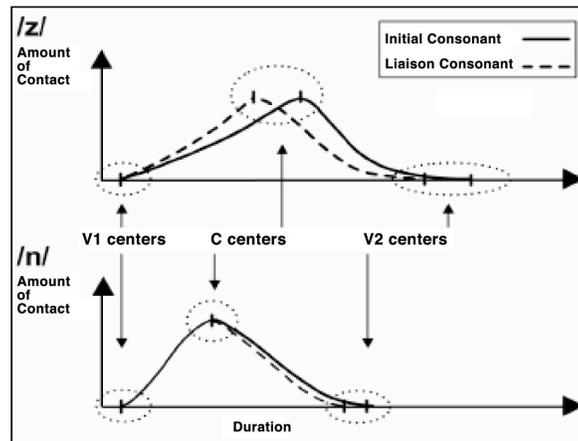
Gaskell et al. (2002) also found durational differences. The segments /t/, /ʁ/ and /z/ were significantly shorter when realized in liaison (mean 73 ms) than in word-initial position (mean 88 ms). Spinelli et al. (2003) found significant durational differences among five consonants that surface in liaison /n, t, ʁ, g, p/. LCs were on average 17% shorter than ICs. Measurements of the pivotal consonants revealed that ICs were on average 10 ms longer (difference range= 6 to 12 ms) than word-final, resyllabified consonants. Similarly, Shoemaker (2006) found that /t/ produced in liaison contexts (*un grand amour* ‘a great love’) were over 20 ms shorter than word-initial /t/ (*un grand tamis* ‘a big sieve’) and that /ʁ/ in liaison context (*le dernier anglais* ‘the last Englishman’) are over 10 ms shorter than word-initial /ʁ/ (*le dernier rancard* ‘the last appointment’). However, Nguyen et al. (2007) who looked at acoustic realizations of /n/ and /z/ in both environments found no significant durational differences between LCs and ICs for either segment.

Even more recently, Douchez and Lancia (2008) analyzed differences in the articulation of /n/ and /z/ when these segments are realized as LCs and ICs. Two native French speakers were fitted with pseudo-palates, a device which measures the amount of contact between the tongue and the palate during speech production. Durations in this study were measured in two intervals based on the amount of linguopalatal contact. The first interval was measured from the middle of the vowel preceding the pivotal consonant to the peak of consonantal contact as measured by the pseudo-palate, while the second interval was measured from this peak to the middle of the following vowel.

Results from the segment /z/ showed that both intervals were significantly shorter when /z/ appeared in liaison environments than when it appeared as word-initial, while the segment /n/ showed a significant difference in the same direction, but for the second interval only. A schematic representation of these authors’ results is shown below in Figure 2-2.

Voice onset time (VOT) of /t/ has also been shown to vary according to the presence or absence of liaison. Initial /t/ was found to have a longer VOT than a /t/ that surfaces in liaison (Dejean de la Bâtie 1993; Wauquier-Gravelines 1996). Additionally, the burst energy of an IC has been found to be greater than that of a LC (Durand 1953; Dejean de la Bâtie 1993).

Figure 2-2: Schematic representation of the articulation of /z/ and /n/ in initial and liaison position as a function of amount of linguopalatal contact. The X axis represents time, while the Y axis represents the amount of contact. (Adapted from Douchez & Lancia 2008).



As in environments of resyllabification, liaison has also been found to affect segments surrounding the pivotal consonants. The production sample in Spinelli et al. (2003) showed that vowels preceding LCs were shorter by 3% than vowels preceding ICs, though no differences were found for post-boundary vowels. Nguyen et al. (2007) measured pre-boundary vowels as well, but found no significant differences for liaison environments.

2.4.3 The Exploitation of Segmental Duration in the Perception of Liaison

Several of the behavioral studies reviewed above allude to segmental duration as a cue to segmentation in the case of liaison. Spinelli et al. (2003) hypothesized that listeners exploit this “subtle but reliable” acoustic cue in French to mark word boundaries and that durational variation facilitates access to representations in the mental lexicon (p. 248). Spinelli et al. suggested that these differences are robust enough to “bias interpretation in the correct direction” (p. 250) in cases of ambiguity, however this suggestion remains conjectural as this study did not directly demonstrate that duration was guiding participants’ responses. Moreover, a post-hoc correlation of consonant durations and priming results for individual stimuli in the Spinelli et al. study did not in fact show a significant correlation between the two. In other words, the length of the consonants did not reliably predict the direction of the priming effects (McQueen, personal communication).

Using the same recordings of 12 pairs of globally ambiguous phrases used in the Spinelli et al. (2003) study, Shoemaker and Birdsong (2008; see also Shoemaker 2005) more directly tested the perception of liaison by employing a forced-choice identification task in which 15 native speakers of French and 15 late learners of L2 French were asked to differentiate ambiguous phonemic content. Participants heard an ambiguous phrase

(e.g. [il.na.o.kœ.nɛʁ] produced as either *Il n'a aucun air* or *Il n'a aucun nerf*) and at the offset of the phrase were presented with both possible parses and asked to indicate which they had heard.

Participants in both groups performed roughly at chance (native speaker mean accuracy, 53.2 %; non-native speaker mean accuracy, 52.7%) and did not consistently identify ambiguous phrases suggesting that, though durational differences may be systematically present in the acoustic signal and may allow for the activation of vowel-initial candidates in the word recognition process, these differences are not robust enough to systematically guide listeners in the disambiguation of globally ambiguous input.

In an attempt to investigate Spinelli et al.'s (2003) claim that duration represents a cue to disambiguation, Shoemaker and Birdsong also looked for correlations between mean accuracy rates and the duration of the pivotal consonants for each individual stimulus pair. If duration does indeed signal to the listener the presence or absence of liaison, longer consonants should be identified more as ICs and conversely, shorter consonants should be identified more as LCs. However, Shoemaker and Birdsong found no significant correlations for any of the stimulus pairs.

Interestingly, though neither participant group performed significantly above chance on the identification task, the two groups did perform similarly as to the distribution of their responses. Though overall mean accuracy rates for both groups were roughly at chance, there were significant biases within most stimulus pairs for both groups. Moreover, the non-native speakers showed biases in the same direction as those of the native speaker group in 10 of 12 minimal pairs. In several post-hoc analyses, lexical frequency and phrase plausibility were ruled out as factors guiding responses. This suggests that there may be other as yet unidentified factors (bottom-up or top-down) that were guiding participants' responses and that both native and non-native speakers were sensitive these factors¹⁵.

2.5 Summary

Taken as a whole, the body of research concerning SWR in French and, specifically, the perception of liaison seems to paint a confusing picture. Though speech recognition is not slowed by resyllabification due to either *enchaînement* or liaison (Gaskell et al. 2002; Spinelli et al. 2003), LCs seem to be assigned a distinctive cognitive status in relation to fixed initial consonants as evidenced by the results of word- and phoneme-detection studies discussed above (Dejean de la Bâtie and Bradley 1995; Nguyen et al. 2007; Wauquier-Gravelines 1996). Moreover, while acoustic differences have been demonstrated between liaison and resyllabified consonants and consonants that occur in other environments, no research to date has directly established that listeners make use of these differences in the online processing of speech. The use of segmental duration as a cue to word boundaries has been demonstrated in English (Davis et al. 2002; Salverda et al. 2003), Dutch (Quené 1992; Shatzman & McQueen 2006; Warner et al. 2004) and Italian (Tabossi et al. 2000), however, whether listeners exploit segmental duration in the processing of liaison in French warrants further investigation.

¹⁵ Results for non-native speakers versus native speakers are discussed in more detail in Chapter Four.

CHAPTER THREE: Processing Speech in a Second Language

*No speaker of any language perceives acoustic reality;
perception is altered in the service of language.*
(Kuhl 2001: 11852)

3.0 Introduction

Thus far we have discussed the cognitive processes involved in SWR from the point of view of the native speaker. Chapters One and Two reviewed a large body of research establishing that acoustic and phonological cues to speech segmentation and word recognition are not exploited to the same extent and in the same manner cross-linguistically (e.g. Cutler & Norris 1988; Cutler et al. 1989; Pallier et al. 1993; Sebastian-Galles et al. 1992; Tabossi et al. 2000). The question therefore arises as to how non-native speakers process a L2 that may not make use of the same recognition and segmentation strategies as the L1. In addressing this question, the current chapter focuses on the perception and acquisition of L2 phonological systems and L2 segmentation strategies. As we will see, the very processing strategies that render so efficient the comprehension of our native language can paradoxically hinder the aural comprehension of a L2 acquired later in life. Before exploring L2 processing, we will briefly discuss the acquisition of L1 phonology and segmentation routines in order to draw comparisons between L1 and L2 phonological development.

3.1 Age Effects in Phonological Processing

Infants are born with acute sensitivity to speech sounds. Drawing on a line of research suggesting that infants are extremely adept at discriminating speech sounds in the early months of life (e.g. Eimas, Siqueland, Jusczyk & Vogorito 1971; see also Jusczyk 1997 for a comprehensive review of this domain), many researchers originally assumed that infants process speech via perceptual mechanisms that are both innate and specifically attuned to human language. Infants of six months are able to distinguish

contrasts from languages to which they have never been exposed in addition to contrasts from the ambient language. Infants born to English-speaking parents, for example, have shown evidence of the discrimination of dental /t/ versus retroflex /t/, a phonemic contrast that exists in Hindi but not in English (Werker & Tees 1984; see also Werker & Tees 1999). Similarly, English-learning infants demonstrated discrimination of contrasts specific to Thai (Aslin, Pisoni, Hennessy & Perey 1981) and Czech (Trehub 1976). Eimas (1975) among others suggested that infants are born endowed with innate 'phonetic feature detectors' endowing them with sensitivity to all possible speech sounds.

By the end of the first year of life, however, the ability to perceive segmental contrasts diminishes as the infant perceptual system abandons or prunes away those phonetic features which are not contrastive in the native language. Early accounts of phonological development were consequently based on the idea that phonetic learning in infants is characterized by the preservation and/or loss of sounds from among all possible speech sounds. For example, Werker and Tees (1984) showed that English-learning infants lost the ability to discriminate the Hindi /t/ contrast referred to above by the age of 12 months. Likewise, infants of six months born into Japanese-speaking homes discriminated /r/ and /l/, a contrast not discriminated by Japanese-speaking adults, but at one year Japanese infants were no longer able to make the same /r-l/ distinction (Tsushima, Takizawa, Sasaki, Siraki, Nishi, Kohno, Menyuk & Best 1994). Recent data also suggest that, not only does sensitivity to non-native contrasts diminish at this point in development, but the ability to discriminate native categorical contrasts is sharpened, indicating further refinement of the L1 phonology (Kuhl, Stevens, Hayashi, Deguchi, Kiritani & Iverson 2006; Polka, Colantino & Sundara 2001; Rivera-Gaxiola, Silva-Pereyra & Kuhl 2005).

At the suprasegmental level a similar sharpening of perception occurs. Research has suggested that infants develop sensitivity to rhythmic structure before acquiring segmental knowledge, possibly even in the womb. Initially, newborns can discriminate languages that differ in basic rhythmic structure. For example, discrimination has been demonstrated between a stress-based language such as English and a mora-based language such as Japanese (Nazzi et al. 1998; see also Nazzi & Ramus 2003). However, infants cannot discriminate languages that exhibit similar metrical structures, for example, between French and Spanish, which are both syllable-based (Mehler et al. 1988) or English and Dutch, which are both stress-based (Nazzi et al. 1998). It has been suggested that a preference for the rhythmic structure of the mother tongue emerges between six and nine months. A head-turning task showed that nine-month-old infants learning English preferred to listen to disyllabic English words with stress on the first syllable, whereas six-month-old infants did not (Jusczyk, Cutler & Redanz 1993), suggesting that infants' perception was sensitized to the regular distribution of English stress.

Once an infant has acquired sensitivity to phonemic contrasts and rhythmic structure, the next challenge is to extract individual words from the speech stream. Corpus research has demonstrated that less than 10% of infant-directed speech in the first year of life consists of isolated words (Siskind 1996; Van de Weijer 1998), therefore infants must develop strategies early on to detect word boundaries. Not coincidentally, as

pointed out by Cutler (2001), the pruning of the phonological perception of segmental contrasts that occurs around one year corresponds to the learning of a child's first words. The developing infant has thus learned to disregard sound variation that does not produce a change in meaning, focusing only on contrasts which are linguistically relevant in the mother tongue.

3.1.1 Becoming 'Adult Listeners'

As the infant perceptual system becomes attuned to the phonological properties of the L1, infants become in effect 'adult listeners' as sensitivity to abandoned contrasts is greatly diminished. This decline in perceptual sensitivity is one factor that has been invoked to account for the striking difference between the facility of L1 acquisition and the difficulties that adult learners experience in acquiring a L2. The domain of L2 acquisition has long been centered on the effects of maturation on L2 attainment as predicted by the Critical Period Hypothesis (CPH; see for example Lenneberg 1967). The CPH holds that nativelike attainment of language is out of reach if learning begins after the closure of a bounded and biologically determined period of heightened sensitivity to language acquisition, usually assumed to coincide with puberty. For proponents of the CPH, language acquisition initiated prior to the closure of this period is assumed to result in nativelike linguistic competence. However, if acquisition begins after closure, deficits are assumed unavoidable, and, according to Lenneberg (1967), acquisition has recourse only to general purpose learning mechanisms as opposed to specialized neural circuitry.

There are differing opinions as to what brings about the putative closure of the critical period (CP). Candidates include a progressive lateralization of cerebral function and a subsequent loss of plasticity (Lenneberg 1967); an irrevocable loss of access to Universal Grammar (Bley-Vroman 1989); an increase in the sophistication of cognitive function that is detrimental to language learning (Newport 1990); the dismantling of neural circuitry needed for language learning in order to cut metabolic costs (Pinker 1994); the atrophy of the language acquisition facility due to lack of use (Bever 1981); or the increasing strength of established L1 neural connections which inhibits the laying down of novel neural pathways in the L2 (McClelland 1996; Elman et al. 1996). See Birdsong (1999, in press) as well as Bowden, Sanz and Stafford (2005) for comprehensive discussions of mechanisms that have been proposed to account for critical period effects.

A stringent interpretation of the CPH predicts a sharp cut-off in levels of L2 attainment at or around puberty (see for example Patkowski 1990). However, numerous studies have demonstrated that this prediction does not accurately reflect observed patterns of learning outcomes (see Birdsong 2006 for a review). Attainment levels across linguistic domains are better represented by a gradual post-maturational linear decline in sensitivity than by an abrupt drop after the assumed closure of the CP. In addition, the slope of this decline can vary as a function of myriad factors that are wholly independent of age. These factors include, but are not limited to, the L1/L2 pairing, the amount of L2 use in relation to L1 use, amount of education in the L2, and language dominance (see for example Birdsong 1998; Bongaerts 1999; Flege 1999; Flege & Liu 2001; Golato 2002 and Piske, MacKay & Flege 2001 for discussion of these and other factors). More recent

refinements of the CPH have included the notion of a ‘sensitive’ or a ‘maturational’ period (Hyltenstam & Abrahamsson 2003), or even an ‘optimal’ period (Werker & Tees 2005), implying that during this period the acquisitional system experiences heightened sensitivity to language input, but that this period is not necessarily essential to nativelike attainment.

Indeed, research over the past two decades has offered strong evidence against strict limitations on learning implied by the CPH by uncovering instances of nativelike behavior (i.e. linguistic performance that is indistinguishable from native controls) in post-pubescent learners. In a review of previous work concerning the effects of age on L2 acquisition and processing, Birdsong (2006) reports rates of nativelike attainment ranging from 0% to 45% of participants in 20 studies testing participants across several linguistic domains. However, while instances of nativelike behavior in morphosyntax hover around 10-15% of participants (see Birdsong 1999 for a review of relevant studies), examples of nativelike pronunciation are less frequent when L2 exposure begins in adulthood. The domain of phonology is thought to be particularly sensitive to the effects of age (Long 1990; Scovel 1988). Some researchers have even suggested that there is a domain-specific CP for phonology that terminates earlier than CPs for other linguistic domains (Moyer 1999; Singleton & Ryan 2004; see also Pallier, Bosch & Sebastian-Galles 1997). Indeed, in a study of 240 native Korean speakers ranging in age of arrival (AOA) in the United States from 1-23 years, Flege, Yeni-Komshian and Liu (1999) observed that degree of foreign accent was significantly more affected by AOA than performance on behavioral tasks involving grammaticality judgments. As AOA increased, foreign accents steadily grew stronger, while performance involving morphosyntactic knowledge correlated more with education level and amount of L2 use than with age.

However, impressive rates of nativelike pronunciation are attested. Recent work has demonstrated that a nativelike accent, that is, L2 speech that is judged indiscernible from that of native speakers, is attainable for more than just a handful of late learners. Bongaerts, van Summeren, Planken & Schils (1997) found three out of nine Dutch learners of L2 French who showed no detectable foreign accent (see also Bongaerts 1999). Bongaerts, Mennen and van der Slik (2000) found nativelike performance for learners of L2 Dutch. Of 30 participants from various L1 backgrounds, two were judged as native speakers by both naïve and expert judges. Birdsong (2003; see also Birdsong 2007) found two out of 22 advanced English-speaking learners of French were deemed indistinguishable from native controls. Of particular note is the fact that these nativelike attainers had biographical factors in common — all had undertaken phonetic or diction training and all had expressed high motivation to integrate into the L2 culture and pass for native speakers, further demonstrating that age is not the sole determinant of degree of foreign accent.

3.2 Modeling the Acquisition and Perception of Second Language Phonology

While early study of the acquisition of L2 phonological systems approached the process almost exclusively from the point of view of *production*, attributing a foreign accent in late learners to *contrast* between the L1 and the L2 (see for example Lado’s *Contrastive Analysis Hypothesis* 1957; Eckman’s *Markedness Differential Hypothesis*

1977, 1981; and Major's *Ontogeny Model* 1994), we limit ourselves here to a discussion of more recent models which focus on *L2 perception*.

The *perception* of a non-native phonological system has long been focused on the notion of *interference*. As early as the 1930s, the idea that the sounds of our native language can affect how we perceive a L2 was prevalent. Both Polivanov (1931) and Troubetzkoy (1949) suggested that the L1 functions as a sort of filter for the L2, causing listeners to map what they hear in the L2 onto established L1 representations. Troubetzkoy (1949) noted that a listener "...emploie involontairement pour l'analyse de ce qu'il entend le « crible phonologique » de sa langue maternelle." (...involuntarily uses the "phonological sieve" of his native tongue to analyze what he hears; p. 54). Researchers have more recently revised the notion of a L1 filter, referring instead to a sort of 'deafness' toward non-native phonological contrasts which arises when two L2 sounds are mapped on the same L1 category (Dupoux, Pallier, Sebastian & Mehler 1997; Dupoux & Peperkamp 2002). Numerous behavioral studies have indeed demonstrated the influence of the native phonological system on perception of the L2 (see for example Cutler 2001, 2002; Floccia & Bertoncini 1993; McAllister, Flege & Piske 2002; Hallé, Best & Levitt 1999; Strange 1995; Strange et al. 1998; Pallier, Christophe & Mehler 1997; Weber & Cutler 2004 among others).

We review below two models which seek to describe the processes by which the L1 'filter' affects the perception of L2 phonology: the *Perceptual Assimilation Model* (PAM; Best 1995; Best et al. 2001) and the *Speech Learning Model* (SLM; Flege 1995, 1997). Significantly, what PAM and SLM share is the view that the accuracy with which non-native sounds are perceived is a direct function of their perceived phonological similarity to established L1 categories.

3.2.1 The Perceptual Assimilation Model (Best 1995)

Drawing on the fact that the world's languages compose their sound systems from a universally shared set of possible articulatory gestures, Best (1995; Best, McRoberts & Goodell 2001) observes that there is inevitably a great deal of overlap in the production of segments across languages. The *Perceptual Assimilation Model* is centered on the notion that the degree of similarity between the L1 and L2 sound systems determines the degree of difficulty that L2 learners will experience in L2 perception. In addition, PAM focuses on the perception of non-native *segmental contrasts*, offering predictions as to where successful discrimination of these contrasts is expected to be achieved. Essential to this idea is the prediction that learners will have problems detecting differences between native and non-native segments if non-native segments are perceived as similar to pre-existing native categories. In these cases, the learner will assimilate the new L2 sound into what is perceived as the closest L1 category instead of creating a novel and distinct category.

Best makes a distinction among four types of assimilation of L2 contrasts. First, *Two-Category Assimilation* describes cases in which a non-native contrast is perceived as similar to two separate native categories. This contrast is thus assimilated into two different native categories, with one L2 sound being assimilated to one native category and the other being assimilated to a different native category, neither of which is a good

fit. An example of this type of assimilation occurs in speakers of French learning the English dental fricative voicing contrast in /θ/ and /ð/. The unvoiced member of this pair is perceived as /s/ while the voiced member is perceived as /z/, both of which are existing members of the French phoneme inventory. As a result, neither of the two non-native sounds is correctly classified, but discrimination is nevertheless achieved.

The second type of assimilation, *Single-Category Assimilation*, refers to cases in which two non-native categories are assimilated into a single existing native category. In this case as well, the native category into which the two non-native sounds are assimilated constitutes a good match for neither. Discrimination of the two sounds is assumed to be difficult as both tend to be (poorly) mapped onto one existing category. Japanese learners of English, for example, have well documented difficulty with the English /l-r/ distinction as both tend to be mis-mapped onto the Japanese phoneme /r/.

The third type of assimilation described by Best is based on *Category Goodness*, in which two non-native categories are also assimilated into one native category, but unlike *Single-Category* assimilation, one of the non-native sounds represents a good exemplar of a pre-existing native category. For example, English-speaking learners of French often assimilate both French vowels /y/ and /u/ into the pre-existing English category /u/, preserving the roundedness of /y/ while losing its frontedness. Portuguese-speaking learners of French on the other hand assimilate French /y/ and /i/ into the pre-existing Portuguese category /i/, preserving the frontedness of /y/ while losing its roundedness. In this type of assimilation, discrimination of the two non-native categories is not achieved.

Finally, Best describes non-native speech sounds which may fall outside the phonetic space of the native phonology. These sounds are not assimilated as the non-native contrasts are sufficiently different from native categories to be perceived as such and the learner establishes new categories for each. Discrimination of these sounds is assumed to be strong.

PAM successfully proposes an explanation as to why certain non-native sounds are perceived with ease, while others are persistently assimilated (badly) to L1 categories. This model accurately predicts that Zulu clicks, for example, will be easily perceived by speakers of European languages (Best, McRoberts & Sithole 1988) as there is no close equivalent. Vowel contrasts among languages on the other hand are notoriously difficult for L2 learners (see for example Flege 1997; 2003; Flege et al. 1999; Polka 1995). A discrete category is established with relative ease for a novel sound in the case of the former, while vowels entail a great deal of cross-linguistic overlap within production parameters and are therefore more likely to be incorrectly assimilated into pre-existing native categories.

One shortcoming of PAM is that it is centered primarily on early contact with and categorization of L2 sound systems. It is a descriptive model that focuses on predicting discrimination difficulties by elucidating how naïve (i.e. inexperienced) listeners may perceive unfamiliar speech sounds; it does not offer developmental predictions for the subsequent learning that inevitably occurs as the learner is increasingly exposed to L2 input. Best (1995) herself admits that the PAM should be expanded in respect to stages of

the development of the L2 phonology, including tracking the particular development of different types of assimilation (p. 198).

3.2.2 The Speech Learning Model (Flege 1995)

The *Speech Learning Model* (SLM) put forth by Flege (e.g. 1987; 1995; 2002) is a more dynamic model in that, unlike PAM, it specifically addresses learning processes and the development of the L2 phonological system in experienced L2 listeners. The SLM offers predictions on long-term phonetic development. Like Best, Flege offers multiple classifications of L2 sounds as they relate to those of the L1. He proposes that the learner classifies an L2 sound as *identical to*, *different from* or *similar to* the sounds of the L1. Those sounds which are classified as identical are thought to pose no problem for the learner as they are subsumed into already established phonological categories. L2 sounds which are sufficiently dissimilar from those of the L1 are also thought to be learned with relative ease as novel discrete categories can be established. Sounds which are perceived as similar to L1 sounds, however, pose the biggest obstacle for the learner as these phonemes are perceived as belonging to neither new nor separate phonological categories and thus are incorrectly assimilated into an existing (but different) category. Flege proposes that in these cases novel L2 phonetic category formation is in effect blocked by mismatch of phonetic features between the L1 and L2. Consequently, features that signal a contrast in the L2, but not the L1, will be difficult to perceive. This assimilation then hinders the creation of a new category throughout learning, and, according to Flege, sounds classified as *similar* are mostly likely to retain an accent.

Unlike PAM, the SLM directly addresses the effects of age on phonological acquisition, predicting that learners are less likely to establish new L2 phonetic categories as the AOA in the L2 environment increases. However, Flege does not attribute this decline to a loss of plasticity resulting in the inability to perceive and pronounce unfamiliar sounds. Instead, he makes the case that difficulty to acquire L2 sounds is correlated with the relative degree of *entrenchment* of the L1 phonological system when L2 learning begins. Flege qualifies entrenchment as the degree to which L1 categories have formed and stabilized, i.e. the older the learner is, the more entrenched the L1 phonology is, and, consequently, the more difficult the creation of new categories becomes. Though Flege does not explicitly make reference to a neurological basis for L1 entrenchment, this notion is consistent with connectionist models (e.g., Elman et al. 1996; Marchman 1993) that have suggested that, as L1 neural representations become increasingly entrenched, the re-wiring or “unlearning” that would be required for the acquisition of L2 phonetic categories becomes increasingly difficult.

Crucially, while Flege (1999) maintains that the likelihood that L2 learners will establish new categories for L2 vowels and consonants decreases as age of exposure to the L2 increases, he maintains that “the mechanisms and processes used in learning the L1 sound system remain intact over the life span” (Flege 1995: 239) and plasticity is never entirely lost. One precept linked to this view is Flege’s observation that the L2 can interfere with L1 performance just as L1 interferences occurs in the L2. Cross-linguistic effects in bilinguals have indeed been demonstrated. Flege (1987) showed that VOT values in both the L1 and L2 of French/English bilinguals were intermediate between

those of monolingual speakers of these languages, suggesting that L1 boundaries had shifted toward L2 values. Further evidence of L2 influence on the L1 comes from Cook (2003) who found L2 interference in the domains of pronunciation, morphosyntax, and collocations.

Both PAM and SLM offer cogent models of the perception of L2 phones, including specific predictions as to mechanisms concerning the perception and classification of sounds along a continuum of similarity to the L1. One limitation that both models share, however, is that the perception of segments is removed from the context of natural speech. Focus is placed on perception of phonetic contrasts at the segmental level only and therefore no predictions are made as to how levels both below the phoneme (e.g. sub-phonemic allophonic detail) and above the phoneme (e.g. prosody or phenomena of external sandhi such as liaison) are treated by the L2 perceptual system.

3.2.3 Re-modeling Phonological Acquisition

More recent work on phonological acquisition in both infants and adults focuses on general perceptual learning rather than language-specific processes. Stemming partially from evidence that phenomena once assumed to be language-specific (e.g. categorical perception) are neither limited to speech (Miller et al. 1976; Pisoni 1977; Jusczyk et al. 1977) nor to humans (Kuhl & Miller 1975, 1978), some recent models of the acquisition of phonetic systems are based on general statistical learning. According to these models, phonological learning in infants is shaped by statistical regularities in the acoustic input to which they are exposed, suggesting that the phonological system in effect organizes itself according to the distributional properties of the phonological environment (see for example Maye, Weiss & Aslin 2008; Saffran, Aslin & Newport 1996). Acquisition is thought to occur via the computation and accumulation of distributional frequencies with which items/events occur in relation to other items/events. Maye, Werker & Gerken (2002) for example showed that infants exposed to different distributions of sounds along the same eight-step continuum of /t/-type sounds categorized the input in different ways. Half of the infants in this study were exposed to more instances of two sounds toward each end of the continuum (Stimuli 2 and 7 along the eight-step continuum), simulating a bimodal distribution, while the other half were exposed to more sounds in the middle of the continuum (Stimuli 4 and 5), simulating a unimodal distribution. After less than three minutes of exposure, the infants exposed to the bimodal distribution discriminated the endpoints of the continuum better than infants in the unimodal condition, suggesting that exposure to a bimodal distribution had provoked a sort of categorical distinction between the two sounds. The authors maintain that this experiment is analogous to the difference between, for example, a child who acquires the Hindi dental /d/ and retroflex /d/ distinction and a child growing up in an English-speaking household who acquires a single category, /d/. In the case of the former, the child is exposed to a clearly delineated bimodal distribution of the two sounds, while in the case of latter, the child may be exposed to /d/s which have many variants, but which are all grouped around one central tendency. Infants have also shown particular sensitivity to the distributional properties of probabilistic phonotactics and lexical stress which may render the identification of word boundaries easier (Jusczyk 1993, 1997). The

ability to exploit the distribution of phonotactic patterns in the detection of familiarized words in running speech, for example, has been demonstrated in eight-month-olds (Aslin, Saffran & Newport 1998).

While early theories of infant phonological development were centered on the idea of a *loss* of perceptual sensitivity to sounds not found in the ambient language, a model of infant learning based on the statistical properties of speech is more centered on the idea of the *reorganization* of phonetic space (Werker 1995; Werker & Tees 2005). Kuhl (2000, 2005) has proposed a view of phonological acquisition that is centered on the effects of early neural commitment based on these regular distributions, proposing that exposure to acoustic input brings about dedicated neural networks in infants' brains that encode the patterns of the native language. This view differs from a traditional notion of a critical period in which acquisition slows as a function of chronological age in that learning in Kuhl's model is independent of time, based instead on continuity of input. Learning ends when the influx of significantly novel input ends and a statistically determined distribution is established. Put differently, phonological acquisition slows only after there has been enough acoustic input to establish a distribution that can reliably and systematically predict the categorization of further input. This notion is supported by evidence suggesting that biologically determined periods of perceptual development in other domains, namely vision, can be prolonged under certain exposure conditions (Cynader, Timney & Mitchell 1980).

Interference in language learning later in life is attributable to this initial mapping of the native phonology rather than by the effects of maturation or the closure of a sensitive acquisitional period. The decline in sensitivity observed in infant perception results therefore from the stabilization of phonetic distributions and not from age-related effects. This model also goes against the notion of the "suppression" of a new (L2) phonological system and instead offers that the acquisition of a new phonological system may be hindered by the relative entrenchment of the native system. In privileging experience over a finite window of time, this viewpoint is similar to that of Flege, who also suggested that age effects on phonological acquisition are a function of the degree of L1 entrenchment rather than chronological age. However, Kuhl (2000) extends Flege's model in that she offers specific predictions based on the behavior of neural networks, proposing that committed neural networks interfere with the processing of a second language when L2 patterns do not conform to already established patterns. In addition, in emphasizing general learning strategies in the acquisition of language, broader predictions can be inferred about multiple levels of phonological organization than those offered by SLM and PAM.

3.2.3.1 Statistical Learning in Adults

A burgeoning body of research on perceptual processing in adults has offered evidence that adult phonological systems retain sensitivity to statistical distributions in the speech signal throughout the lifespan (see for example Eisner & McQueen 2005; Kraljic & Samuel 2005; Maye et al. 2003; Vroomen et al. 2007 among others). In laboratory environments, adults have shown sensitivity to distributional patterns of prosody and transitional probabilities between syllables in the learning of artificial

language (Saffran, Newport & Aslin 1996). Adult listeners have also demonstrated plasticity in adapting to synthetic speech (Greenspan, Nusbaum & Pisoni 1988) and artificially time-compressed speech (Dupoux & Green 1997) after just minutes of exposure. Moreover, learning in both of these contexts transferred to novel words and speakers.

The plasticity of the adult perceptual system is evidenced by the ease with which listeners are able to normalize variability in their L1 due to different speakers (native or non-native), unfamiliar dialects, change in speaking rate, etc. Norris et al. (2003) designed an experiment meant to be analogous to a listener's adaptation to a new (L1) speaker or dialect. They showed that even extremely short exposure to ambiguous sounds can cause a native phoneme boundary shift when ambiguity is coupled with lexical information. The authors used a lexical decision task to investigate whether native categorical boundaries for the fricatives /f/ and /s/ could shift after exposure to ambiguous sounds falling somewhere along a continuum between the two sounds. They began by establishing participants' categorical boundaries for /f/ and /s/ using a phoneme identification task along the /f-s/ continuum. Then, in a lexical decision task, participants heard real words ending with either ambiguous /f/ or ambiguous /s/. One half of the participants heard words ending in ambiguous /s/ but also words ending in unambiguous /f/; the other half heard words ending in ambiguous /f/ but also words ending in unambiguous /s/. They then performed a second phoneme identification task with the same continuum of /f/ to /s/. Listeners who had heard the ambiguous /f/ and unambiguous /s/ were more likely to categorize a sound as /f/ and vice versa — listeners who had heard the ambiguous /s/ and unambiguous /f/ were more likely to categorize a sound as /s/. This showed that /f/ and /s/ categories had been broadened by very limited exposure (20 ambiguous words among 100 real words and 100 non-words). Participants had succeeded in identifying the ambiguous sounds according to lexical information and had changed phonemic categories accordingly.

More recently, Maye, Aslin and Tanenhaus (2008) showed that speakers can adapt to shifted vowels. Participants in this study first listened to a 20-minute sample of speech in a normal (American) English accent, after which they completed a lexical decision task that contained lexical items included in the story, lexical items not included in the story, as well as non-words. In a second session one to three days later, the same participants again heard the same story produced in a voice in which certain vowels had been lowered (e.g. *witch* /witʃ/ produced as [wɛtʃ]). In a second lexical decision task, participants accepted items with lowered vowels as real words that had been rejected in the first lexical decision task (e.g. [wɛtʃ]). Moreover, increased acceptance rates transferred to items that had not been included in the speech sample, but that exhibited a similar vowel shift from a real English word (e.g. [kɛŋ] accepted as a real word, presumably from *king* /kɪŋ/), suggesting that participants had not simply memorized the pronunciation of words in the story, but that the actual vowel space had shifted.

Similarly, Clarke and Garrett (2004) found in a test of adaptation to non-native speech that native English speakers can adapt to non-native productions of English sentences after exposure to just two to four sentence-length utterances, further showing

evidence of rapid perceptual learning of phonetic categories and the readjustment of established native contrasts by adult speakers. The above studies offer strong evidence for adult perceptual learning and attest to the plasticity of the adult phonological system. The study of statistical learning in adults sheds light on possibilities for L2 acquisition and phonetic training, which we discuss in further detail below as well as in Chapter Nine.

3.3 Spoken Word Recognition in a Second Language

We return now to a discussion of the word recognition and segmentation strategies elaborated upon in Chapters One and Two (e.g. competition-based segmentation, probabilistic phonotactics, prosodic structure, and allophonic variation). In this section we will re-examine implementation of these strategies as they apply to the processing of a L2. Cutler (2001), in a review of previous research on ‘non-native listening’ processes, notes three areas of speech processing which can constrain the L2 listener. First, as discussed in detail in Chapter One, segmentation strategies are language-specific. This presents problems for the listener in that L1 strategies may not be efficient when applied to the processing of the L2. Second, Cutler (2001) points out that segmentation strategies are located in the listener, not in the signal. In other words, it is not the inherent nature of the acoustic signal in a particular language that triggers the use of a particular segmentation routine, but rather the language experience of the listener. For this reason, according to Cutler (2001), not only do adult learners lack recourse to an appropriate L2 strategy, but inexperienced learners may employ L1 segmentation routines in the comprehension of the L2 “even when the speech signal discourages it” (p. 9). Finally, though strategies are language-specific, Cutler (2001) does propose that the inappropriate use of a particular strategy is avoidable with sufficient experience, pointing out that some bilinguals are able to inhibit the implementation of an inefficient segmentation strategy in the L2. Crucially, however, she does not entertain the possibility that a late learner can acquire a novel L2 strategy. Thus, even though listeners may avoid the use of an inefficient listening strategy, they may not be capable of acquiring the strategy employed by native speakers. These three factors led Cutler (2001) to presume that the second language learner is “disabled” by the L1 listening strategy and that it is unlikely, if not impossible, that a L2 learner can achieve the listening efficiency of a native listener (p. 4). With these factors in mind, we now consider specific segmentation strategies as they apply to L2 listening, including a discussion of where Cutler’s predictions are borne out and where they fall short.

3.3.1 Lexical Competition

Within the framework of competition-based models of SWR, the limited vocabulary of L2 learners would seem to render L2 comprehension less complicated in that fewer candidates are available for competition than in the L1. Recall that words are recognized more slowly when there are more competitors and, conversely, more rapidly when there are fewer competitors (Luce & Pisoni 1998; Norris, McQueen & Cutler 1995; Vroomen & de Gelder 1995; Dahan, Magnuson & Tanenhaus 2001). However, evidence from behavioral tasks has suggested that the potential for lexical competition is actually

greater in L2 listening due to significant interaction between the L1 and L2 lexica, rendering aural comprehension of the L2 less efficient.

While semantically related items in bilingual lexica have failed to systematically exhibit priming effects across languages (Gerard & Scarborough 1989; Scarborough, Gerard & Cortese 1984; Watkins & Peynircioglu 1983), phonetically related items have been shown to activate similar-sounding competitors in a listener's languages. Weber and Cutler (2004) used the tracking of eye movements to investigate interaction between L1 and L2 lexica in Dutch-speaking participants with high proficiency in English (mean 7.8 years of study). The experiment was conducted completely in English. Participants heard a word and were asked to locate and click one of four images on a computer monitor. Images included items that were phonetically similar in both languages and evidence of activation of L1 items was found. For example, when asked to *click on the desk* participants initially fixated as often on a picture of a lid (*deskel* in Dutch) as on a picture of a *desk*. Similarly, participants fixated on a picture of a church (*kerk* in Dutch) when asked to click on a *carrot*, suggesting that L1 lexical items with similar phonemic content were activated. Parallel activation has also been found for bilinguals of different language pairings such as Russian/English (Spivey & Marian 1999), German/English (Blumenfeld & Marian 2005) and French/Dutch (Van Wijnendaele & Brysbaert 2002).

These results are in line with research on visual word recognition, which has provided evidence that both native and non-native lexical items are activated by visual input. Utilizing a cross-modal priming paradigm Schulpen, Dijkstra, Schriefers and Hasper (2003) investigated inter-lingual homophones and found that, for example, the English visual target LEASE was primed by the Dutch auditory stimulus *lies* 'groin' /li:s/ for Dutch-English bilinguals. This result is particularly interesting in that a related experiment showed that listeners could differentiate productions of the homophones in a gating task, demonstrating that they were sensitive to the fine-grained acoustic differences between the language-specific productions. The fine-grained detail did not, however, inhibit the activation of the homophone in the priming task.

Nas (1983) showed that Dutch/English bilinguals were slower to reject non-words in English (e.g. SNAY) that have homophones in Dutch (*snee* is a Dutch word meaning 'slice') than words with no similar sounding Dutch counterpart such as ROLM. Similarly, Dijkstra, van Jaarsveld and ten Brinke (1998), using a visual lexical-decision task, found that Dutch/English bilinguals responded less quickly to written words having homographs in the two languages (e.g. *brand*, 'fire' in Dutch) than to words that exist uniquely in one language. Increased reaction times suggest that recognition was slowed by increased competition between the listeners' two lexica.

The miscategorization of L2 sounds has also been shown to increase the number of lexical competitors. As discussed above, unfamiliar L2 phonemic contrasts are often misperceived and subsequently miscategorized into existing L1 categories. Take for example the difficulties of Japanese speakers in distinguishing between the English segments /l-r/, which are both subsumed into one Japanese category /r/. Consequently, Japanese learners of English, upon hearing the sequence /bal-/ from *balance* or *balcony*,

could find that lexical candidates beginning with /bar-/ (*barren*, *barely*, etc) are spuriously activated, thus increasing competition.

Indeed, in a second portion of the eye-tracking study by Weber and Cutler (2004) reviewed above, these authors demonstrated that a lack of discrimination of L2 phonemic contrasts erroneously activates competitors. Dutch learners of English have well documented difficulty distinguishing between English /ɛ/ and /æ/, both of which are assimilated to the Dutch category /ɛ/. Upon hearing the English word *panda* Dutch-speaking participants fixated as long and as often on a *pencil* than on the *panda*, presumably due to a lack of distinction between /ɛ/ and /æ/. However, these results were asymmetrical. After hearing *pencil*, there was no difference in fixation between *panda* and two other distracter images, suggesting that input containing the non-native phoneme /æ/ activates words containing the native /ɛ/, but not vice versa. To account for this discrepancy, Weber and Cutler hypothesize that a contrast between these two English vowels may indeed be discriminated at the lexical level, but that at the input level words containing /ɛ/ are activated no matter which of the two vowels is heard. In other words, top-down lexical information helped listeners differentiate *panda* from *pencil* when input is /pændə/, but bottom-up acoustic input failed to activate *panda* when input is /pensl/.

Inter-lingual interference can be bi-directional as well. In line with Flege's predictions discussed above, a line of research has shown that, not only does the L1 provide additional candidates in the L2 SWR process, but the L2 can influence on the activation of L1 lexical candidates. In an eye-tracking study, Ju and Luce (2004) showed L2 interference influence on the activation of L1 lexical items in native Spanish speakers who were advanced learners of English. When VOT values in Spanish words were altered toward more English-like values, participants fixated significantly more on inter-lingual distracters (e.g. images of items whose English names were similar sounding to Spanish items). This effect was not found when words were presented with normal Spanish VOT values.

Furthermore, research suggests that the degree of parallel activation of bilingual lexica can depend on L2 proficiency. For example, Marian and Spivey (2003) found activation of the L1 (Russian) given L2 (English) input in highly advanced bilinguals living in the L2 environment, while Weber and Cutler (2004) found little activation of L1 (Dutch) lexical items given L2 (English) input in less proficient bilinguals living in a L1-speaking environment.

3.3.2 Prosodic Structure

Cutler, Mehler, Norris and Segui (1986) illustrated that adult speech segmentation processes are conditioned by the rhythmic structure of the L1. This study (reviewed in §1.4.2.1), found that monolingual French and English listeners differed in their use of processing strategies while listening to words in both languages. French listeners employed a syllable-based strategy in listening to both French and English, while English listeners showed no evidence of sensitivity to syllable boundaries in either language. Several studies have since confirmed this phenomenon. Van Zon (1997) found that

French-dominant French-Dutch bilinguals did not employ a stress-based segmentation routine in listening to Dutch; they employed a syllable-based routine in both languages. Similarly, Bradley, Sanchez-Casas & Garcia-Albea (1993) found that Spanish-dominant Spanish/English bilinguals did not use a stress-based strategy in English. Otake, Hatano, Cutler and Mehler (1993) found evidence that Japanese listeners retain sensitivity to moraic structure when listening to English, while French listeners employ the syllable in the processing of Japanese. Furthermore, research has shown that listeners whose languages do not incorporate lexical stress (e.g. French, Hungarian) are not sensitive to stress placement changes in non-words (e.g. *bópelo* versus *bopélo* versus *bopeló*), while speakers of languages which do make use of lexical stress distinctions are (e.g. Spanish, English, Dutch; Dupoux, Pallier, Sebastián-Gallés & Mehler 1997; Dupoux & Peperkamp 2002; Dupoux et al. 2008). These authors speculate that speakers of French and Hungarian have no need to encode stress as linguistically relevant as infants and therefore lose sensitivity early in life.

Further research tested highly proficient bilinguals in order to ascertain whether increased L2 exposure could affect sensitivity to rhythmic structure. Cutler, Mehler, Norris and Segui (1989) tested participants who were bilingual in French and English from birth, most having grown up with one English-speaking parent and one French-speaking parent. All participants avowed that they were equally fluent and at ease in both languages. Cutler et al. asked participants to make a hypothetical decision as to which of their languages they would keep to save their lives, which the authors took to be their dominant language. In syllable-monitoring tasks these authors found that the language a participant had chosen predicted his/her use of a segmentation routine. English-dominants used stress-based segmentation in English, but did not employ syllable-based segmentation in French. French-dominants employed syllable-based segmentation in French, but did not use stress-based segmentation in English. Of particular note is the fact that French-dominant bilinguals did not show evidence of applying a syllable-based routine while listening to English, unlike the monolingual French listeners in the previous Cutler et al. (1986) study. While early bilinguals seemed to avoid the use of inappropriate segmentation routines, they point out that the French-dominants in this study did not show evidence of employing a stress-based routine in English. Cutler and colleagues therefore suggest that higher proficiency can lead to an ability to suppress an inefficient listening strategy in a language with a differing rhythmic structure, but they also maintain that the acquisition of a novel rhythmic strategy is not likely.

Several researchers have proposed that the acquisition of rhythmic structure is an indispensable building block for further language development, specifically for the acquisition of individual lexical items. On this view, infants exploit rhythmic properties of the native language in order to identify word boundaries in the initial stages of word acquisition (see Jusczyk 1997 for a review; see also Cutler et al. 1992). Supporting this stance, evidence has indeed suggested that infants form units based on rhythmic regularities before analyzing the distributional properties of segments (Morgan & Saffran 1995). Taking this position to its extreme, Cutler (2001) has proposed that infants, even those born into bilingual environments with equal exposure to languages of differing rhythmic structure, make use of just one rhythmic routine in the acquisition of speech.

She proposes that “the initial launching of segmentation is unique — the assistance is effectively only needed once” (p. 11). Citing evidence from Cutler et al. (1986, 1989, 1992), Cutler and colleagues claim that even listeners who are bilingual from birth remain functionally monolingual and can efficiently command only one rhythmic segmentation procedure as adults.

However, Golato (2002) found that limits on prosodic processing in late learners can be asymmetrical depending on language dominance. While Cutler et al. (1989) tested early bilinguals, Golato undertook a partial replication of this study with late-learning French/English bilinguals. All participants had begun study of the L2 as adults. Contra Cutler et al., Golato found that late-learning bilinguals who were English dominant were in fact able to implement multiple segmentation strategies, using a stress-based strategy when listening to English and a syllable-based strategy when listening to French. French dominants on the other hand showed evidence of syllable-based segmentation in both languages. These findings bring into question the inevitability of late-learners' inability to utilize multiple segmentation strategies based on prosody. Golato's work suggests that late learners may indeed be able to learn and apply alternate rhythmic segmentation procedures.

3.3.3 Probabilistic Phonotactics

As discussed in Chapter One, languages differ as to which groupings of phonemes constitute licit sequences in the same syllable or word. Phonotactic patterns have been shown to influence word recognition and the identification of word boundaries in the L1 (McQueen 1998; see §1.4.2.2). In a study investigating the influence of L1 phonotactic patterns on L2 processing, Weber and Cutler (2006) looked at the identification of English words embedded in nonsense strings by both native speakers of English and German. The German participants had an advanced level of English (mean 15 years of study), while the English participants had no knowledge of German. English words were embedded in four conditions: Condition 1: syllable boundaries aligned according to phonotactics of both languages (e.g. *loft* in /fumlɔft/; /ml/ is an illegal onset in both languages); Condition 2: syllable boundaries misaligned according to phonotactics of both languages (e.g. *loft* in /zarplɔft/; /pl/ is an allowed onset in both languages); Condition 3: syllable boundaries aligned according to phonotactics English only (e.g. *loft* in /prarfɔft/; /ʃl/ is an illegal onset in English, but not German); and Condition 4: syllable boundaries aligned according to phonotactics of German only (e.g. *loft* in /forslɔft/; /sl/ is an illegal onset in German, but not English). Previous findings by McQueen (1998) predict that response latencies will be shorter when phonotactic constraints force a syllable boundary to be placed before the word to be identified. In other words, response latencies for English participants (with no knowledge of German) are expected to be shorter in Conditions 1 and 3 relative to the other conditions. German participants, on the other hand are expected to show shorter responses latencies in Conditions 1 and 4. However, if German participants have also acquired the use of English phonotactic patterns, they are expected to react more quickly to sequences in Condition 3 as well.

Results for the English speakers replicated McQueen's (1998) earlier findings on L1 phonotactic patterns. Reaction times were shorter when syllable boundaries aligned due to phonotactic constraints. German listeners showed evidence of exploiting both English and German phonotactic constraints when monitoring for English words. Response latencies were similar to English speakers in the English-specific condition (3) as relative to the control condition (1), but were also speeded by German phonotactic patterns (Condition 4). This pattern of results suggests that though participants could not inhibit L1 constraints, they had gained sensitivity to L2 constraints.

Also noteworthy is the fact that, though the native English speakers were significantly faster and more accurate overall, reaction time and error rate distributions had significant overlap for both groups, demonstrating that some German participants were performing at native levels on a task involving exclusively L2 materials.

Further evidence of the inhibition of native phonotactic constraints in the L2 comes from vowel harmony. Suomi, McQueen and Cutler (1997) demonstrated that Finnish participants exhibit sensitivity to vowel harmony constraints in the segmentation their native language. A subsequent word-spotting study showed that Finnish listeners with only low to medium proficiency were able to inhibit vowel harmony effects in listening to English (Cutler, McQueen & Suomi 1997). Participants, for example, spotted *charge* with equal speed in *daepcharge* (where the two English vowels violate harmony constraints) as in *darpccharge* (where vowels are identical and thus harmonious). These results suggest that Finnish speakers are in fact able to 'turn off' a native phonotactic segmentation strategy that would be inappropriate to the L2 even at low proficiency levels.

3.3.4 Non-Contrastive Allophonic and Phonological Variation

The acquisition of L2 phonemic contrasts has generated an extensive body of work, including the phonological acquisition models of Best (1995) and Flege (1995) reviewed above. Much of this work has focused on the fact that novel phonemic contrasts in the L2 are initially represented as within-category distinctions in the L1. This necessitates that L2 learners create phonemic boundaries where none exist in the L1, as is the case for the English /l-r/ distinction for Japanese learners. However, far less research focus has been placed on the acquisition of within-category allophonic variation in the L2. Models of the acquisition of L2 phonology discussed thus far offer no predictions as to how non-contrastive phonetic detail is perceived and/or exploited by the L2 learner. We have discussed the fact that allophonic variation and fine-grained acoustic detail play a significant role in L1 spoken word recognition (see § 1.4.2.3). Furthermore, evidence from gating experiments has shown that bilingual listeners are indeed sensitive to fine-grained acoustic differences between their languages. Both Grosjean (1988) and Li (1996) showed that bilinguals can detect in which language a word is produced given very little acoustic information. Nonetheless, the capacities of L2 learners in the exploitation of non-contrastive detail in the perception of continuous speech have received little attention to date, with the exception of the studies we turn to now.

Altenberg (2005) examined acoustic cues to word juncture in L2 English in a partial replication of the Nakatani and Dukes' (1977) study reviewed above in §1.4.2.3.

Altenberg's study looked at cues to the segmentation of potentially ambiguous sequences in English (e.g. *keep stalking* versus *keeps talking*) by both native speakers of English and native speakers of Spanish learning English. She found that Spanish speakers at both intermediate and advanced levels of proficiency in English performed significantly above chance in the identification of ambiguous stimuli (mean accuracy 76%), but that they also performed significantly worse than native speakers (mean accuracy 97%). In the production sample used in this study, acoustic analyses showed many of the same juncture cues found in the original Nakatani and Dukes' (1977) study (e.g. duration, glottalization before vowel-initial words, aspiration of word-initial stops). Stimuli which included glottalization as a cue (e.g. *an ice man* vs. *a nice man*) were identified with significantly more precision by the Spanish-speaking group (mean accuracy 88.4%) than those which included aspiration (e.g. *keeps ticking* vs. *keep sticking*; mean accuracy 58.5%). Altenberg attributed this discrepancy to L2 transfer, noting that glottal stops can occur in emphatic speech in Spanish, whereas the aspiration of Spanish plosives does not occur in any environment. Though participants performed significantly worse on stimuli incorporating aspiration as a cue, it should be noted that they did perform significantly above chance. This is significant in that these L2 learners acquired the use of a non-contrastive acoustic-phonetic cue that is not used in the L1. However, as Altenberg herself points out, this study did not include a monolingual Spanish control group and therefore it remains uncertain if Spanish speakers with no knowledge of English would also be sensitive to this cue.

We have already discussed language-specific assimilation as one factor that introduces variation into the speech signal (see §1.0). Evidence from Gaskell and Marslen-Wilson (1998) showed that listeners compensate only for changes resulting from licensed processes in their language. These authors explored place assimilation in English using a phoneme-monitoring task in which listeners were asked to monitor for the segment /t/. Listeners identified /t/ in phonologically viable assimilation environments such as *freight* [pb] *bearer*, where /t/ is assimilated to the following bilabial /b/ and is physically realized as /p/. This suggests that listeners had in effect compensated for assimilation and recovered the underlying form. However, listeners did not recover /t/ in a phrase such as *freight* [pk] *carrier*, where its realization as /p/ is not phonologically motivated. Following these results, Gaskell and Marslen-Wilson (1998) propose that compensatory processing for phonological variation in speech is language-specific and that listeners only compensate for variation that is relevant to their language. Subsequent studies have shown similar effects in German (Coenen, Zwitserlood & Bölte 2001) and Dutch (Mitterer & Blomert 2003).

Expanding on this line of research, Darcy, Peperkamp and Dupoux (2007) examined learners' sensitivity to L2 assimilation processes that differ from those of the L1. Recall that while segments in English tend to be assimilated for place of articulation, French tends to assimilate more for voicing. In this study, Darcy et al. employed an identification task to probe late learners' sensitivity to L2 assimilation processes. Native English-speaking learners of L2 French (beginning and advanced) and native French-speaking learners of L2 English (beginning and advanced) were all tested on both English and French stimuli in a word-identification task. Participants heard a word in isolation

and then a sentence. They were asked to indicate if the word presented in isolation was identical to a word in the sentence. Half the stimuli presented licensed instances of assimilation in that language (i.e. voicing assimilation in French as in *botte* [dg] *grise* ‘grey boot’, or place assimilation in English, see above) while the other half presented instances of assimilation where an inappropriate process was applied to that particular language. For example, French stimuli representing unlicensed contexts included the sequence *lune pâle* ‘pale moon’ produced as [lympal], where the final /n/ of *lune* is assimilated to the place of articulation of following /p/, while in spoken French no assimilation would occur. English stimuli representing unlicensed contexts included sequences such as *big fountain* produced as [bɪkfauntɪn], in which the final /g/ of *big* is assimilated to following unvoiced /f/.

Darcy et al. found that beginning learners of both French and English were not sensitive to L2 assimilation processes. In addition, they continued to compensate for L1 assimilation processes when listening to the L2. However, these researchers also found that very advanced L2 learners of French not only showed sensitivity to the L2 assimilatory processes, but they were also able to inhibit the L1 assimilatory processes when processing the L2. Darcy et al. do point out that some universal compensatory mechanisms could be involved and that the processes investigated in this study are not entirely language-specific. For example, some degree of voicing assimilation can occur in English. However, the fact that advanced learners were more sensitive to voicing assimilation in French than less proficient learners suggests that this sensitivity is due to prolonged exposure to the L2 and not to the exploitation of native compensation strategies. These results go against claims that listeners possess a single phonological grammar and are inconsistent with claims of strict limitations on the plasticity of phonological learning and perceptual processing.

3.4 Summary

The evidence on the attainment and exploitation of L2 perceptual strategies discussed thus far is mixed. At the segmental level, research has shown that advanced L2 learners are sensitive to extremely fine-grained detail differentiating their languages (e.g. Bohn & Flege 1993; Hazan & Boulakia 1993), including the ability to differentiate putatively homophonous lexical items in the two languages (Schulpen et al. 2003). On the other end of the spectrum, however, evidence has also been offered that some advanced bilinguals remain ‘deaf’ to non-native contrasts even when these contrasts are characterized by considerable physical variation. For example, Pallier, Colomé, Sebastián-Gallés (2001; see also Pallier et al. 1997; Sebastián-Gallés & Soto-Faraco 1999) found that native Spanish speakers who had learned Catalan as early as primary school and lived in bilingual Spanish/Catalan environments remained less sensitive to Catalan-specific segmental contrasts (e.g. /s-z/, /ʃ-ʒ/, /ɔ-o/, /e-ɛ/) than native Catalan speakers. How can we account for this discrepancy?

Avoiding L1 interference and ‘language-specific listening’ appears to be dependent on myriad factors, including age of acquisition, language dominance, and proficiency level, among others. We have already noted inter-lingual lexical competition

can be mitigated by L2 proficiency (Marian & Spivey 2003; Weber & Cutler 2004). The degree of phonetic overlap between languages can affect the level of parallel activation as well. Blumenfeld and Marian (2005) found evidence that manipulating the degree of phonetic overlap at a word's onset can affect the degree of inter-lingual activation. One can also imagine that language setting and amount of L2 use can greatly influence parallel activation. A L2 speaker surrounded by monolingual speakers of the target language could experience less L1 interference than L2 speakers in bi- or multi-lingual settings.

Concerning the segmentation of L2 speech, one additional factor affecting the degree of L1 interference appears to be the nature of the particular perceptual cue under investigation. The above body of research suggests that varying degrees of plasticity and entrenchment exist for particular cues. Listeners of Finnish, for example, with only moderate proficiency in English are able to inhibit vowel harmony effects in English listening tasks (Cutler et al. 1997), while highly competent French/English bilinguals have shown mixed results in the use of prosody-based processing (Cutler et al. 1986, 1989).

Cutler (2001) suggested that, in the case of prosodic processing, advanced learners can inhibit the native strategy, but yet cannot learn a non-native strategy. However, research on the exploitation of L2 phonotactic constraints showed the inverse pattern. Highly proficient German learners of English were able to exploit the phonotactic patterns of English, but at the same time did not seem able to inhibit sensitivity to German phonotactic constraints (Weber & Cutler 2006).

Ongoing research on L2 processing will thus need to address the question of which L2 perceptual strategies are learnable to nativelike levels, and, furthermore, the degree to which learnability is dependent on exposure and/or training. The amount of exposure to language input is clearly a major factor differentiating L1 from L2 acquisition. The study of perceptual learning has demonstrated the plasticity of L1 categories even after extremely limited exposure (e.g. Clarke & Garrett 2004; Maye et al. 2008; Norris, McQueen & Cutler 2003). However, in the L2, several studies have suggested that even exposure from childhood is not always sufficient for nativelike perception (Pallier et al. 1997; Cutler et al. 1989; Peperkamp et al. 1999).

Phonetic training has shown promising results for the perception of segmental contrasts (e.g., Bradlow, Pisoni, Akahane-Yamada & Tohkura 1997; Lively, Pisoni, Yamada, Tohkura & Yamada 1994; McCandliss, Fiez, Protopapas, Conway & McClelland 2002; McClelland, Fiez & McCandliss 2002). Furthermore, the attainment of nativelike pronunciation has been linked to targeted phonetic instruction (Birdsong 2003; Bongarts 1997, 1999). Nonetheless, to our knowledge, no training studies concerning L2 segmentation routines have been carried out. We will return to this apparent gap in the research in Chapter Nine, including speculation as to which segmentation cues could be amenable to training.

CHAPTER FOUR: Processing French as a Second Language

*On sait la peine qu'a un nouvel élève à comprendre les mots
qu'il a appris isolément, dès qu'ils sont réunis dans une phrase.
L'élosion, la liaison et l'enchaînement aidant, le groupe rythmique
lui donne l'impression d'une succession de syllabes sensiblement égales
dont le rythme ininterrompu cache malicieusement les limites des mots.*
(Delattre 1966:141)

[One knows the difficulty that a new student faces in understanding words learned in isolation once they are put together in a sentence. Because of elision, liaison and *enchaînement*, the [French] rhythmic group gives him/her the impression of a sequence of equal syllables in which word boundaries are maliciously hidden by the uninterrupted rhythm.]

4.0 Introduction

The perception and processing of French phonological phenomena such as liaison and resyllabification by *native speakers* has been the focus of an extensive body of work, the bulk of which we reviewed in Chapter Two. The perceptual capacities of learners of L2 French, however, have received markedly less attention. The current chapter reviews what little work has been undertaken to date on this topic and explores the question as to whether non-native speakers of French employ the same strategies in the processing of continuous speech as their native speaker counterparts.

4.1 The Acquisition of L2 French Phonology

An analysis of the acquisition of L2 French phonology reveals that adult learners experience difficulties with the same phonological processes in spoken French that pose problems in L1 development, namely phenomena which are characterized by variability including schwa deletion, elision, liaison, and *enchaînement*. Saunders (1988) conducted a study investigating the systematicity of comprehension errors in spoken French in

which 45 third- and fourth-year American university students of French phonetics were tested on their perception of a three-minute recording of radio French. The students' task was to write what they heard after repeatedly listening to the extract over a 30-minute period. What makes this analysis particularly noteworthy is the fact that the students, who were advanced students of phonetics, were able to employ phonetic transcription for sequences that they did not recognize as French words or for which they were unsure of the orthography, thereby allowing for a precise analysis of perceptual errors.

Saunders found a great deal of consistency among the students regarding which sequences were misperceived, a fact she took to be indicative of the specific areas of perception that may pose problems for (English-speaking) learners. Saunders grouped these errors into 11 areas of French phonology ranked by their frequency of occurrence among the students. The three most frequent errors arose in instances of schwa deletion, the misanalysis of syllable boundaries, and what Saunders refers to as 'linking phenomena' (liaison and *enchaînement*¹⁶). For example, in cases of the resyllabification of fixed final consonants as in *le monde entier* [lə.mɔ̃.dɑ̃.tje] 'the whole world', where word-final /d/ is resyllabified as a word onset, several students either misanalyzed /d/ as a word-initial elided form (e.g. **le mot d'entier* or **le mois d'entier*) or failed to perceive the presence of /d/ at all (e.g. *le mot entier* or **le moitié*). Errors in the perception of resyllabified liaison consonants were of a similar vein. In the sequence *il vous devient* [t] à ce point nécessaire... 'it becomes so necessary to you...', where the speaker produced an optional liaison with the final /t/ of *devient*, only six of the 45 students heard the phoneme /t/ at all, three of which correctly identified it as a liaison /t/. Seven students perceived it as word-initial /d/ and four as word-initial /l/. The phrase *on est élu* [ɔ̃.ne.te.ly] 'one is elected', also a case of optional liaison with /t/, generated the same misperception for over a third of the students; the phrase was perceived as *on était lu* 'one was read', where resyllabified /t/ is analyzed as word-medial.

In attempting to account for learners' difficulties in the perception of variable phonological phenomena, Saunders (1988) frames perceptual difficulties in terms of depth of processing. She proposes, in effect, that non-native speakers are "unable to penetrate the surface variability in French speech to hear the underlying entities" as native speakers are (p. 94). Her claim is that this inability renders learners overly reliant on the surface form of the signal.

4.2 Production of L2 French

Saunders' (1988) findings on perceptual errors are consistent with production data from L2 French, which indicate that late learners of French struggle with resyllabification of both fixed coda consonants as well as consonants that surface in liaison. Both Mastromonaco (1999) and Thomas (2004) examined the production of liaison by university-level students of French across the three categories of liaison realization (obligatory, optional, and prohibited) and found that students systematically failed to

¹⁶ Saunders does not make explicit what differentiates problems with syllable boundaries from 'linking phenomena'. Given that 'linking phenomena' involve resyllabification across word boundaries resulting in the misalignment of syllable boundaries, there is presumably some overlap between these two categories.

resyllabify consonants that surface in liaison¹⁷. Seven percent of liaison consonants were not resyllabified as the onset of the following word in Mastromonaco's sample while 8.5% were not resyllabified in Thomas' sample. Furthermore, Thomas found that errors in the realization of liaison accounted for roughly 20% of all pronunciation errors committed by the students in his production sample.

4.3 The Perception of Liaison in L2 French

The above studies underscore the fact that the frequent misalignment of syllable and word boundaries in spoken French can render the segmentation of the acoustic signal difficult for learners of the language. Contrary to an extensive body of work examining the perceptual capacities of French native speakers (reviewed in §2.2 above), to our knowledge, there is little existing work specifically focusing only on the perception and processing of liaison by adult learners of L2 French. A search of the literature produced four studies published over the past two decades examining the perception of liaison by non-native speakers. We review each in turn here.

The first study to examine the perception of liaison by non-native speakers of French was undertaken by Matter (1989) who compared the effects of liaison on word recognition between native French speakers and Dutch learners of L2 French in word-monitoring tasks. In the first of two experiments, Matter examined the recognition of vowel-initial words in optional environments of liaison (e.g. recognition of the target *assez* in *vous en prenez assez* 'you take enough of it' where final /z/ in *prenez* is optionally realized). Participants heard phrases produced both with and without liaison (e.g. [vu.zã.pʁə.ne.zã.se] or [vu.zã.pʁə.ne.a.se]) and were asked to monitor for the presence of /a/-initial words. An analysis of response latencies revealed no differences in word detection between realization and non-realization of liaison; /a/-initial words were recognized equally fast in both conditions by both participant groups, suggesting that the process of liaison does not retard word recognition (as would be later confirmed for native speakers in studies by Gaskell et al. 2002 and Spinelli et al. 2003; see §2.2). However, an analysis of reaction times between the two groups did reveal that native speakers were faster overall at executing the task than the non-native speaker group.

In a second experiment, Matter (1989) looked at the effects of local lexical ambiguity in cases of potential liaison. Participants were asked to detect the presence of /t/-initial words preceded either by a word that could trigger liaison (e.g. *grand* in *C'est un grand tableau* 'It's a big painting') or by a word where liaison is not possible (e.g. *beau* in *C'est un beau tableau* 'It's a beautiful painting'). Mean reaction times showed significant differences both between conditions and between participant groups. Both native and non-native speakers found it significantly more difficult to detect /t/-initial words in locally ambiguous phrases (e.g. *C'est un grand tableau*) than in unambiguous phrases (e.g. *C'est un beau tableau*). In addition, as in the first experiment, native speakers performed that task significantly faster than non-native speakers.

¹⁷ Both Mastromonaco (1999) and Thomas (2004) code a lack of resyllabification of liaison consonants as a production error, assuming that student participants are not intentionally producing *liaison sans enchaînement* as would be found in higher registers of speech such as political discourse (Encrevé 1988).

Matter's (1989) results suggest that the non-native speakers process liaison (and potential liaison) employing strategies similar to those of native speakers, as evidenced by similar differences in reaction times for both groups between potential and impossible liaison. The non-native speaker group showed clear evidence of having developed perceptual strategies for the recognition of environments of liaison. The difference in overall reaction times between the two participant groups, however, indicates that this processing may incur a greater processing load for learners than for native speakers.

Dejean de la Bâtie and Bradley (1995) extended Matter's work in comparing the perceptual performance of native French speakers and native English speakers who were university students in second year French. This study was reviewed in §2.2 above, however we return to it here for further discussion of non-native participant performance. In the first part of the study, both groups of participants (native and non-native speakers of French) were asked to monitor for the presence of word-initial /t/ in four conditions (while the Matter (1989) study had employed just two conditions): a /t/-initial word preceded by a word that could potentially trigger liaison with /t/ (e.g. *grand théâtre* 'big theater'); a /t/-initial word preceded by a word where liaison is not possible (e.g. *vrai théâtre* 'real theater'); a vowel-initial word preceded by liaison with /t/ (e.g. *grand éléphant* 'big elephant'); and finally a vowel-initial word preceded by a word where liaison would not be possible (e.g. *vrai éléphant* 'real elephant'). Confirming Matter's results as to the perception of potential liaison, an analysis of response latencies showed that both native and non-native participants were significantly slower to recognize /t/-initial words in potential liaison contexts (e.g. *grand théâtre* 'big theater') relative to the other conditions. The cost of potential ambiguity for the native speakers was an average of 64 ms, while the non-native speakers were an average of 97 ms slower in potential liaison environments. The difference in reaction times between the groups, however, was not significant and thus did not confirm Matter's results showing an increased processing load for non-native speakers.

An analysis of mean accuracy rates, however, did reveal a difference between the two groups. Native speakers showed error rates of less than 1% in the detection of initial /t/ in potential liaison environment and 0% where liaison was not possible. Late learners on the other hand exhibited error rates of over 10% in potential liaison environments and 2% in cases where liaison was not possible. The authors attributed this discrepancy to a lack of sensitivity to liaison environments on the part of the non-native speakers. However, as in Matter's (1989) study, the fact that non-native responses were delayed by the potential liaison environment indicates that the non-native speakers had developed processing strategies for dealing with local ambiguity due to this particular phonological process.

In the second portion of the study, the same stimuli were presented in a contextualizing frame (e.g. *excellent acteur* 'excellent actor' and *vrai acteur* 'real actor' preceded by *Dans ce film, j'ai découvert un...* 'In this film, I discovered a ...'). These results revealed significant processing differences between the two groups. The difference in reaction times between potential and impossible liaison environments for the native speakers disappeared, while non-native speakers' responses continued to be delayed by instances of potential liaison. In other words, locally ambiguous acoustic

input initially delayed processing for native speakers in a context-free environment. In contrast, processing costs were eliminated by higher level information gleaned from semantic context, suggesting that top-down information was able to override bottom-up ambiguity. The non-native speakers, on the other hand, were not able to use the contextual information as efficiently as native speakers and processing costs were still incurred by potential ambiguity. It is worth noting however that overall error rates for the non-native group were lower in the context condition (8.9% error in biasing context versus 12.8% without context¹⁸), suggesting that they were making use of contextual information albeit to a lesser extent than the native speakers.

Further examining the effects of phonemic ambiguity due to liaison, Stridfeldt (2003) tested native French speakers and speakers of Swedish studying first year university French on the perception of ambiguous pairs of non-words in liaison environments. The stimuli consisted of ambiguous minimal pair phrases containing non-words produced by a native speaker as either vowel-initial or consonant-initial beginning with one of the LCs /n, v, t, or z/ (e.g. *un avas/un navas*, *les avas/les zavas*), which were presented to participants in the carrier phrase *Je vois ___*, 'I see ___'. The participants' task was to report what they heard. Neither participant group consistently classified the stimuli as intended by the speaker. Moreover, the non-native participants showed an overwhelming predisposition for liaison situations; a great majority of the phrases were reported as vowel-initial by the non-native group. The native speaker group showed a similar bias toward perceiving the stimuli as vowel-initial, but to a lesser degree than the non-natives. The learners' bias toward vowel-initial words suggests that they had developed processing routines and mental representations of liaison consonants, but that these representations were overgeneralized to all possible instances of liaison. If no mental representation of liaison environments had been established by these learners, we would expect a bias in the opposite direction, namely the majority of words would have been perceived as consonant-initial as the surface forms imply.

The three studies reviewed above tested participants with low to medium proficiency levels in L2 French, the majority of whom were undergraduate university students. More recent work has examined the perception of L2 liaison by highly advanced learners. Shoemaker and Birdsong (2008), discussed in §2.4.3 above, compared the perception of globally ambiguous phrases by native French speakers with that of native English speakers who were professors of university-level French or graduate students of French literature and linguistics.

Recall that in this study neither the native nor non-native speakers showed any systematic ability to discriminate minimal pair phrases rendered globally ambiguous by the possibility of liaison (e.g. *Il n'a aucun air* or *Il n'a aucun nerf*), however, the two groups did behave similarly in the distribution of their responses for the majority of the pairs. Overall mean accuracy rates for both groups were roughly at chance (native French participants: 53.2%; L2 French participants: 52.7%) and there was no significant difference between the two groups ($t = 0.236$, $df = 28$, NS), however, mean accuracy by individual stimulus item was strongly correlated between the two participant groups in

¹⁸ The statistical significance of this difference is not given in the original text.

both a simple regression analysis ($r = .572$, $p = .004$) as well as a Spearman rank correlation ($\rho = .605$, $p = .004$). A further investigation of the similarities between the two participant groups revealed that they agreed considerably in the directionality of their response distribution as well. That is, participants consistently chose one member of a minimal pair significantly more than the other in 10 out of 12 of the stimulus pairs. This distribution suggests that both groups' responses were guided by similar information present in the acoustic signal. It should be noted, however, that non-native responses were characterized by greater variability than those of the native speakers.

Shoemaker and Birdsong (2008; see also Shoemaker 2005) performed several post-hoc analyses in an attempt to identify the particular factor that was guiding responses (e.g. correlation of individual responses with pivotal consonant duration, lexical frequency, plausibility), none of which revealed a significant effect. Though Shoemaker and Birdsong were not able to identify the factors underlying the response biases found in both participant groups, the results suggest that the highly advanced learner group had developed some sort of perceptual strategy that mirrors that used by the native speakers.

These results taken together with the findings of Darcy et al. (2007; see §3.3.4) and Golato (2002; see §3.3.2), both of which showed evidence of nativelike processing strategies on the part of advanced adult learners of L2 French, underscore the need for further research on the perceptual abilities of advanced learners in this domain.

PART TWO:
AN EMPIRICAL INVESTIGATION OF
ACOUSTIC CUES TO
SPEECH SEGMENTATION IN FRENCH

PART TWO: The Current Study

The second portion of this dissertation presents an empirical investigation into the saliency of acoustic cues to liaison in spoken French and the processing of lexical ambiguities by both native French speakers and adult learners of L2 French. Though research has shown that both LC and IC candidates in liaison environments are activated in SWR processes (Gaskell et al. 2002; Spinelli et al. 2003), the question remains as to whether this activation is triggered solely by durational differences that arise between the two realizations of these pivotal consonants. Spinelli et al. (2003) suggested that these durational differences are robust enough to “bias interpretation in the correct direction” in cases of ambiguity (p. 250). Recall, however, that the priming results of this study did not demonstrate a direct relationship between durational differences and processing.

Given the nature of the stimuli used in the Spinelli et al. (2003) study, it is impossible to isolate durational variation as the sole factor allowing for the activation of candidate words in the priming task. Each token of the ambiguous minimal pairs used in this study was produced as a separate utterance (e.g. *Il n'a aucun air* and *Il n'a aucun nerf*) and it is therefore conceivable that there were other acoustic cues available to listeners that were not addressed by the authors. These authors reported durational measurements from three segments in each phrase (the pivotal LC or IC consonant and preceding and following vowels), but they did not report measurements from other acoustic parameters such as formant transitions, F0, VOT in plosives, etc. Therefore, the authors' statement that listeners exploit “subtle but reliable [durational] cues” in cases of global ambiguity remains conjectural (p. 248). Spinelli et al. (2003) conclude that

...further research is required to confirm that consonant duration is indeed the only cue which French listeners use to distinguish between liaison and non-liaison utterances. Nevertheless, it seems reasonable to assume on the basis of the current evidence that, while other cues may be involved, durational differences are at least an important part of this distinction (p. 248).

One way to verify the use of duration as a segmentation cue is to manipulate this one acoustic factor in the same physical utterance, thus holding all other acoustic factors in the signal constant. To this end, the current study employs both an AX discrimination task and a forced-choice identification task which utilize sequences in which the pivotal consonants in ambiguous environments of liaison (i.e. /n/ in [œ.nɛʁ], *un air* or *un nerf*) are instrumentally shortened and lengthened while the rest of the utterance remains unaltered. An AX discrimination task is employed to tap lower-level acoustic processing, while a forced-choice identification task is used to investigate the use of segmental

duration in higher-level lexical decision processes. In this way we can test whether the durational variation of the pivotal consonants represents a sufficient acoustic cue for segmentation.

The specific aims of the study presented in the second portion of the dissertation are as follows:

- 1) to provide new production data from native speakers of French which will shed light on the systematicity of acoustic differentiation between consonants that surface in liaison environments and consonants in word-initial position (Chapter Five);
- 2) to establish thresholds of perceptual saliency of durational differences between consonants that surface in liaison and word-initial consonants for native and non-native speakers of French (Chapter Six);
- 3) to determine the extent to which segmental duration modulates lexical access and the segmentation routines of native and non-native speakers of French (Chapter Seven);
- 4) to expand the current body of knowledge on the plasticity of the adult perceptual system and upper limits of L2 speech processing through an investigation of the exploitation of non-contrastive phonetic detail by late L2 learners (Chapter Eight);
- 5) to bring new data to bear on the exploitation of fine-grained phonetic detail in SWR processes and to examine how the use of such detail can be accommodated by currently accepted models of SWR (Chapter Nine).

CHAPTER FIVE: Materials Production

5.0 Introduction

The current chapter presents the creation, measurement and instrumental manipulation of a production sample containing globally ambiguous phrases in spoken French. These phrases incorporate both real lexical items (e.g. *un air* ‘a melody’ and *un nerf* ‘a nerve’, both [œ.nɛʁ]) and non-word items (*un upe* and *un nupe*, both [œ.nyp]). Production tokens from this sample are subsequently manipulated and employed in the perception portion of the current study, which includes an AX discrimination task (presented in Chapter Six) and a forced-choice identification task (presented in Chapter Seven). Both perceptual tasks utilize sequences in which the pivotal consonants (i.e. the /n/ in *un air/nerf* [œ.nɛʁ]) are instrumentally shortened and lengthened while all additional acoustic information is held constant.

5.1 Creation of Production Sample

Though ultimately the recordings from one native French speaker were selected to serve as stimuli in the behavioral tasks of the current study, it was decided to record and take measurements of multiple tokens from six native French speakers so as to gather enough production data to warrant descriptive generalizations to a larger speaker population. In this way, this production sample offers further evidence of systematic durational differences that arise between LCs and ICs in spoken French.

5.1.1 Production Participants

The participants in the production portion were six native speakers of French (5 female and 1 male) aged 25-32 years old (mean 27.3 years). All were graduate students at the University of Texas at Austin. Participants had all had extensive exposure to English (range: 14-20 years of study; mean: 15.7 years) and had lived in the United States for an extended period at the time of recording (range: 2-6 years; mean: 3.67 years). Regional and dialectal differences were not controlled for¹⁹. Four speakers were from the south of France (Lyon, Toulouse, Montpellier and Garonne) and two speakers were from the Paris metropolitan area. Participants were naïve to the purpose of the recordings and were paid \$5 for their participation.

¹⁹Dialectal differences may have in fact influenced some participants in the perception portion of the study. This methodological issue is addressed in Chapter Nine.

5.1.2 Materials

Of the six consonants that surface in liaison environments in French, /g, n, p, ʁ, t, z/, three, /n, t, z/, were chosen to be included in this study for two reasons. First, these three segments represent three different degrees of obstruence (nasal, plosive and fricative respectively). Employing segments with varying degrees of obstruence allows us to investigate whether the systematicity and/or robustness of durational differences that arise in environments of liaison vary as a function of consonant class.

Second, these three segments were chosen due to their frequency of occurrence in environments of liaison in contemporary spoken French. According to *Phonologie du français contemporain*, a corpus based on speech samples from 600 native French speakers from various regions (www.projet-pfc.net; see Durand, Laks & Lyche 2002, 2005 and Durand & Lyche 2008 for a full description of this corpus), these three consonants are the most commonly realized in liaison environments. Table 5-1 shows the frequency of occurrence of each of the six liaison consonants out of the 9920 realizations of liaison in this corpus.

Table 5-1: *Frequency of occurrence of six liaison consonants in Phonologie du français contemporain corpus* (www.projet-pfc.net).

Liaison Consonant	Frequency of Occurrence in PFC
/g/	0
/p/	9
/ʁ/	13
/t/	1665
/n/	3689
/z/	4544
Total	9920

Previous studies have shown similar proportions. According to Léon (1992) roughly 50% of liaisons are realized with /z/, while /n/ and /t/ each account for approximately 25% of liaison occurrences. Realizations of liaison with /g, p, ʁ/ make up less than 1% combined.

Furthermore, the liaison consonants /g, p, ʁ/ were excluded not only due to their relatively low frequency of occurrence in liaison environments in contemporary usage, but for the following reasons as well. The voiced velar plosive / g / is rarely realized in liaison contexts in contemporary French (and in fact showed zero occurrences in the PFC database as noted above). Furthermore, the realization of /g/ in liaison is limited to the

adjective *long* ‘long’ and some fixed expressions (e.g. *sang* [gɛ̃] *impur* ‘impure blood’ as sung in the French national hymn *La Marseillaise*). In addition it can be realized as either /g/ or /k/ depending on the speaker. There have also been documented instances of speakers inserting epenthetic liaison consonants for /g/ in a liaison environment. For example, one participant in the Shoemaker (2005) study cited above reported that she would pronounce *long hommage* ‘long homage’ as [lɔ̃tomaʒ] with an epenthetic /t/ serving as the LC. The realization of /g/ in environments of liaison is therefore both rare and inconsistent.

The voiced uvular fricative /ʁ/ was excluded because it can be accompanied by a change in the quality of the preceding vowel. For example, in isolation the word *dernier* ‘last’ (masculine) is pronounced [dɛʁnjɛ] with a final close /e/ whereas in contexts of liaison this vowel is often produced as it would be in its feminine form *dernière* [dɛʁnjɛʁ] with a final open /ɛ/. This vocalic alternation is attributed to the Closed Syllable Adjustment rule in spoken French as proposed by Tranel (1984) which refers to the process by which close-mid vowels, /e, ø, o/, in open syllables are lowered to open-mid vowels, /ɛ, œ, ɔ/ respectively, in closed syllables. There was therefore some concern that if speakers in the production portion of the study treated the /ʁ/ in liaison environments as a word-final segment instead of a (resyllabified) word-initial segment this would entail a difference in vowel quality which could potentially signal the intended word boundary to a listener in the perception portion of the study.

The voiceless bilabial plosive /p/ was excluded mainly for syntactic as opposed to phonological reasons. As discussed in §2.0 above, the realization of consonants in liaison environments is syntactically as well as phonologically conditioned. The realization of /p/ in liaison is restricted to the two adverbs *beaucoup* ‘a lot’ and *trop* ‘too much’, both of which represent cases of optional liaison. Given that /p/ is realized uniquely in optional liaison contexts, it was excluded from investigation. Specifically, it would be difficult to ensure that participants in the production portion consistently produce /p/ in liaison environments. Moreover, if a listener tends not to realize liaison with /p/ in his/her own speech, h/she could be biased toward perceiving a /p/ in a liaison environment as an initial /p/ in the perception portion of the study.

For the reasons elucidated above, the three liaison consonants that were included in this study are /n/, /t/, and /z/. Two lists of phrases (real and non-words) including these three segments in globally ambiguous phrases were created. For the real-word phrases, four vowel-initial words for each of the three consonants were selected such that the realization of these words preceded by a word that triggers liaison (in this case *un*, *grand* and *les*) give rise to a sequence that is ostensibly homophonous. For example, the word *air* ‘melody’ [ɛʁ] preceded by *un* [œ̃], the singular masculine indefinite article, yields a phonemic sequence consistent with both *un air* ‘a melody’ and *un nerf* ‘a nerve’, [œ̃.nɛʁ]. This selection process resulted in a total of 24 real-word targets ranging from 2 – 4

syllables (mean 2.83 syllables) and from 4 – 9 phonemes (mean 6.42 phonemes), which are presented below in Table 5-2.

Table 5-2: *Real-word target pairs.*

Consonant	Vowel-initial (liaison) target	Consonant-initial target	Pronunciation
/n/	<i>un hectare</i> 'a hectare'	<i>un nectar</i> 'a nectar'	[œ̃.nɛk.taʁ]
	<i>un aval</i> 'a support'	<i>un naval</i> 'a naval officer'	[œ̃.na.val]
	<i>un air</i> 'a melody'	<i>un nerf</i> 'a nerve'	[œ̃.nɛʁ]
	<i>un œuf</i> 'an egg'	<i>un neuf</i> 'a nine'	[œ̃.nœf]
	/t/	<i>un grand assaut</i> 'a big assault'	<i>un grand tasseau</i> 'a big bracket'
<i>un grand ami</i> 'a great friend'		<i>un grand tamis</i> 'a big sieve'	[œ̃.gʁɑ̃.ta.mi]
<i>le grand Est</i> 'the big East'		<i>le grand test</i> 'the big test'	[lə.gʁɑ̃.tɛst]
<i>un grand acte</i> 'a great act'		<i>un grand tact</i> 'a great (sense of) tact'	[œ̃.gʁɑ̃.takt]
/z/	<i>les aunages</i> 'the measurements' (by aune)	<i>les zonages</i> 'the zonings'	[le.zo.naʒ]
	<i>les ailés</i> 'the winged ones'	<i>les zélés</i> 'the zealous ones'	[le.zɛ.le]
	<i>les ailes</i> 'the wings'	<i>les zèles</i> 'the zeals'	[le.zɛl]
	<i>les aines</i> 'the groins'	<i>les Zens</i> 'the Zens'	[le.zɛn]

Due to the limitations of using real lexical items, syllable structure could not be held constant across real-word stimuli. The ambiguous pairs include the following syllabic structures, where (C) represents the onset of the consonant-initial item of each minimal pair: (C)VC, (C)VCC for one-syllable items and (C)VC.CVC, (C)V.CVC, (C)V.CV for two-syllable items.

A second list of non-words was also created. The use of non-words in addition to real lexical items is warranted given evidence that lexical frequency can affect the production of a word. Whalen (1991, 1992) demonstrated that low frequency words are produced more slowly and articulated more carefully than high frequency words. This introduces the possibility that durational differences produced by speakers do not reflect the intended segmentation of the sequence, but rather the lexical frequency of the target

word. Second, as pertains to L2 learning, the use of non-words more closely simulates L2 learning and segmentation processes in that the L2 learner is often confronted with unknown words that must be parsed out of the signal.

Finally, the use of nonsense words is important in that it could offer a clearer picture of the perceptual effects of durational differences in segments by removing the interference of top-down factors. There was some concern, for example, that the perception of the real-word targets could be affected by lexical frequency and/or syntactic and semantic acceptability within each real-word minimal pair in the behavioral tasks to follow. For example, given the real-word pair *un grand ami* ‘a great friend’ and *un grand tamis* ‘a big sieve’, both [œ̃.gʁɑ̃.ta.mi], *ami* has a considerably higher lexical frequency than *tamis*, which could potentially bias listeners in the segmentation of this ambiguous sequence. In investigating this concern, the lexical frequency of each real-word target was determined from the *Lexique 3* database (www.lexique.org; New, Pallier, Ferrand & Matos 2001). This database searches 218 literary texts published between 1950 and 2000 (14.7 million words) as well as the subtitles of 9474 films (50.4 million words). Given that this database places more focus on the subtitles of films than literary works, it is thought to give a more reliable picture of spoken French than databases based primarily on literary works or newspapers and magazines. Frequencies of the target words in this study were calculated by searching for both the singular and plural form (where applicable) of each target. The mean frequency of liaison (vowel-initial) items was 119.34 occurrences per million words, while the mean frequency of non-liaison (consonant-initial) items was 14.25 occurrences per million words. The frequencies for each individual real-word target per million words are given below in Table 5-3.

Table 5-3: *Frequency of occurrence per million words of each real lexical item in the Lexique 3 database (www.lexique.org). Mean frequency for each condition (vowel-initial and consonant-initial) is given at the bottom of the table.*

Vowel-initial (liaison) stimulus		Consonant-initial stimulus	
air	588.00	nerf	30.30
oeuf	44.77	neuf	86.32
hectare	4.65	nectar	1.55
aval	3.04	naval	6.58
Est	65.74	test	30.50
acte	57.06	tact	3.74
ami	551.05	tamis	1.18
assaut	22.14	tasseau	0.01
ailles	46.96	zèles	7.80
aines	1.71	Zens	0.87
ailés	2.12	zélés	2.08
aunages	0.00	zonages	0.05
Mean Frequency	119.34	Mean Frequency	14.25

As can be seen, target words in several of the 12 real-word pairs differ markedly in relative lexical frequency. As noted above, these differences in frequency caused concern for the perception portion of the study given that frequency could impose a bias on the perception of ambiguous pairs in the forced-choice identification task. Though Shoemaker and Birdsong (2008) did not find a significant correlation between response biases and lexical frequency in the identification of globally ambiguous phrases in that study, a large body of research has demonstrated that lexical frequency plays a significant role in the degree of activation of candidate words (see for example Cluff & Luce 1990; Luce & Pisoni 1998; Marslen-Wilson 1990).

Specifically in the case of liaison in spoken French, Bybee's (2001) exemplar model discussed above in §2.3, proposes that the processing of liaison is based largely on the frequency of co-occurrence of words in that words that occur frequently together in environments of liaison (e.g. *un ami* [œ.na.mi] 'a friend', *les amis* [le.za.mi] 'the friends') are stored in the mental lexicon as phonological chunks, much like individual lexical items are stored. Bybee's model is mainly centered on the idea of mental representations and does not make specific predictions about lexical access; however given the empirically demonstrated role of frequency in SWR processes, it is reasonable to infer that her model would predict that, for example, the phrase *un grand ami* 'a great friend' would be given preference over *un grand tamis* 'a big sieve' given the acoustic input [œ.gʁɑ̃.ta.mi].

For these reasons, a second list of non-words was created. For the non-word targets, four vowel-initial sequences with transparent orthography and conforming to French phonotactics were created for each of the three liaison consonants such that the realization of these sequences in liaison environments creates two homophonous non-word sequences. For example, the nonce sequence *épeu* [e.pø] when preceded by *un* [œ] yields phonemic content consistent with both *un épeu* and *un népeu* [œ.ne.pø]. Syllable structure was controlled for in the non-word sample and consisted of two syllabic structures, where (C) represents the consonant-initial item of each minimal pair: (C)VC for one-syllable items and (C)V.CV for two-syllable items. This resulted in a total of 24 non-word targets ranging from 2 – 4 syllables (mean 2.83 syllables) and from 4 – 8 phonemes (mean 5.83 phonemes), which are presented below in Table 5-4.

One further concern that was addressed for both word types (real and non-words) was syntactic environment. All targets in the current production sample were produced in obligatory liaison environments. This stipulation was deemed necessary given the possibility that if a listener were presented with a syntactic environment where liaison were optional or prohibited in the identification task his/her response could be biased toward a non-liaison (consonant-initial) segmentation of the sequence. Accordingly, all target phrases consisted of either determiner + noun (*un air* 'a melody'/ *un ade*) or determiner + adjective + noun (*un grand ami* 'a great friend'/ *un grand ade*) both of which represent obligatory environments of liaison.

Table 5-4: *Non-word target pairs.*

Consonant	Vowel-initial (liaison) target	Consonant-initial target	Pronunciation
/n/	<i>un auvis</i>	<i>un nauvis</i>	[œ̃.no.vi]
	<i>un épeu</i>	<i>un népeu</i>	[œ̃.ne.pø]
	<i>un upe</i>	<i>un nupe</i>	[œ̃.nyp]
	<i>un ade</i>	<i>un nade</i>	[œ̃.nad]
/t/	<i>un grand auvis</i>	<i>un grand tauvis</i>	[œ̃.gʁã.to.vi]
	<i>un grand épeu</i>	<i>un grand tépeu</i>	[œ̃.gʁã.te.pø]
	<i>un grand upe</i>	<i>un grand tupe</i>	[œ̃.gʁã.typ]
	<i>un grand ade</i>	<i>un grand tade</i>	[œ̃.gʁã.tad]
/z/	<i>les auvis</i>	<i>les zauvis</i>	[le.zo.vi]
	<i>les épeus</i>	<i>les zépeus</i>	[le.ze.pø]
	<i>les upes</i>	<i>les zupes</i>	[le.zyp]
	<i>les ades</i>	<i>les zades</i>	[le.zad]

A final consideration in the selection and formation of target pairs was the possible effect of lengthening due to stress placement. Stress accent, as discussed above in §2.4, is fixed (i.e. never lexical) in French and consistently falls on the final syllable of a word in isolation or the final syllable of a phrase. Furthermore, stress in French is most prominently signaled by duration (see for example, Delattre 1951, 1966). Recall as well that Banel and Bacri (1994) demonstrated that duration is a robust cue to whether a syllable is word-final or word-medial. For these reasons, there was some concern that the number of syllables of the target word in which the consonant under investigation appears could present an additional lengthening factor in the production process. In order to address this possibility, this factor was incorporated into the design of the experiment as an additional independent variable. Half of the target words consisted of one-syllable words and the other half consisted of two-syllable words.

5.1.3 Recording procedure

Participants first read and signed a consent form for their participation in the experiment (see Appendix I). Each of the six production participants recorded the two word lists separately — the list containing real-word phrases was read first followed by the list containing non-word phrases. The speakers read the target phrases embedded in the carrier sentence, *je vais dire un/les _____ bleu(s)* ‘I am going to say a/the blue _____’, so as to maintain intonation as constant as possible. An adjective of color was chosen for two reasons. First, given that color adjectives follow nominal items in French, lengthening effects that might occur if the target word appeared phrase-finally were

avoided. Second, the use of a post-nominal adjective allows us to contain the target word in a single intonational unit, further avoiding any lengthening effects that could occur at prosodic boundaries above the word.

Each target appeared in each list three times, giving a total of 72 real-word tokens in the first list and 72 non-word tokens in the second list (3 consonants x 8 tokens x 3 readings) for a total of 144 tokens read by each speaker. This resulted in a total of 864 tokens read by all six speakers. Each list also contained 24 distracter items interspersed throughout as well as three distracter phrases at both the beginning and end of the lists to avoid possible list effects.

Speakers read through both lists with the principal investigator before recording to ensure that they were familiar with all of the real lexical items and to ensure uniform pronunciation of the non-word items. Speakers were asked to maintain a constant speech rate to the best of their ability throughout the recording process.

Participants were recorded in a sound-attenuated booth onto a solid-state recorder at a sampling rate of 44 kHz. Recordings were then transferred in the form of .mp3 files onto a Dell Inspiron 600m laptop computer for storage and subsequent measurement and instrumental manipulation.

5.2 Acoustic Analyses

Four tokens (all /t/s; two real words and two non-words) were removed from analysis because the speaker inserted either a pause or a glottal stop before the consonant under investigation. In total measurements were thus taken from 860 tokens. All acoustic measurements were made from spectrogram and waveform displays in Praat sound-editing software (Boersma & Weenink 2007). Measurements of the segment /n/ were taken from the offset of the preceding vowel to the onset of the following vowel where there was an obvious reduction in amplitude and waveform complexity. Measurements of /t/ were taken from the beginning of the closure to the beginning of the release. Measurements of /z/ were taken where there was a clear transition in amplitude complexity and periodicity between the preceding and following vowels. Measurements were made at zero crossings wherever possible.

First, durations of all tokens were analyzed comparing segmental durations of pivotal consonants in the two Word Types (real and non-words) in order to ascertain whether this factor had an overall effect on production and thus whether real-word targets and non-word targets could be analyzed as one group in subsequent analyses. A factorial analysis of variance (ANOVA) revealed no significant difference in segmental duration between the two Word Types ($F(1, 858) = .064, NS$). Pivotal consonants appearing in real words were an average of 92.30 milliseconds (ms) and pivotal consonants appearing in non-words were an average of 92.76 ms as shown below in Figure 5-1.

Mean segmental durations and standard deviations (in parentheses) for LCs and ICs in both real-word tokens and non-word tokens are presented in Table 5-5.

Figure 5-1: Mean segmental durations in milliseconds by Word Type (real and non-word).

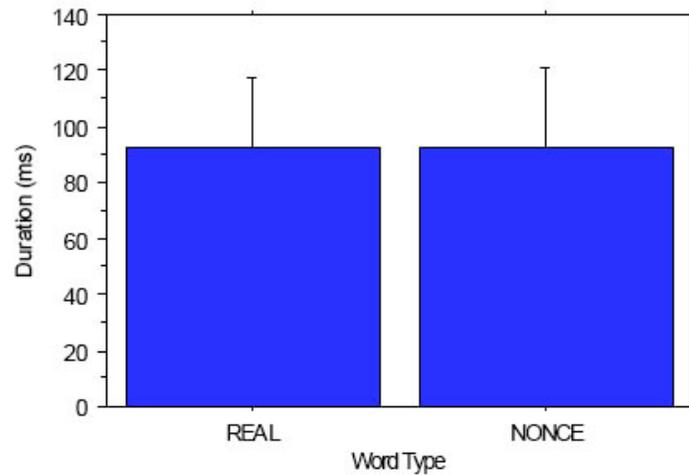
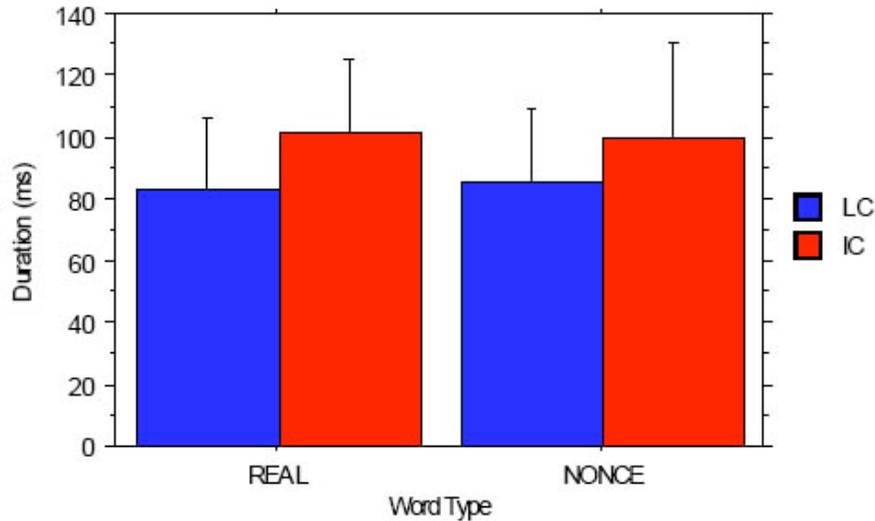


Table 5-5: Mean segmental durations in milliseconds and standard deviations (in parentheses) of liaison consonants and initial consonants in real-word and non-word tokens. Each cell represents 215 data points.

	Liaison Consonant	Initial Consonant
Real Words	83.44 (22.86)	101.08 (24.04)
Non-Words	85.35 (24.20)	100.18 (30.68)

In addition, durations were submitted to a factorial ANOVA with the factors Consonant Type (two levels: liaison consonant (LC) and initial consonant (IC)) and Word Type (two levels: real and non-word) in order to investigate whether these two factors interacted. Again, no significant difference was observed between the durations of consonants appearing in the two Word Types ($F(1, 856) = .084, NS$), though there was a significant difference between Consonant Types ($F(1, 856) = 86.278, p < .0001$). Furthermore, there was no interaction between the two factors ($F(1, 856) = .649, NS$). Mean durations are shown graphically in Figure 5-2.

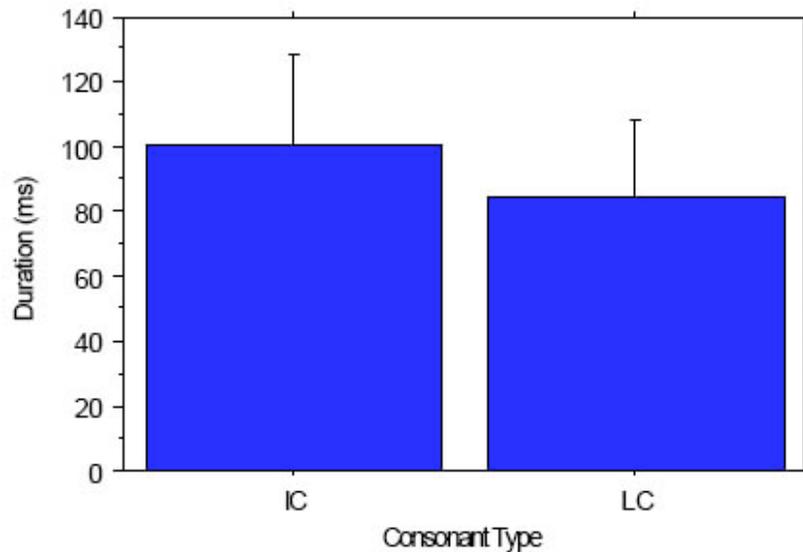
Figure 5-2: Mean segmental durations in milliseconds by Consonant Type (LC and IC) and Word Type (real and non-word).



Given that no significant differences and no interaction were observed between the segmental durations of pivotal consonants appearing in real-word tokens and non-word tokens, both were considered as one group in all subsequent analyses.

The effect of Consonant Type (LC and IC) was then examined across all six speakers, and, subsequently, by individual speaker, by consonant, and by number of syllables in the target word. An analysis of duration by Consonant Type across all tokens and speakers revealed a significant difference between LCs and ICs ($F(1, 858) = 86.39, p < .0001$). The mean duration for LCs was 84.44 ms and the mean duration for ICs was 100.66 ms as shown below in Figure 5-3.

Figure 5-3: Mean segmental durations in milliseconds of initial consonants (IC) and liaison consonants (LC).



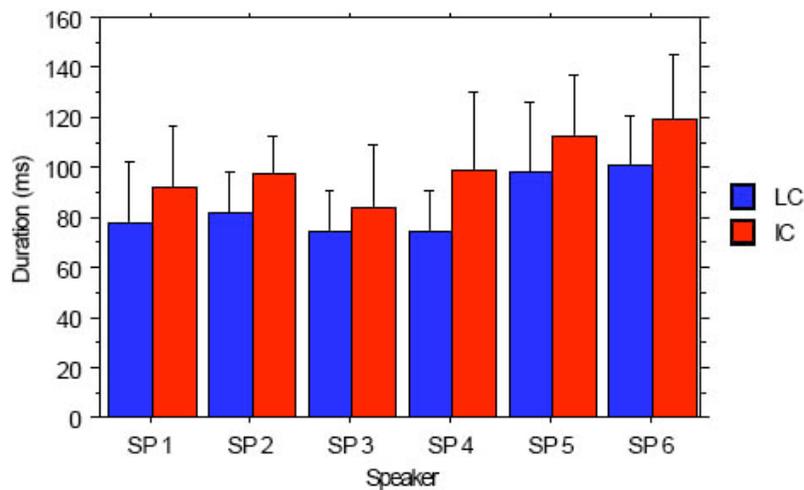
Durational values for each Consonant Type (LC and IC) were then analyzed for each of the six speakers. Table 5-6 displays mean segmental durations and SDs for each speaker for both consonant types.

Table 5-6: Mean segmental durations in milliseconds and standard deviations (in parentheses) of liaison consonants and initial consonants by speaker. Each cell represents 71-72 data points, depending on the speaker.

	Liaison Consonant	Initial Consonant
Speaker 1	77.73 (24.94)	91.92 (25.11)
Speaker 2	81.98 (16.54)	97.49 (15.41)
Speaker 3	73.78 (17.22)	84.04 (25.16)
Speaker 4	74.01 (16.96)	99.02 (31.06)
Speaker 5	98.73 (27.22)	112.68 (24.25)
Speaker 6	100.43 (20.06)	118.82 (26.35)
MEAN	84.44 (12.11)	100.66 (12.96)

A factorial ANOVA showed a main effect of Consonant Type ($F(1,848) = 106.47, p < .0001$) as well as significant differences among the six Speakers ($F(5,848) = 40.7, p < .0001$). No interaction was observed between the two factors ($F(1, 848) = 1.716, NS$). Crucially, each of the six speakers produced significantly longer consonants in initial position than in resyllabified liaison position as can be seen below in Figure 5-4.

Figure 5-4: Mean segmental durations in milliseconds by Speaker and Consonant Type (LC and IC).



We also wanted to investigate whether the number of syllables in the target word affected the duration of pivotal consonants. Table 5-7 presents mean segmental durations and SDs for both consonant types in words of one and two syllables.

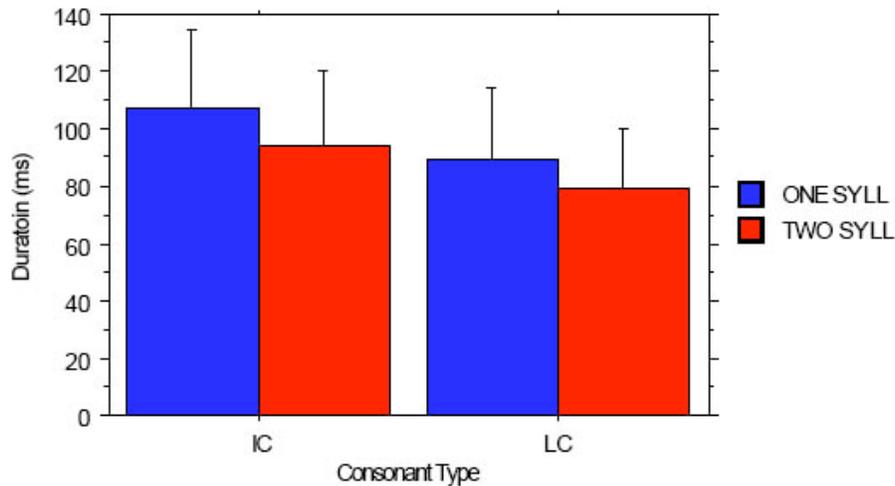
Table 5-7: Mean segmental durations in milliseconds and standard deviations (in parentheses) of liaison consonants and initial consonants in one- and two-syllable tokens. Each cell represents 215 data points.

	Liaison Consonant	Initial Consonant
One-syllable token	89.69 (25.08)	106.94 (27.74)
Two-syllable token	79.13 (20.64)	94.29 (25.86)

Durations were then submitted to analysis with the factors Consonant Type (two levels: LC, IC) and number of Syllables (two levels), which also revealed significant differences. Pivotal consonants appearing in one-syllable words were significantly longer than the same target consonants in two-syllable words ($F(1, 856) = 46.44, p < .0001$).

Furthermore, this effect was consistent for both LCs and ICs ($F(1, 856) = 90.61, p < .0001$). No interaction was found between the two factors ($F(1, 856) = .38, NS$). This analysis is shown below in Figure 5-5.

Figure 5-5: Mean segmental durations in milliseconds by Consonant Type (IC and LC) and number of Syllables.



5.2.1. Acoustic Analyses by Individual Consonant

Analyses were then conducted for each of the three segments under investigation individually. Several studies have shown that, though durational differences are systematically present at word boundaries, the robustness of the differences can vary as a function of the segment. As noted above in §2.4.2, Shoemaker (2006) found more robust differences between LCs and ICs for /t/ than for /ʁ/. In addition, Wauquier-Gravelines (1996) found significant differences for /t/ but none for /n/. Nguyen et al. (2007) also failed to find durational differences in /n/ according to word position, but found that /z/ was shorter in liaison position than in both word-initial and word-final position.

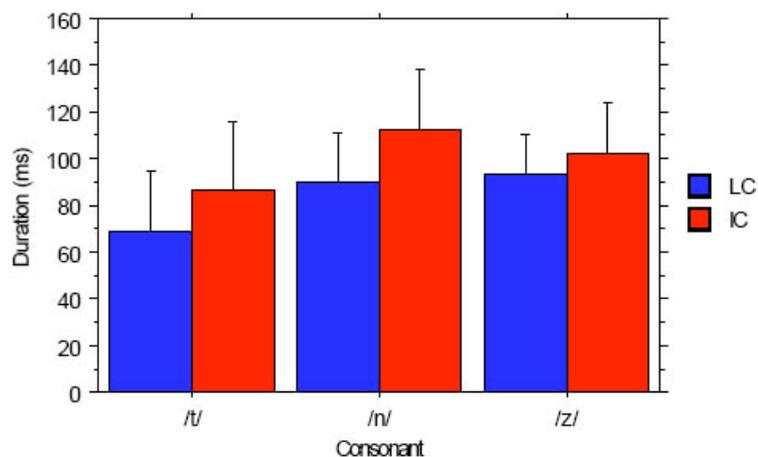
Given the variation in the existing literature on these segments, the possibility of differences in the robustness of durational differences was also addressed in the current production sample. Table 5-8 below presents mean segmental durations in milliseconds and SDs for each of the three consonants in both liaison and initial position.

Table 5-8: Mean segmental durations in milliseconds and standard deviations (in parentheses) of LCs and ICs for each of three consonants. Cells for /n/ and /z/ represent 144 data points. Cells for /t/ represent 141 and 143 data points for liaison and initial tokens, respectively.

	/n/	/t/	/z/
Liaison	86.68	69.49	93.71
Consonant	(20.78)	(25.39)	(16.12)
Initial	112.15	86.74	102.91
Consonant	(25.89)	(28.99)	(21.02)

Segmental durations were further submitted to factorial ANOVAs with the factors Consonant (three levels: n/, /t/, and /z/) and Consonant Type (two levels: IC, LC). The analysis revealed a significant main effect of both Consonant Type ($F(1, 854) = 104.42, p < .0001$) and Consonant ($F(2, 854) = 47.05, p < .0001$) as well as an interaction between the two factors ($F(2, 854) = 5.88, p = .0029$), showing that the robustness of difference in duration between these two consonant environments varies as a function of the particular segment. Post-hoc pair-wise comparisons showed that the difference between mean durations of LCs and ICs was significant for each individual segment: /n/ ($F(1, 286) = 65.98, p < .0001$), /t/ ($F(1, 282) = 26.64, p < .0001$), and /z/ ($F(1, 286) = 35.23, p < .0001$). Mean durations of LCs and ICs for each of the three consonants are shown below in Figure 5-6.

Figure 5-6: Mean segmental durations in milliseconds of three pivotal consonants by Consonant Type (IC and LC) and Consonant.



Mean segmental durations were then analyzed as a function of the three factors Speaker, Consonant and Consonant Type. Table 5-9 below shows mean segmental durations in milliseconds and SDs for each of the three consonants realized as both LCs and ICs for each of six speakers.

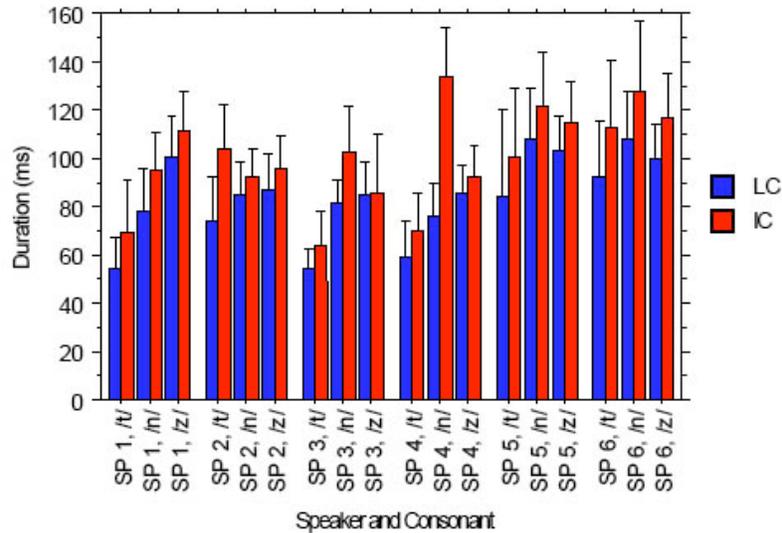
Table 5-9: Mean segmental durations in milliseconds and standard deviations (in parentheses) of liaison consonants and initial consonants for six speakers for each segment. Cells for /n/ and /z/ represent 24 data points. Cells for /t/ represent between 22 and 24 data points, depending on the speaker.

	/n/		/t/		/z/	
	Liaison Consonant	Initial Consonant	Liaison Consonant	Initial Consonant	Liaison Consonant	Initial Consonant
Sp 1	78.54 (17.16)	94.76 (15.76)	53.85 (13.62)	69.13 (21.81)	100.78 (17.04)	111.86 (16.03)
Sp 2	85.03 (13.60)	92.53 (11.76)	74.08 (18.23)	104.15 (18.20)	86.81 (15.06)	95.78 (13.72)
Sp 3	81.93 (8.73)	102.22 (19.71)	54.20 (8.03)	64.32 (14.22)	85.21 (12.99)	85.57 (24.69)
Sp 4	76.17 (13.69)	134.10 (19.89)	59.50 (14.37)	70.36 (15.83)	85.77 (11.45)	92.58 (12.12)
Sp 5	108.34 (21.12)	121.57 (22.43)	83.78 (36.25)	100.96 (28.56)	103.44 (14.44)	115.02 (16.76)
Sp 6	108.04 (19.37)	127.69 (28.73)	92.66 (23.43)	112.10 (28.86)	100.26 (14.26)	116.65 (18.76)

A factorial ANOVA revealed a main effect of Speaker ($F(5, 824) = 61.13, p < .0001$), a main effect of Consonant ($F(2, 824) = 125.01, p < .0001$) as well as a main effect of Consonant Type ($F(1, 824) = 162.39, p < .0001$). Significant interactions were also found between the factors Speaker and Consonant ($F(10, 824) = 13.49, p < .0001$), Speaker and Consonant Type ($F(5, 824) = 2.68, p = .0207$), and Consonant and Consonant Type ($F(2, 824) = 9.16, p = .0001$). There was also a significant interaction among all three factors ($F(10, 824) = 5.58, p < .0001$), reflecting the degree of variance both among speakers and among the three consonants. This analysis is shown graphically in Figure 5-7 below.

It is important to note that, though significant differences were observed among the six speakers, each speaker consistently produced durational differences between the two consonant types in the same direction for each of the three consonants — LCs were shorter than ICs.

Figure 5-7: Mean segmental durations of pivotal consonants in milliseconds by Speaker, Consonant and Consonant Type (IC and LC).



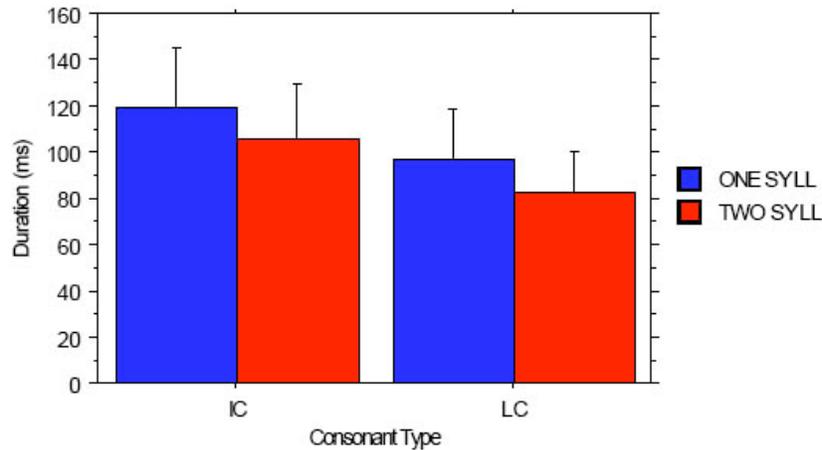
We also wanted to explore whether the number of syllables of the target word affected the durations of each individual segment. Further analyses were therefore conducted in which durations for each of the three individual consonants were submitted to analyses with the factors Consonant Type and number of Syllables. Mean segmental durations and SDs for each Consonant by Consonant Type (LC and IC) and number of Syllables are shown in Table 5-10.

Table 5-10: Mean segmental durations in milliseconds and standard deviations (in parentheses) of liaison consonants and initial consonants for each of three segments in both one- and two-syllable tokens. Cells for /n/ and /z/ represent 72 data points. Cells for /t/ represent between 70 and 72 data points.

	/n/		/t/		/z/	
	Liaison Consonant	Initial Consonant	Liaison Consonant	Initial Consonant	Liaison Consonant	Initial Consonant
One-syll Tokens	96.81 (21.28)	118.76 (26.47)	75.86 (29.77)	94.41 (29.74)	96.00 (17.09)	107.65 (21.05)
Two-syll Tokens	82.54 (17.71)	105.53 (23.67)	63.21 (18.33)	78.96 (26.19)	91.42 (14.86)	98.17 (20.02)

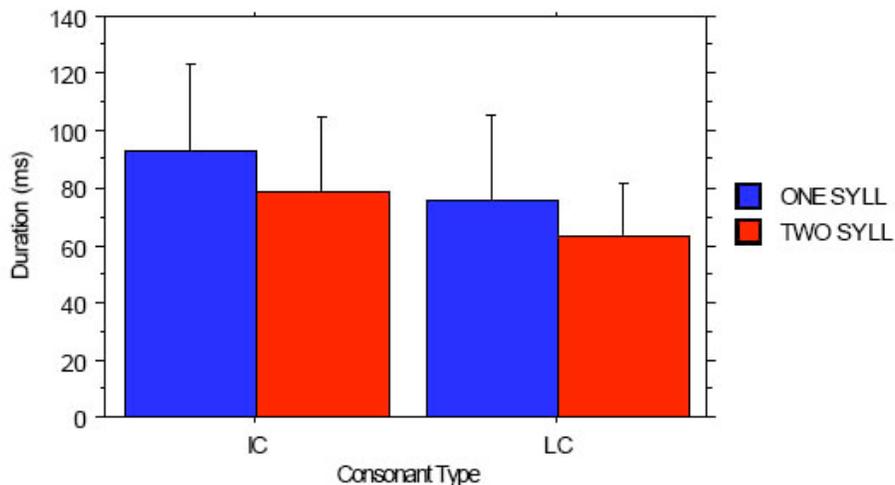
We first analyzed only those tokens containing /n/, which revealed a significant difference between one- and two-syllable words in both Consonant Types. There was a main effect of Consonant Type ($F(1, 284) = 71.72, p < .0001$) as well as a main effect of number of Syllables ($F(1, 284) = 26.83, p < .0001$), but no interaction between the two ($F(1, 284) = .039, NS$). These results are displayed in Figure 5-8.

Figure 5-8: Mean segmental durations in milliseconds of tokens containing /n/ by Consonant Type (IC and LC) and number of Syllables.



Analyses were then conducted for only those tokens containing /t/, which revealed a significant difference between one- and two-syllable words for both Consonant Types. There was a main effect of Consonant Type ($F(1, 280) = 27.88, p < .0001$) as well as a main effect of number of Syllables ($F(1, 280) = 18.45, p < .0001$), but no interaction between the two ($F(1, 280) = .074, NS$) as shown below in Figure 5-9.

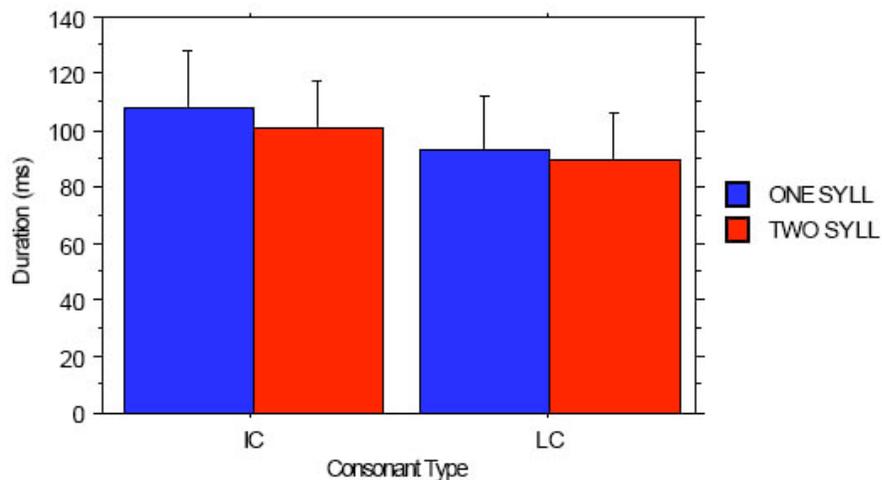
Figure 5-9: Mean segmental durations in milliseconds of tokens containing /t/ by Consonant Type (IC and LC) and number of Syllables.



Finally, analyses were conducted for only those tokens containing /z/, which also revealed a significant difference between one- and two-syllable tokens. An ANOVA revealed a main effect of Consonant Type ($F(1, 284) = 36.02, p < .0001$) as well as a main effect of number of Syllables ($F(1, 284) = 7.70, p = .0059$), but no interaction between the two ($F(1, 284) = .73, NS$). These results are presented in Figure 5-10 below.

Pivotal consonants appearing in one-syllable words were significantly longer than the same consonant appearing in two-syllable words for both LCs and ICs for each of the three consonants, indicating that factors other than the presence or absence of liaison affect the duration of pivotal segments. It should be noted however that the effect of number of syllables on segmental duration was somewhat less robust for /z/ than for both /n/ and /t/.

Figure 5-10: Mean segmental durations in milliseconds of tokens containing /z/ by Consonant Type (IC and LC) and number of Syllables.



5.2.2 Discussion

Though the majority of the classical literature on spoken French has maintained that consonants are identical at the acoustic level whether they surface in liaison or appear as lexical word-initial (e.g. Encrevé 1988; Grammont 1960), the current results are in line with more recent research that has shown that there are systematic durational differences in consonants in liaison environments and their lexical-word-initial counterparts (e.g. Douchez & Lancia 2008; Nguyen et al. 2007; Shoemaker 2006; Spinelli et al. 2003; Wauquier-Gravelines 1996). Consonants that surface in environments of liaison have been shown to be systematically shorter than the same consonant in initial position.

The current results are also consistent with production studies dealing with *enchaînement* in spoken French, which have shown that fixed coda consonants that are resyllabified as the onset of a following vowel-initial word are shorter than the same

consonant when produced as lexical-word initial (Fougeron et al. 2002, 2003; Fougeron 2007). This body of work suggests that durational differences between segments that are attributable to their position in a prosodic hierarchy are not completely neutralized by resyllabification.

Concerning the saliency of differences for each of the three consonants, the current study found the most robust durational differences for the segment /n/, which showed an average difference of 25.47 ms between LCs and ICs. The segment /t/ was an average of 17.25 ms shorter in liaison environments than in initial position, while the segment /z/ was an average of 9.2 ms shorter in liaison environments than in initial position. Regarding the segment /n/, this finding is contra previous studies which have failed to find significant differences for /n/ between LCs and ICs (Nguyen et al. 2007; Wauquier-Gravelines 1996). Though Spinelli et al. (2003) reported that LCs were significantly shorter than ICs for each of the five segments investigated in that study (/n,t,ʁ,g,p/), unfortunately, these authors did not report mean differences for individual consonants, so we are unable to draw specific comparisons with segmental measurements of /n/ in relation to other segments. At present, we are unable to account for the discrepancy between the current production sample and previous research on the segment /n/. The difference could be attributable to the phonetic content of the particular lexical items used in each study, which may have differed significantly in syllabic structure, preceding and following vowels, or other factors; however, a comprehensive analysis of the stimulus sets used in all of the above-mentioned studies would be needed to make this claim.

Regarding the segment /t/, which showed a mean difference of 17.25 ms between LCs (mean 69.49 ms) and ICs (mean 86.74 ms), durational values from the current study are in line with previous work. Wauquier-Gravelines (1996) found that /t/ showed an average closure duration of 50 ms in liaison position and 70 ms in initial position. Shoemaker (2006) also found that /t/ produced in liaison contexts (mean 49.95 ms) were an average of 23.78 ms shorter than word-initial /t/ (mean 73.73 ms).

Comparisons with existing literature for the segment /z/ are difficult. Spinelli et al. (2003) did not include /z/ in their stimulus set. Gaskell et al. (2002) reported significant durational differences between LCs and ICs for /t/, /ʁ/, and /z/, but did not report measurements for individual segments. Douchez and Lancia (2008) found that /z/ was shorter in liaison environments, however recall that the values reported in this study included parts of the preceding and following vowels, and thus comparisons are not possible. Recall as well Nguyen et al. (2007) failed to find systematic durational differences between LCs and ICs for /z/.

The current production sample showed an average durational difference between LCs and ICs of 9.2 ms for the segment /z/, which is the least robust of the three consonants studied here. It is also worth noting that in Shoemaker's (2006) production study, durational differences for the segment /ʁ/, which is also a voiced fricative that surfaces in liaison environments, were also less robust than differences found for /t/. The production sample in that study, based on four native French speakers, showed that /ʁ/ in

liaison contexts were 11.96 ms shorter than word-initial /ʁ/. These two results taken together suggest that durational variation due to the presence or absence of liaison may be less robust in fricatives. This finding is also in line with the work of Fougeron (2001), who found that the segment /s/ in spoken French showed less variation at prosodic boundaries than the other consonants investigated in that study, namely /k,l,n,t/. More recently, Fougeron (2007) examined environments of *enchaînement* and found differing results for the two fricative segments investigated in that study. The segment /s/ did not vary significantly in duration when produced as a resyllabified (*enchaînée*) consonant or as word-initial, while /ʁ/ followed the same pattern as /t/ and /k/. These latter three segments exhibited patterns found in previous research, namely they were longer when produced in word-initial position than when produced as word-final, resyllabified segments. This work offers further evidence of inconsistent variation at word boundaries in fricative segments.

In sum, durations from the current production sample offer further evidence that native speakers of French mark the presence or absence of liaison through the systematic variation of segmental duration of ICs and LCs. However, the degree of variance among the three consonants coupled with mixed evidence from previous research regarding these particular segments suggests that these differences are not only subtle, but may be inconsistently produced in natural (as opposed to laboratory) speech. Note as well that other factors may contribute to the lengthening of these segments at word boundaries. For example, pivotal consonants in one-syllable words were longer than pivotal consonants in two-syllable words, introducing further variation into the signal.

5.3 Post-Hoc Acoustic Analyses

The existing literature on acoustic-phonetic cues to word juncture motivated the measurement of additional acoustic parameters in the current production sample as multiple acoustic cues have been shown to vary at word and syllable boundaries (e.g. Lehiste 1960; Nakatani & Dukes 1977). Though the current study focuses on the durational variation of pivotal consonants at word boundaries, several additional analyses were conducted post-hoc in order to explore acoustic information other than the duration of pivotal consonants that could also signal the presence or absence of liaison to listeners.

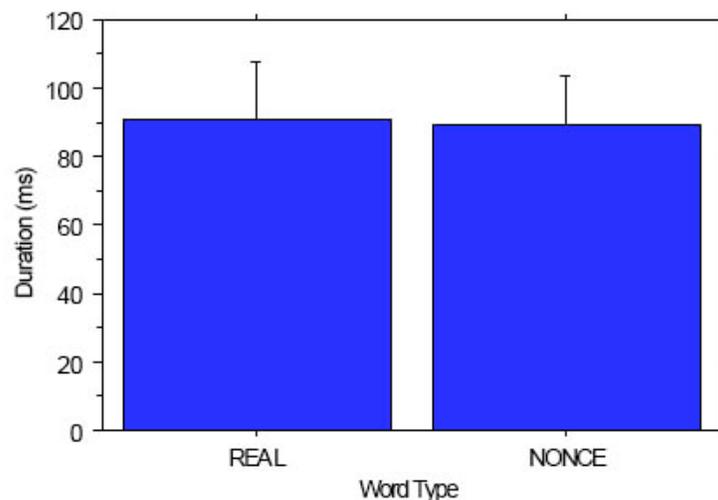
5.3.1 Vowel Duration

In addition to reporting segmental durations of the pivotal consonants in globally ambiguous phrases, Spinelli et al. (2003) also reported segmental durations for vowels preceding pivotal consonants (hereafter V_1) as well as vowels following pivotal consonants (hereafter V_2). These authors found that V_1 s appearing before LCs (mean 97 ms) were an average of 3% shorter than V_1 s appearing before ICs (mean 100 ms). These authors did not, however, find significant durational differences in V_2 s. Nguyen et al. (2007), however, found no significant durational differences in their production sample between V_1 s in liaison environments and V_1 s in non-liaison environments. (Nguyen et al. did not report values of V_2 s.) The possibility of significant differences in duration in both vocalic positions was also addressed in the current production sample.

5.3.1.1 Duration of V₁

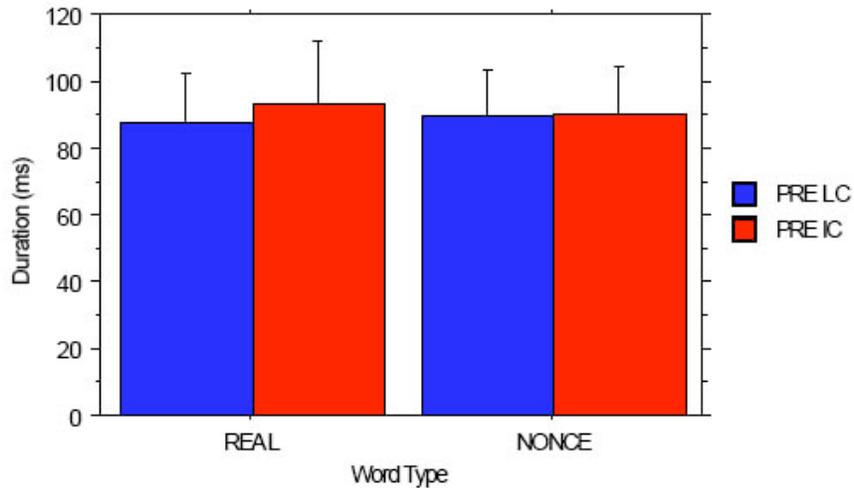
Mean durations of V₁ were analyzed first. As with other acoustic analyses, we first wanted to determine whether there was a significant difference between vowels produced in real-word tokens and vowels produced in non-word tokens so as to ascertain whether the two word types could be treated as one group in subsequent analyses. Mean segmental durations for V₁ were thus submitted to factorial ANOVAs with the factor Word Type (two levels: real word and non-word). This analysis revealed no significant difference between the mean durations of V₁ in real words (mean 90.33 ms, SD 17.38) and in non-words (mean 89.54 ms, SD 14.22; $F(1, 863) = .529$, NS). Mean durations in each condition are shown below in Figure 5-11.

Figure 5-11: Mean segmental durations in milliseconds of vowels preceding pivotal consonants in real-word and non-word tokens.



In addition, values were submitted to a factorial ANOVA with the factors Word Type and Vowel Condition (two levels: vowel preceding a liaison consonant, henceforth pre-LC, and vowel preceding initial consonant, henceforth pre-IC). This analysis also failed to show a significant effect for Word Type ($F(1, 857) = .532$, NS) but did reveal a main effect of Vowel Condition ($F(1, 857) = 10.570$, $p = .0012$). Furthermore, a significant interaction between the two factors was observed ($F(1, 857) = 5.980$, $p = .0147$). These results are shown in Figure 5-12 below.

Figure 5-12: Mean duration in milliseconds of vowels preceding pivotal consonants by Word Type (real and non-word) and Vowel Condition (Pre IC and Pre LC).



Specific comparisons revealed the source of the observed interaction; the difference in segmental duration between V_1 s appearing pre-LC and V_1 s appearing pre-IC was significant for real-word tokens only ($F(1, 429) = 13.728, p = .0002$; see Figure 5-13 below). The difference in segmental duration did not reach significance for non-word tokens ($F(1, 429) = .397, NS$; see Figure 5-14 below). For this reason, subsequent analyses of V_1 durations were conducted separately for real-word tokens and non-word tokens.

Figure 5-13: Mean segmental durations in milliseconds of vowels preceding a liaison consonant (Pre LC) and vowels preceding an initial consonant (Pre IC) in real-word tokens.

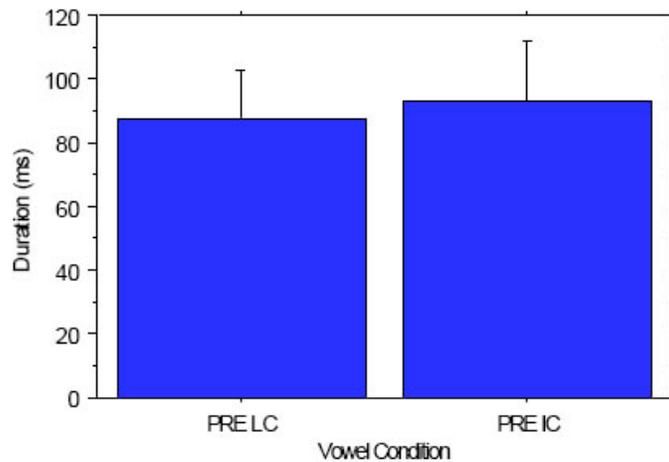
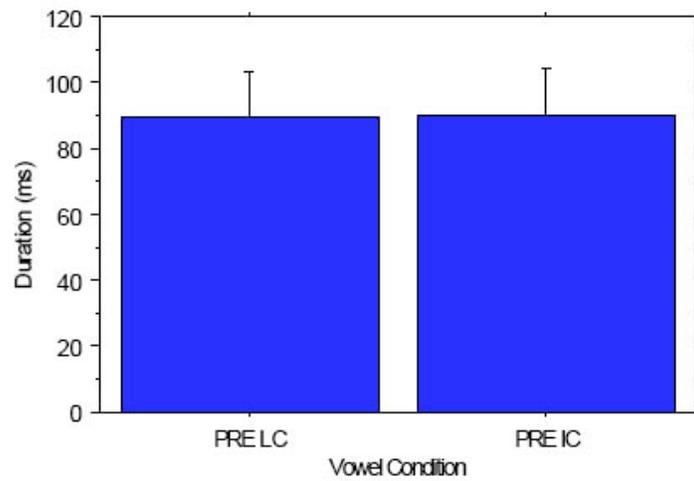
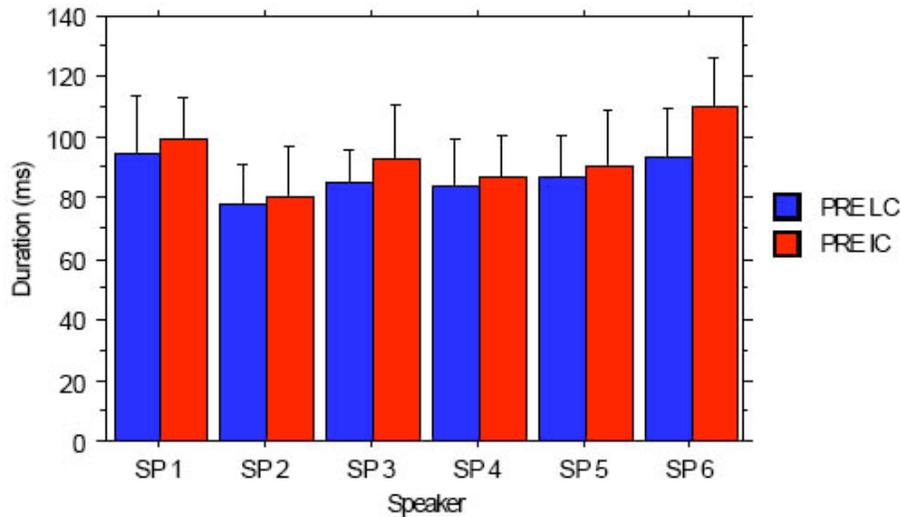


Figure 5-14: Mean segmental durations in milliseconds of vowels preceding a liaison consonant (Pre LC) and vowels preceding an initial consonant (Pre IC) in non-word tokens.



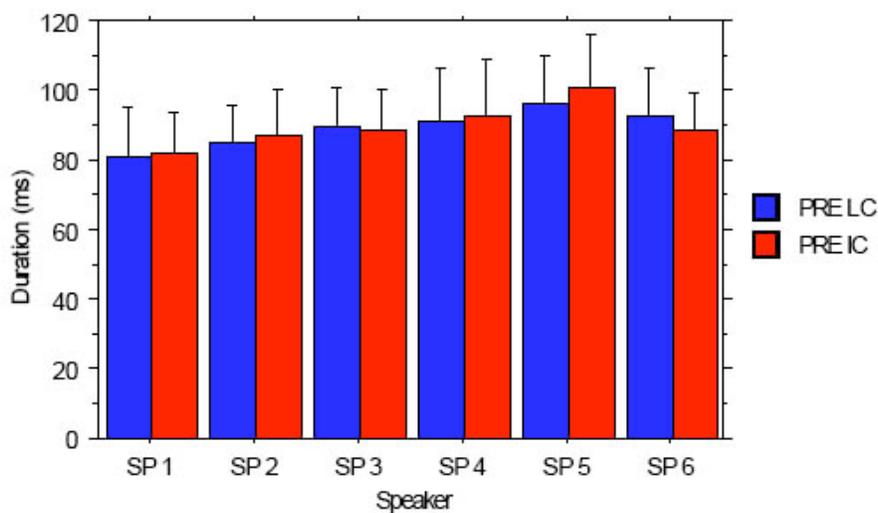
Durations of V_1 for each individual speaker were also examined, which revealed a significant amount of inter-speaker variability. For real-word tokens, there was a main effect of Speaker (six levels; $F(5, 419) = 20.458, p < .0001$) as well as an effect for Vowel Condition (two levels: pre-LC and pre-IC; $F(1, 419) = 17.294, p < .0001$), and a significant interaction ($F(5, 419) = 2.351, p = .0402$) indicating that the durational variation in V_1 depends on the speaker. See Figure 5-15 below. Of particular note is the fact that the durational differences between pre-LC and pre-IC conditions were in the same direction for each of the six speakers — vowels preceding LCs were shorter than vowels preceding ICs though this difference did not reach significance for each speaker.

Figure 5-15: Mean segmental duration of V_1 in milliseconds by Speaker and Vowel Condition (Pre LC and Pre IC) in real-word tokens.



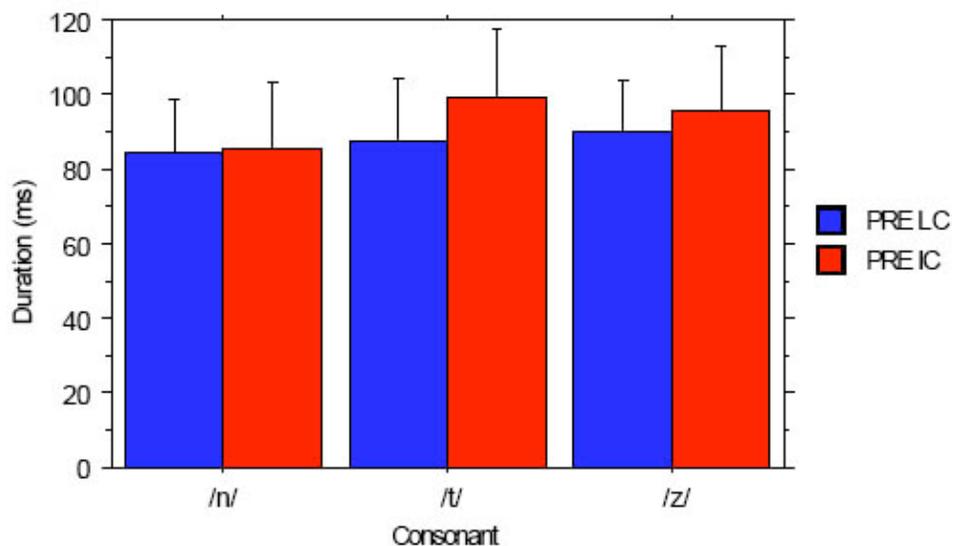
For the non-word tokens, there was also a main effect of Speaker (six levels; $F(5, 418) = 12.945$, $p < .0001$), but no effect for Vowel Condition (two levels: pre-LC and pre-IC; $F(1, 418) = .535$, NS), and no significant interaction ($F(5, 418) = .678$, NS). See Figure 5-16 below. Significant differences were observed among the six speakers, but no significant differences in duration were observed between the two V_1 conditions.

Figure 5-16: Mean segmental durations of V_1 in milliseconds by Speaker and Vowel Condition (Pre LC and Pre IC) in non-word tokens.



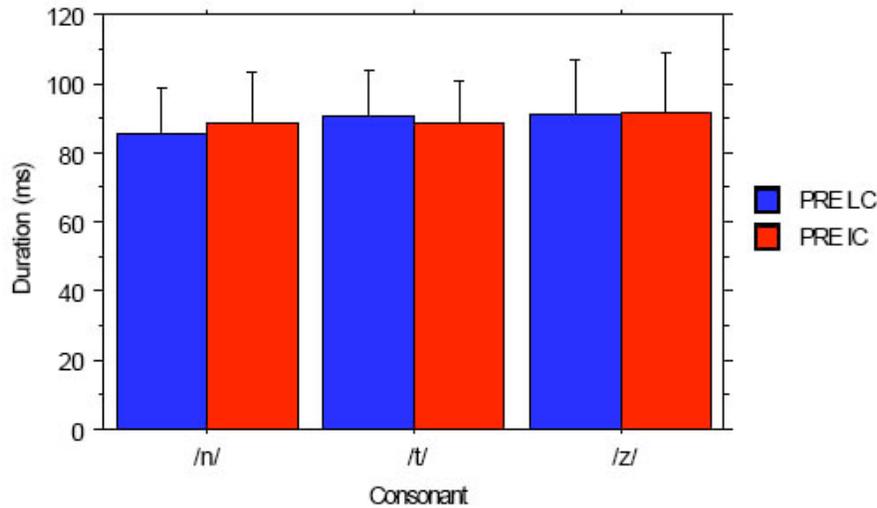
We also considered the possibility that V_1 durations could be affected by the particular consonants that they preceded, thus mean segmental durations for V_1 were also submitted to ANOVAs with the factors Consonant (three levels: /n/, /t/, /z/) and Vowel Condition (two levels: pre-LC and pre-IC). For the real-word tokens, this analysis revealed a significant main effect of Consonant ($F(2,425) = 11.118, p < .0001$) as well as a significant main effect of Vowel Condition ($F(1,425) = 14.485, p = .0002$). A significant interaction was also observed ($F(2,425) = 3.723, p = .0249$) indicating that the duration of V_1 also depends on the consonant it precedes. For all three segments, V_1 s preceding LCs were shorter than V_1 s preceding ICs. Figure 5-17 below shows mean V_1 durations for each consonant in each condition.

Figure 5-17: Mean segmental durations of V_1 in milliseconds by Vowel Condition (Pre LC and Pre IC) and Consonant in real-word tokens.



An effect of Vowel Condition was not observed for the non-word tokens ($F(1,424) = .411, NS$), however. A significant main effect was observed for Consonant ($F(2,424) = 3.405, p = .0341$). There was no significant interaction between the two factors ($F(2,424) = 1.036, NS$). Figure 5-18 below shows mean V_1 durations for each consonant in each condition for non-word tokens.

Figure 5-18: Mean segmental durations of V_1 in milliseconds by Vowel Condition (Pre LC and Pre IC) and Consonant in non-word tokens.



5.3.1.2 Duration of V_2

The possible effect of the presence or absence of liaison on the duration of post-boundary vowels (V_2) was also analyzed. First, as in previous analyses, V_2 produced in real-word targets and V_2 produced in non-word targets were compared so as to ascertain whether the two groups could be treated as one group in subsequent analyses. Mean segmental durations for V_2 were therefore submitted to factorial ANOVAs with the factor Word Type (two levels: real words and non-words). This analysis revealed a significant difference between the two types of words ($F(1, 863) = 18.802, p < .0001$). Mean durations of V_2 s in real-tokens were 93.62 ms (SD 27.60), while mean durations of V_2 s in non-tokens were 85.13 ms (SD 29.84). These means are shown below in Figure 5-19.

Durational values were subsequently submitted to a factorial ANOVA with the factors Word Type and Vowel Condition (two levels: vowel following a liaison consonant, hereafter post-LC, and vowel following an initial consonant, hereafter post-IC), which revealed a significant effect for Word Type ($F(1, 857) = 18.771, p < .0001$). However, this analysis failed to reveal either a main effect of Vowel Condition ($F(1, 857) = .004, NS$) or a significant interaction ($F(1, 857) = .110, NS$). These results are shown in Figure 5-20 below.

Figure 5-19: Mean segmental durations in milliseconds of vowels following pivotal consonants in both real-word and non-word tokens.

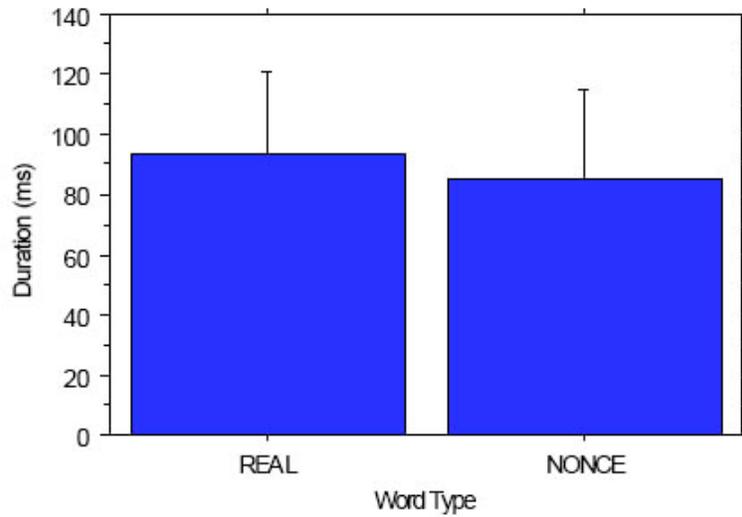
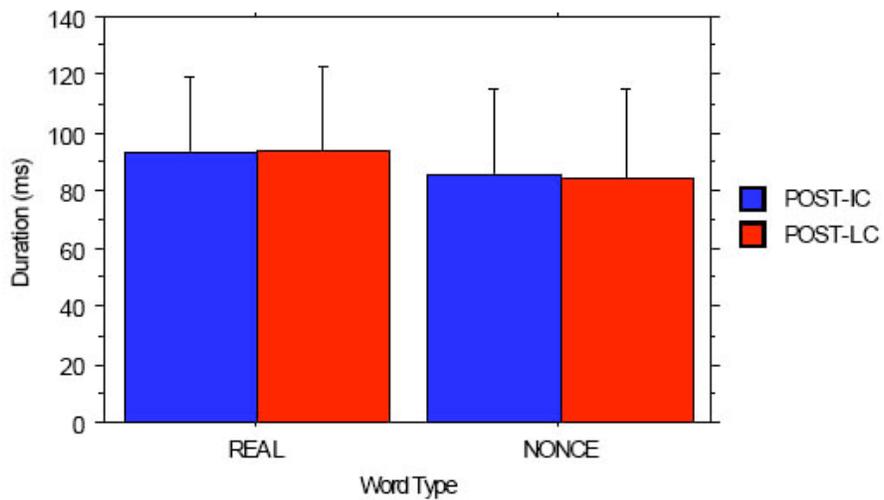


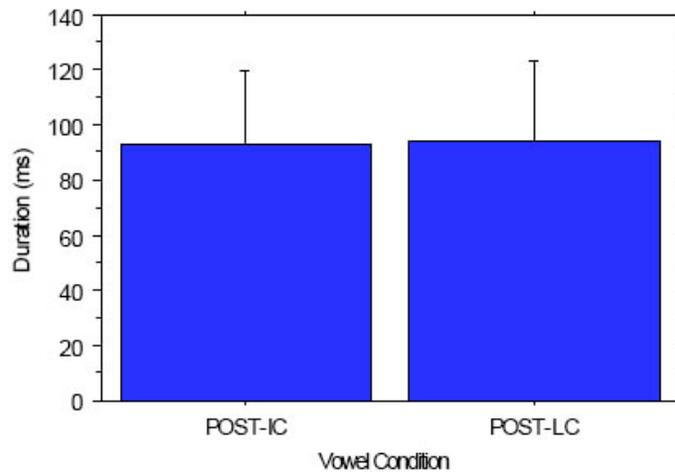
Figure 5-20: Mean segmental durations of vowels in milliseconds preceding pivotal consonants by Word Type (real word and non-word) and Vowel Condition (Post IC and Post LC).



Given the differences observed in V_2 durations between real- and non-word tokens, subsequent analyses of V_2 were conducted separately for the two word types.

Durations of V_2 in real-word tokens were then analyzed according to Condition (two levels: post-LC and post-IC), which did not reveal a significant effect ($F(1, 429) = .038$, NS). This analysis is displayed below in Figure 5-21.

Figure 5-21: Mean segmental durations in milliseconds of vowels following pivotal consonants in real-word tokens by Vowel Condition (Post IC and Post LC).



Durations of V_2 in non-word tokens were then analyzed, which also failed to reveal a significant main effect of Vowel Condition ($F(1, 428) = .073$, NS). This analysis is displayed below in Figure 5-22.

Durations of V_2 for each individual speaker for real-word tokens were then examined. This analysis revealed a significant amount of inter-speaker variability, as was observed in previous analyses. A main effect of Speaker (six levels; $F(5, 419) = 7.300$, $p < .0001$) was observed, but this analysis failed to show either an effect of Vowel Condition (two levels: post-LC and post-IC; $F(1, 419) = .036$, NS) or a significant interaction between the two factors ($F(5, 419) = .920$, NS). See Figure 5-23 below. Significant differences were found among the speakers in durations of V_2 , however no significant differences in duration were observed between post-LC vowels and post-IC vowels.

Figure 5-22: Mean segmental durations in milliseconds of vowels following pivotal consonants in non-word tokens by Vowel Condition (Post IC and Post LC).

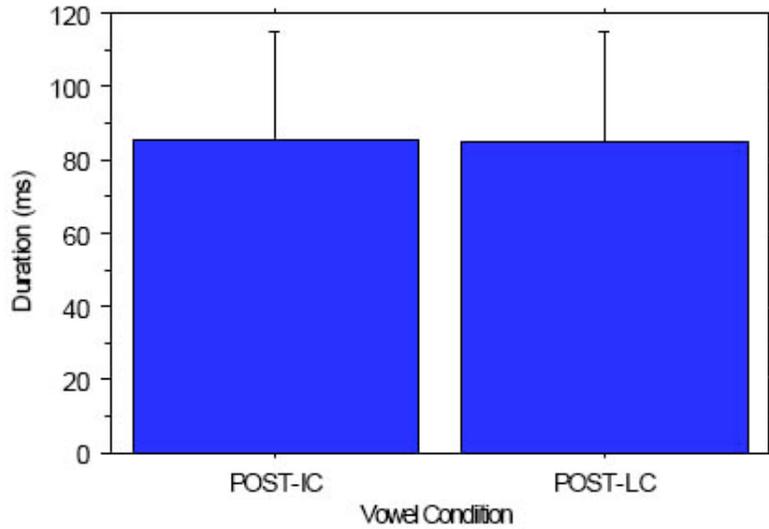
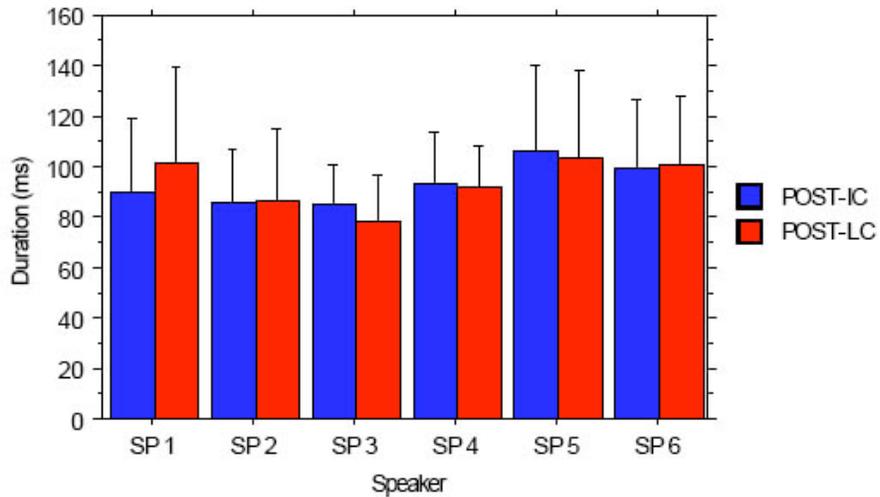
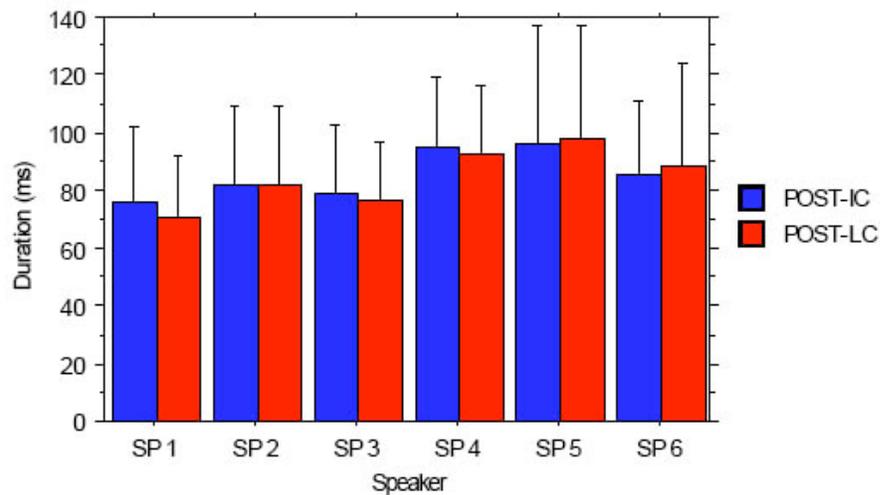


Figure 5-23: Mean segmental durations of vowels following pivotal consonants in milliseconds by Speaker and Vowel Condition (Post IC and Post LC) in real-word tokens.



Analyses of the durations of V_2 for each individual speaker for were then conducted non-word tokens. A main effect of Speaker was again observed ($F(5, 418) = 7.180, p < .0001$). No significant effect of Vowel Condition was found ($F(1, 418) = .106, NS$), nor was there a significant interaction between the two factors ($F(5, 418) = .202, NS$) as shown below in Figure 5-24. Again, a large degree of inter-speaker variability was observed, but no significant differences in duration emerged between post-LC vowels and post-IC vowels.

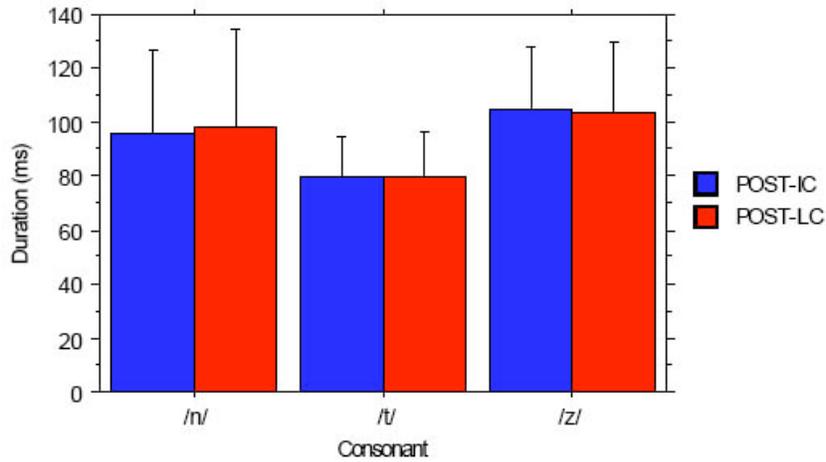
Figure 5-24: Mean segmental durations of vowels following pivotal consonants in milliseconds by Speaker and Vowel Condition (Post IC and Post LC) in non-word tokens.



We also considered the possibility that V_2 durations could be sensitive to the preceding consonants, thus mean segmental durations for V_2 s were also submitted to ANOVAs with the factors Consonant (three levels: /n/, /t/, /z/) and Vowel Condition (two levels: post-LC and post-IC).

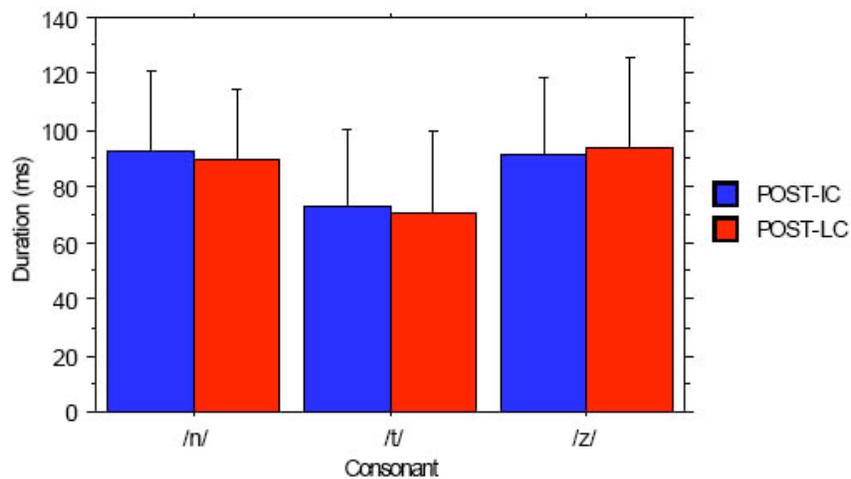
For the real-word tokens, this analysis revealed a significant main effect of Consonant ($F(2,425) = 34.675, p < .0001$), but no effect of Vowel Condition ($F(1,425) = .052, NS$) and no interaction ($F(2,425) = .208, NS$) as shown below in Figure 5-25.

Figure 5-25: Mean segmental durations of vowels following pivotal consonant in milliseconds by Vowel Condition (Post IC and Post LC) and Consonant in real-word tokens.



For the non-word tokens, a significant main effect was again observed for Consonant ($F(2,424) = 24.114, p < .0001$), but no effect of Vowel Condition was found ($F(1,424) = .090$ NS) and no interaction ($F(2,424) = .438$, NS) as shown below in Figure 5-26.

Figure 5-26: Mean segmental durations of vowels following pivotal consonant in milliseconds by Vowel Condition (Post IC and Post LC) and Consonant in non-word tokens.



5.3.1.3 Discussion

The current results concerning vocalic duration in liaison environments are consistent with those of Spinelli et al. (2003). These authors found that pre-LC vowels were shortened by 3% relative to pre-IC vowels. The current production sample shows a 3.82% shortening of pre-LC vowels relative to pre-IC vowels, though this difference was significant for real-word tokens only. In addition, Spinelli et al. found no significant durational differences in vowels following pivotal consonants. This result was also replicated in the current production sample.

It should be noted, however, that due to the nature of the recorded tokens in this sample, V₁ was held constant for each of the three consonants in both real- and non-word targets. The segment /n/ was always preceded by /œ/ as all sequences in this set began with the singular masculine indefinite article *un* /œ/. Likewise, the segment /t/ was always preceded by /ã/ since sequences in this group included the adjective *grand* ‘big’ /gʁã/. Finally the segment /z/ was always preceded by /e/ since all sequences in this group began with the plural definite article *les* /le/.

V₂ on the other hand varied within each consonant set and included tokens of /a, e, ε, o, œ, y/. This could account for the fact that more variation (and subsequently a lack of significant durational differences) was observed in productions of V₂ than in those of V₁.

Regarding the possibility that vowel duration could serve as a reliable perceptual cue to the presence or absence of liaison in spoken French, the relatively small degree of variation (3%) found in pre-boundary vowels makes this particular segment an unlikely candidate as a robust acoustic cue to liaison. However, more research would be needed to explore this possibility.

5.3.2 Voice Onset Time

We also wanted to investigate whether the presence or absence of liaison gives rise to significant differences in voice onset time (VOT) in the plosive /t/. Dejean de la Bâtie (1993) found that this segment had a longer occlusion as well as a longer VOT when produced as an IC than when produced as an LC. Wauquier-Gravelines (1996) found the same pattern; VOTs in /t/s that surface in liaison in that study were shorter than VOTs in initial /t/s.

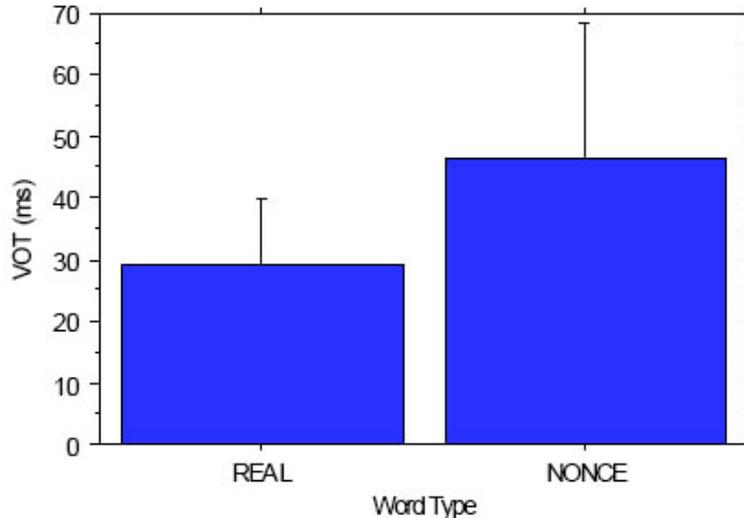
Durational values of the segment /t/ in the current production sample (reported above in §5.2) were taken only from the occlusion portion of the segment, and thus did not include VOT values. To investigate the possibility the VOT could also be sensitive to environments of liaison in the current sample, VOT values of both real- and non-word tokens containing the segment /t/ were analyzed. Mean VOTs for each /t/ token are summarized below in Table 5-11.

Table 5-11: Mean voice onset time in milliseconds for targets containing /t/. Each cell represents between 15-18 data points.

IC real target	VOT	LC real target	VOT	IC non-word target	VOT	LC non-word target	VOT
grand test	29.80	grand Est	34.04	grand tupe	76.77	grand upe	79.35
grand tact	22.72	grand acte	25.53	grand tade	28.17	grand ade	28.92
grand tamis	31.74	grand ami	37.58	grand tauvis	40.66	grand auvis	37.68
grand tasseau	25.56	grand assaut	26.27	grand tépeu	38.82	grand épeu	39.77
Mean VOT (SD)	27.45 (8.62)	Mean VOT (SD)	30.85 (12.67)	Mean VOT (SD)	46.45 (22.22)	Mean VOT (SD)	46.53 (22.41)

First, as with previous analyses, we wanted to investigate whether there was a difference between real-word and non-word stimuli. A factorial ANOVA revealed significant differences between these two Word Types ($F(1, 281) = 68.330, p < .0001$). The mean VOT for real-word targets was 29.06 ms (SD 10.83) and for non-word targets was 46.31 ms (SD 22.26). VOTs in non-word stimuli were significantly longer than in real-word stimuli. These results are shown in Figure 5-27.

Figure 5-27: Mean voice onset time in milliseconds by Word Type (real word and non-word) for targets containing /t/.

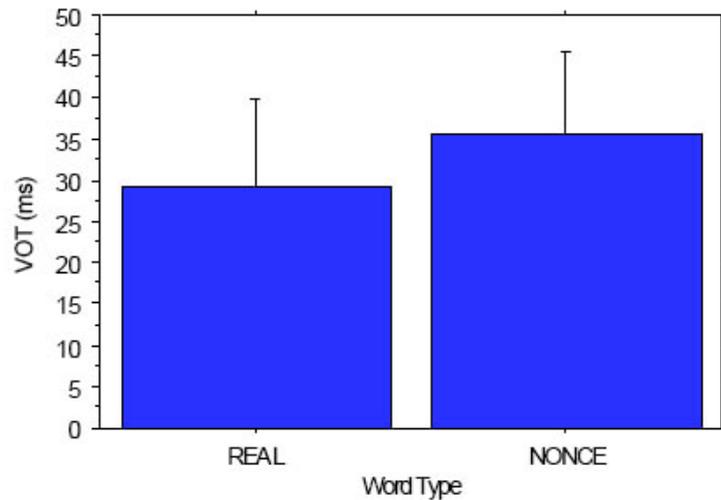


Closer inspection of non-word tokens containing /t/ revealed that one pair (*tupe/upe*) had VOT values more than twice those of other tokens, as can be seen in Table 5-11 above. This difference can be attributed to the fact that this is the only target pair in which the pivotal consonant precedes a high vowel, /y/, which has been shown to increase the VOT of plosive segments (Klatt 1975).

We therefore considered the possibility that this single pair of tokens may have skewed the comparison of real and non-word items. This pair was removed from analysis

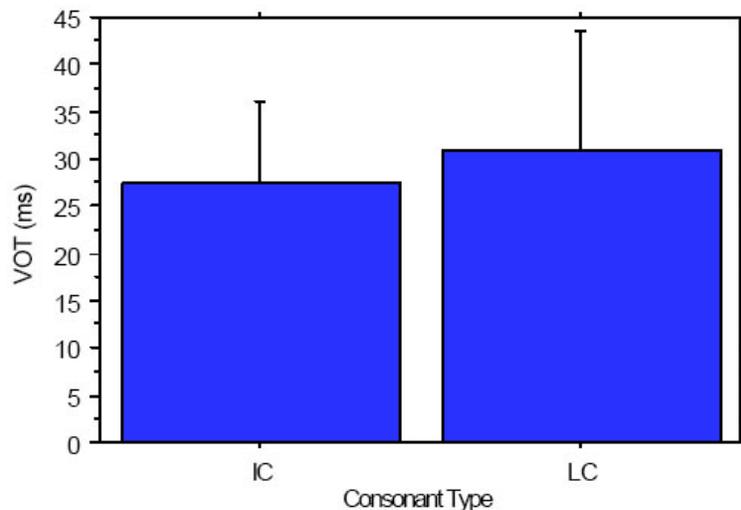
and a second factorial ANOVA was conducted. This second analysis reduced mean VOT values for non-word targets to 35.63 ms (SD 9.86), but continued to show a significant difference between VOT values in real-word tokens and in non-word tokens ($F(1, 245) = 24.158, p < .0001$) as shown below in Figure 5-28.

Figure 5-28: Mean voice onset time in milliseconds by Word Type (real words and non-words) for targets containing /t/ for all targets excluding the non-word target pair tupe/upe.



Taking into account the observed difference between real- and non-word tokens, subsequent analyses of VOT values were conducted on each of the two word types separately. Mean segmental durations for real-word stimuli including /t/ were submitted to a factorial ANOVA examining the factor Consonant Type (two levels: LC and IC), which fell just short of significance ($F(1, 141) = 3.450, p = .0654$). For real-word stimuli, mean VOT of /t/ in liaison position was 30.85 ms (SD 12.67), while mean VOT of /t/ in initial position was 27.45 ms (SD 8.62) as shown below in Figure 5-29. It should be noted as well that, not only did the difference between LCs and ICs in real-word tokens just miss significance, but this difference is also in the opposite direction of what has been previously demonstrated in the literature (Dejean de la Bâtie 1993; Wauquier-Gravelines 1996); VOT in the current sample was shorter in ICs than in LCs.

Figure 5-29: Mean voice onset time in milliseconds by Consonant Type in real-word targets containing /t/.



VOT values for LCs and ICs tokens of /t/ in real-word stimuli were then examined for each individual speaker. This analysis showed a main effect of Speaker ($F(1, 127) = 10.909, p < .0001$) as well as an effect of Consonant Type ($F(5, 127) = 6.689, p = .0108$) and a significant interaction between the two factors ($F(5, 127) = 3.239, p = .0087$). Note that, though a significant effect of Consonant Type was observed in this analysis, not all speakers showed durational differences between VOTs in LCs and ICs in the same direction; two out of six speakers showed VOT values for ICs that were superior to those of VOT in LCs, while four speakers showed VOT values for LCs that were superior to those of VOT in ICs. Mean VOTs for each speaker in each consonant type are shown below in Figure 5-30.

VOT values for non-word tokens were then examined. Again, productions of the pair *upe/tupe* were removed from analysis. Results of an ANOVA revealed no effect for Consonant Type ($F(1, 105) = .069, NS$) as shown in Figure 5-31 below. The mean VOT for non-word stimuli containing /t/ in liaison position was 35.38 ms (SD 9.25). The mean VOT for non-word stimuli containing /t/ in initial position was 35.88 ms (SD 10.50).

Figure 5-30: Mean voice onset time in milliseconds by Consonant Type (IC and LC) and Speaker in real-word targets containing /t/.

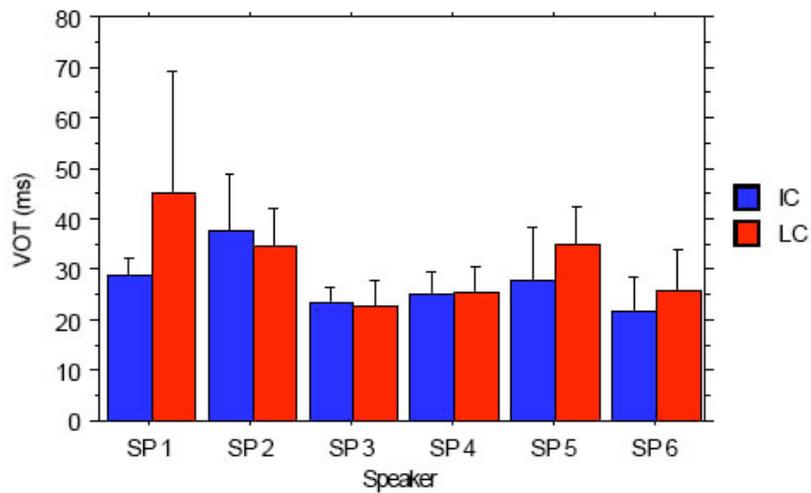
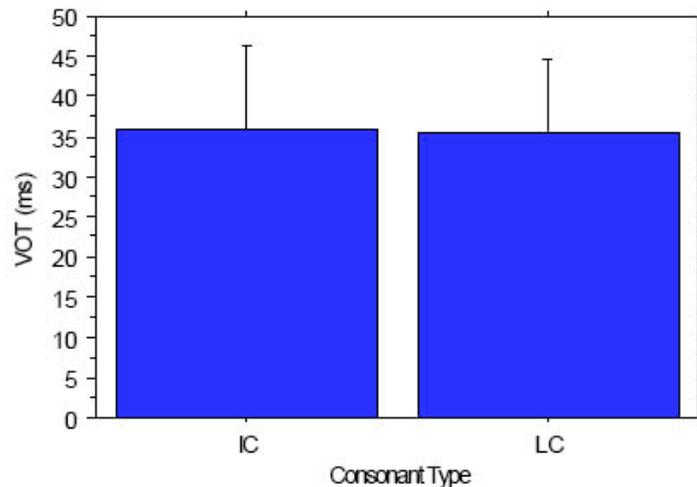


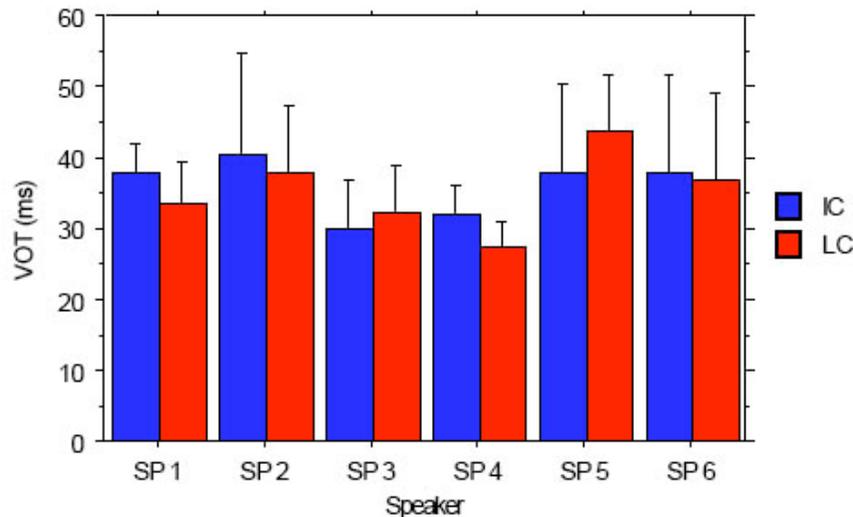
Figure 5-31: Mean voice onset time in milliseconds by Consonant Type (IC and LC) in non-word targets containing /t/, excluding the pair upe/tupe.



VOT values for non-word LCs and ICs in tokens containing /t/ were then examined for each individual speaker. An ANOVA examining the factors Speaker and Consonant Type was conducted, which showed a significant effect of Speaker ($F(5, 95)$

= 3.903, $p = .0029$), no effect of Consonant Type ($F(1, 95) = .129$, NS), and no interaction ($F(5, 95) = .808$, NS). Note as well in Figure 5-32 below that speakers again did not consistently produce differences in the same direction.

Figure 5-32: Mean voice onset time in milliseconds by Consonant Type (IC and LC) and Speaker in non-word targets containing /t/.



An analysis of VOT values in the segment /t/ in the current production sample reveals little consistent variation between realizations of LCs and ICs. The current data have not replicated previous findings concerning VOT in LCs and ICs (Dejean de la Bâtie 1993; Wauquier-Gravelines 1996). Differences not only failed to reach significance, but individual speakers also failed to exhibit differences in same direction, i.e. some speakers produced longer VOT in LCs while others produced longer VOT in ICs. This result is in line with Fougeron (2001) who did not find consistent variation in VOT in either /t/ or /k/ at different prosodic levels (word-initial, syllable-initial, phrase-initial). Keating, Cho, Fougeron and Hsu (2003) also demonstrated that VOT does not represent a reliable cue to prosodic level in French. Fougeron and colleagues' data on acoustic variation within a prosodic hierarchy combined with inconsistent differences in the current production sample render it unlikely that VOT represents a systematic and reliable cue to word boundaries in spoken French.

5.3.3 Discussion

A series of acoustic analyses was conducted post-hoc on the current production sample in order to investigate whether acoustic factors other than the segmental duration of pivotal consonants are sensitive to the presence or absence of liaison in spoken French. Previous work on liaison has found acoustic variation in both the duration of vocalic

segments surrounding pivotal consonants (Spinelli et al. 2003) and voice onset time in /t/ (Dejean de la Bâtie 1993; Wauquier-Gravelines 1996). The current sample has offered mixed support of these findings.

The current results support previous evidence of durational differences in pre-boundary vowels. Consistent with Spinelli et al. (2003), vowels appearing before LCs in the current production sample were significantly shorter than vowels appearing before ICs, though this effect was observed in real-word tokens only. Also consistent with Spinelli et al. (2003), no durational differences in post-boundary vowels were observed. Results regarding VOT of /t/ in environments of liaison, on the other hand, are inconsistent and fail to provide support for consistent variation in this particular acoustic cue.

One further result that emerged from these post-hoc analyses is that acoustic cues to the presence or absence of liaison appear to be more robust in real-words than in non-word. Where significant (or just short of significant) differences were observed in these post-hoc acoustic analyses (pre-boundary vowels, VOT), these differences were observed uniquely in real-word tokens.

Of particular note is the fact that durational variation in pivotal consonants, the main focus of this dissertation, showed consistent patterns in both real- and non-word tokens, namely consonants were significantly shorter in liaison position than in initial position across both word types. In other words, of the four acoustic factors investigated in this production sample (segmental duration of pivotal consonants, of pre-boundary vowels, of post-boundary vowels, and VOT), the segmental duration of pivotal consonants was the only acoustic factor to show consistent and significant variation between liaison and non-liaison environments across both our entire sample (real and non-words) as well as across all three segments /n, t, z/, suggesting that this single acoustic factor is significantly more robust than other cues.

5.4 Instrumental Manipulation of Stimuli

From this production sample, a set of experimental stimuli to be used in the perception portion of the current study was created by exaggerating the durational differences between LCs and ICs through instrumental manipulation. A three-step durational continuum was created in which shortened pivotal consonants represent instances of LCs, pivotal consonants with durations intermediate to LCs and ICs represent a baseline value, and lengthened pivotal consonants represent instances of ICs.

In order to determine which value the duration of the manipulated consonants should take, the distribution of durations from the production sample discussed above was examined. The following mean durations and their respective SDs were calculated for each of the three consonants: the mean duration of LCs, the mean duration of ICs, and the mean duration of all tokens of that consonant (LCs and ICs combined). Given that significant differences were also observed between the durations of consonants in one- and two-syllable words, these durational means and their SDs were taken into account as well. Thus, for each of the three segments, /n, t, z/, six separate means and SDs were calculated:

- 1) mean duration of LCs in one-syllable words,
- 2) mean duration of all pivotal consonants in one-syllable words,
- 3) mean duration of ICs in one-syllable words,
- 4) mean duration of LCs in two-syllable words,
- 5) mean duration of all pivotal consonants of two-syllable words and
- 6) mean duration of ICs in two-syllable words.

Mean segmental durations and SDs for each consonant in each condition are given below in Table 5-12.

Table 5-12: *Mean segmental durations in milliseconds and standard deviations (in parentheses) of each of three pivotal consonants in six conditions.*

	/n/	/t/	/z/
Two-syllable liaison consonant	82.54 (17.71)	63.21 (18.33)	89.50 (16.68)
All two-syllable tokens	94.04 (23.81)	71.08 (23.87)	94.97 (17.53)
Two-syllable initial consonant	105.53 (23.67)	78.95 (26.19)	100.44 (16.74)
One-syllable liaison consonant	96.81 (21.28)	75.86 (29.77)	93.59 (18.48)
All one-syllable tokens	107.78 (26.34)	84.71 (31.04)	100.86 (20.52)
One-syllable initial consonant	118.76 (26.47)	93.32 (30.00)	108.14 (19.97)

A three-step durational continuum of values was then created based on this distribution to serve as the values used in the instrumental manipulation of stimuli. Following methodology laid out in Shatzman and McQueen (2006), the factor by which the shortened and lengthened segments were manipulated was the standard deviation in each respective condition. Previous behavioral studies employing similar methodologies utilized stimuli whose segments were altered by a fixed factor of duration (see for example Huggins, 1972; Quené 1992; Warner et al. 2004). For the current study, we decided to manipulate the stimuli by a factor of the SD for two reasons. First, given that the objective of the current study is to examine the perception of allophonic variation in pivotal consonants, the use of the SD ensures that the durations of manipulated stimuli, though exaggerated, represent points that fall within the distribution of durations and therefore represent reasonable instances of allophonic durational variation in spoken French. Second, SDs are calculated for each particular consonant in each condition. The SD factor is therefore more sensitive and context-specific than a fixed durational factor in that it takes into account any possible variation among the different consonant classes.

Using this factor for each individual consonant allows more sensitivity to inherent durational differences due to consonant class.

The three values of duration used for the continuum were as follows: The shortened (*liaison*) version of each token represented the mean duration for all instances of that consonant in the *liaison* environment *minus* one SD from that particular mean. The value for the midpoint of the continuum (baseline version) represented simply the mean duration across all instances (LCs and ICs) of each consonant. Finally, the value for the lengthened (*word-initial*) version of the consonant represented the mean duration for that consonant in *word-initial* position *plus* one SD from that particular mean.

Again, since significant differences were found between the segmental durations of one-syllable and two-syllable words, a different continuum of durations was calculated for each of these conditions. The resultant durations used as target values in the manipulated stimuli are presented in Table 5-13. Values in parentheses represent the percentage difference from the production mean in that condition (Table 5-12 above).

Using these target durations, tokens were subsequently edited using Praat speech-editing software. The recordings from one of the six speakers who participated in the production procedure were chosen to be manipulated for use in the behavioral tasks. Speaker Three was chosen because it was judged that she maintained a more constant speech rate and intonation throughout both lists than the other five speakers and because she made no errors during the recording procedure.

Table 5-13: *Segmental durations in milliseconds used in the manipulation of experimental stimuli. Percentage difference from the production mean in each condition is given in parentheses.*

		/n/	/t/	/z/
Two-syllable tokens	Short (LC) version	64.83 (-21.46%)	44.88 (-29%)	72.82 (-18.64%)
	Baseline (average) version	94.04 (0)	71.08 (0)	94.97 (0)
	Long (IC) version	129.20 (+22.43%)	105.14 (+33.12%)	117.17 (+14.26%)
		75.53 (-21.94%)	46.09 (-39.24%)	75.11 (-22.21%)
One-syllable tokens	Short (LC) version	107.78 (0)	84.71 (0)	100.86 (0)
	Baseline (average) version	145.23 (+22.29%)	123.32 (+32.15%)	128.11 (+18.47%)
	Long (IC) version			

Of the three tokens of each target recorded by this speaker one token which was judged as being articulated fluently, clearly, and at a normal rate was chosen to be instrumentally manipulated. Though recordings were made and measurements were taken of both the vowel-initial and consonant-initial member of each lexically ambiguous minimal pair, only the consonant-initial member of each pair was chosen for the sake of limiting the number of stimuli to be employed in the perception portion of the

experiment. In other words, though, for example, tokens of both *un air* and *un nerf* were recorded and included in acoustic analysis, only one token of *un nerf* was chosen to be instrumentally altered. Optimally, manipulated versions of both vowel- and consonant-initial tokens would have been manipulated to be included in the perceptual tasks, however, in the interest of feasibility, only the consonant-initial tokens were chosen so as not to render the behavioral tasks excessively long and taxing for participants. The consonant-initial version of each minimal pair was chosen mainly for reasons of practicality in instrumental manipulation. The relatively longer durations of initial consonants facilitate the manipulation process in that a longer segment is available to work with.

Durations of /t/ were manipulated by either deleting a portion of the closure as needed to shorten the consonant or by inserting a segment of silence into the closure as needed to lengthen the consonant. Durations for /n/ and /z/ however were manipulated by cross-splicing. Again, following methodology laid out in Shatzman and McQueen (2006), middle portions of /n/ and /z/ were deleted leaving approximately 20 ms of the initial and final portions of the segment. A portion of a version of the same segment from another version of the same word from the same speaker was then spliced into the recording in order to attain the desired duration. All splices were made at zero crossings in an effort to avoid any acoustic artifacts such as clicks, buzzes or other audible distortions that could occur in the splicing process. Due to the constraints of splicing and inserting only at zero crossings, it was usually impossible to manipulate the tokens to exactly match the desired durations given above in Table 5-13, but great effort was made to keep the manipulated durations within 5 ms of the desired duration.

The manipulation of these phrases resulted in 36 real-word sequences and 36 non-word sequences (12 tokens x 3 manipulated versions) that are therefore phonemically identical in their content but differ as to the precise acoustic phonetic realization of the individual consonants under investigation. These manipulated stimuli were then utilized in the behavioral tasks presented in the following two chapters.

CHAPTER SIX: AX Discrimination Task

6.0 Introduction

In this chapter, the first of two behavioral studies probing the perception of segmental duration in phrases rendered ambiguous by the possibility of liaison in spoken French is presented. An AX discrimination task incorporating pairs of instrumentally manipulated stimuli taken from the three-step continuum of duration described in §5.4 above was used to investigate the saliency of durational differences by establishing thresholds of noticeability between LCs and ICs for both native speakers (NS) and non-native speakers (NNS) of French. The use of a discrimination task is motivated by the assumption that segmental duration represents an effective cue to segmentation and lexical access in cases of ambiguity only to the extent that this cue is perceptually salient to listeners. Experiment 1 employs real-word manipulated stimuli, while Experiments 1.1 and 1.2 employ non-word manipulated stimuli.

6.1 Predictions

The acoustic-phonetic literature on *just noticeable differences* (JNDs), i.e. the amount by which a stimulus must be altered in order to produce a noticeable variation in the sensory experience of the perceiver, offers some predictions as to how salient the durational differences in the current study will be. Lehiste (1976) holds that differences in segmental duration on the order of 10 – 40 ms are perceivable in *optimal* conditions. Hawkins (1977) proposes that a difference of 25 ms in a segment is perceivable to the listener. Klatt (1976) on the other hand gives a range of 10 – 20 ms, partially based on the following two studies: Fujisaki, Nakamura & Imoto (1975), who showed that a change of 10 ms in a segment is discriminable in a language which makes use of contrastive duration such as Japanese; and Huggins (1972), who showed that a change of 20 ms in a segment provokes a discriminable change in the rhythm of a (English) sentence. Furthermore, Huggins (1972) showed that listeners are more sensitive to changes in duration in vocalic segments than in consonants.

Table 6-1 below presents the pairings of stimuli to be used in the AX discrimination tasks in the present investigation, where 1 represents the shortened token on the continuum, 2 represents the baseline token, and 3 represents the lengthened token. This table also gives the mean difference in absolute duration (in milliseconds) between the two stimuli in each experimental trial, i.e. between the first member (A) the second member (X) for each of the three segments, /n, t, z/. (Refer to Table 5-13 above for values used in the manipulation of stimuli.) Note that, according to what are considered to be absolute JNDs, these absolute measurements are relatively large.

Table 6-1: Mean durational differences in milliseconds within nine stimulus pairs used in AX discrimination task for each of three consonants.

PAIR	Mean	Mean	Mean	Mean
	durational difference within /n/ pairs	durational difference within /t/ pairs	durational difference within /z/ pairs	durational difference across segments
1_1	0	0	0	0
1_2	+30.73	+32.41	+23.95	+29.03
1_3	+67.04	+68.75	+48.68	+61.49
2_1	-30.73	-32.41	-23.95	-29.03
2_2	0	0	0	0
2_3	+36.31	+36.34	+24.73	+32.46
3_1	-67.04	-68.75	-48.68	-61.49
3_2	-36.31	-36.34	-24.73	-32.46
3_3	0	0	0	0

Perceptible variation in segmental duration can also be characterized in relational terms rather than measured in absolute values. *Weber's law* is a general law of psychophysics which states that JNDs are determined by a proportion of the original stimulus value. This law predicts, for example, that the same physical change in duration would be more perceptible in a shorter segment than in a longer segment due to the fact that the absolute value of the change represents a larger proportion of the former than the latter. In terms of relational change, Klatt (1976) proposes that listeners can discriminate differences on the order of 20%.

The perception of duration can also vary as a function of the duration of surrounding segments (Diehl et al. 1980; Miller 1987; Summerfield 1981), speech rate (Summerfield 1975) as well as the length of the utterance in which the segment appears (Kawai & Carrell 2005). These studies are all consistent with research showing that listener ratings of the goodness of a stimulus are dependent on speaking rate (Allen & Miller 2001). Klatt and Cooper (1975) also showed that JNDs depend on a segment's position within a sentence; durational changes in initial segments, for example, tend to be more perceptually salient and therefore exhibit smaller JNDs than utterance-medial and utterance-final segments.

Table 6-2 below presents the relational change between each of the stimulus pairs used in the current task for each of the three segments, /n, t, z/ (i.e. the percentage change in absolute duration from stimulus A to stimulus X in the AX discrimination task). As can be seen, the percentage change between manipulated stimuli in the current task ranges from 20.16 – 151.14 % depending on the pairing, values which are well above the 20% considered to be discriminable.

Table 6-2: Mean percent change in duration within nine stimulus pairs used in AX discrimination task for each of three consonants.

PAIR	Mean	Mean	Mean	Mean
	% change in duration within /n/ pairs	% change in duration within /t/ pairs	% change in duration within /z/ pairs	% change in duration across segments
1_1	0	0	0	0
1_2	+43.79	+71.25	+32.38	+49.14
1_3	+95.52	+151.14	+65.81	+104.16
2_1	-30.45	-41.61	-24.46	-32.17
2_2	0	0	0	0
2_3	+35.98	+46.65	+25.26	+35.96
3_1	-48.85	-60.18	-39.69	-49.58
3_2	-26.46	-31.81	-20.16	-26.14
3_3	0	0	0	0

Taking into account thresholds of noticeability of both absolute duration and percentage change, the durational differences in the current stimulus sample are substantial and therefore discrimination would not be predicted to be challenging. However, there are other factors which could mitigate perceptual saliency and which should be considered. One key factor is that the durational differences in the current stimulus sample represent an allophonic (within-category) distinction. It has long been known that variation crossing categorical boundaries is more readily perceived than the same physical difference within the same category (Liberman et al. 1957). Recent research employing AX and AXB discrimination tasks to test the perception of allophonic variation by native speakers has further established that the discrimination of allophonic variation is significantly worse than discrimination of phonemic contrasts (Pegg & Werker 1997; Shea & Curtin 2005; Whalen, Best & Irwin 1997). Specifically, Peperkamp, Pettinato and Dupoux (2003) showed that physical differences between segments (in this study, a voicing contrast) are more easily distinguished when removed from phonological context than when presented in a context where the variation is phonologically licensed. Peperkamp et al. tested native French-speaking participants on the discrimination of the voiced uvular fricative /ʁ/ and its voiceless allophone /χ/. This voicing contrast is not phonemic in French, but rather results from voicing assimilation (e.g. *perde* [pɛʁd] ‘lose’, 3rd person singular present subjunctive, versus *perte* [pɛχt] ‘loss’). Participants discriminated the two segments significantly better in single VC syllables (e.g. [aʁ] versus [aχ]) than in VC.CV sequences which represented a phonological context that licenses the allophonic change (e.g. [aʁ.do] versus [aχ.sa]). This result suggests the perceptual system is less sensitive to physical variation in the signal when this variation is motivated by the phonological environment.

The durations of the manipulated consonants used in the current task fall within the range and distribution of the production sample discussed in Chapter Five, and therefore represent valid tokens of allophonic variation in spoken French. Furthermore, participants are hearing these durational differences presented in the phonological environment that licenses the variation under investigation, i.e. in possible liaison environments. These factors are likely to depress perceptual saliency.

One further factor that could mitigate the perceptual saliency of duration in the current stimulus sample is the length of the sequences and the location in the sequences in which the manipulated consonants appear. A large portion of the JND literature is based on the perception of individual segments or syllables, which are usually removed from any semantic context. The manipulated consonants in the current stimulus sample appear in phrases of 2 – 3 words incorporating from 2 – 4 syllables and between 4 – 9 phonemes. Recall as well that Klatt and Cooper (1975) showed that variation in segments that occur utterance-medially are discriminated less readily than utterance-initial segments. Participants in the current task therefore must attend to a signal that is more complex, both semantically and acoustically, than is usually seen in the study of JNDs.

In light of the above factors, we predict that the thresholds of discrimination of these stimuli will be greater than observed JND thresholds of both absolute duration and percent change would predict, even though the durational differences in the current stimulus sample are significantly above what is assumed to be noticeable. Given the range of durations in the stimulus sample, we predict that tokens differing by one degree of separation will be difficult to distinguish. Tokens separated by two degrees, on the other hand, should be relatively easy to distinguish.

The current task also includes three different consonants that surface in environments of liaison, /n/, /t/ and /z/. One additional question that the current study seeks to examine is whether perceptual saliency of segmental duration varies as a function of the particular segment. If differences in perceptual saliency are found among the three segments, we foresee two plausible outcomes. First, following Bybee's (2001a/b, 2005) usage-based model of mental representations of liaison discussed above in §2.3, the perceptual saliency of the individual consonants could be affected by their frequency of occurrence in liaison environments in spoken French. Corpus research (Durand & Lyche 2008; Léon 1992) has found that /z/ accounts for the majority of occurrences of liaison, followed by /n/ and /t/. A frequency-based model would lead to the prediction that perceptual saliency among the consonants will follow this same pattern, with /z/ being discriminated significantly better than /n/ and /t/, and /n/ in turn being discriminated better than /t/.

However, the production sample on which the manipulated durations in the current task are based yielded significant differences among the three consonants in an alternate direction. Differences between LCs and ICs were most robust for /n/ (mean difference 22.47 ms), followed by /t/ (mean difference 17.25 ms) and /z/ (mean difference 9.20 ms). If differences in the perceptual saliency among the three consonants are found, a second possibility is that differences will be based solely on the acoustic properties of the signal and will therefore follow the durational distribution of the production sample.

This leads to the prediction that /n/ will be discriminated significantly better than /t/, which will in turn be discriminated significantly better than /z/.

Regarding predictions as to the performance of the two participant groups, to our knowledge there is little literature investigating the discrimination of within-category durational variation by non-native speakers, and thus predictions based on previous findings are not possible. However, several possible outcomes present themselves. If both groups perform comparably, this could be attributable to two explanations. The first is that sensitivity to these durational differences is purely acoustic, i.e. based entirely on the acoustic properties of the signal and independent of language experience. In this case, we would conclude that perception of segmental duration is based purely on general auditory processes. The second possibility in this case is that sensitivity is indeed conditioned by language experience and that the NNS participants are performing in a nativelike manner. This result would suggest that the NNS participants had acquired nativelike sensitivity to fine phonetic differentiation in this phonological environment.

Alternatively, if significant differences between the two participant groups are observed, we predict that native performance will be superior to that of late learners. This outcome would suggest that sensitivity to durational differences in this particular context (i.e. environments of possible liaison) is conditioned by language exposure and that the NNS participants have not acquired processing strategies to a nativelike degree.

We now turn to the first of three AX discrimination tasks, which employs real-word manipulated stimuli.

6.2 Experiment 1: AX Discrimination Task (real-word stimuli)

6.2.1 Participants

Thirty-six participants comprising two groups took part in Experiment 1. The control group consisted of 18 NS of French (15 female, 3 male) ranging in age from 19-54 years (mean: 30.2 years), all of whom lived in or around Paris, France at the time of testing. All NS participants had studied languages other than French to varying degrees of proficiency, however none were bilingual from birth.

The experimental group consisted of 18 native speakers of English (11 female, 7 male; mean age: 42.2 yrs, range: 26-71) all of whom met a minimum immersion requirement of five years in France or a French-speaking country at the time of testing (mean residency: 13.8 yrs; range: 5 – 44 yrs). Meeting this requirement is not intended to predict a certain level of proficiency among participants, but rather is taken as an indication that participants have reached end state in their attainment of L2 French. (See Birdsong 2004 and Johnson & Newport 1989 for a discussion of residency requirements).

Mean age of arrival in France for the NNS group was 28.4 years (range: 18-59 years). Mean age of first exposure to French (e.g. either through classroom instruction or time spent in a French-speaking country) was 17.2 years of age (range: 6 – 54 years of age). The variety of English spoken by each NNS participant was not controlled for. The NNS group consisted of 14 speakers of American English and 4 speakers of British English. Almost all NNS participants spoke second languages other than French to varying degrees of proficiency, but none were bilingual from birth.

Participants were paid eight Euros to take part in the experiment. All had normal or corrected vision and none reported any hearing impairment. Handedness was not controlled for.

6.2.2 Stimuli

Stimuli consisted of real-word pairs of phrases drawn from the three-step continuum of manipulated phrases described in §5.4. Each token on the three-step continuum was paired with a duplicate version of the same token as well as with the other two manipulated versions of that token. This resulted in nine pairings for each of the 12 stimuli. Of the nine pairs, three were identical (1_1, 2_2, 3_3) and six were different (1_2, 1_3, 2_1, 2_3, 3_1, 3_2). Of the six different pairs, two pairs were separated by two degrees on the durational continuum (1_3, 3_1) and four were separated by one degree (1_2, 2_1, 2_3, 3_2).

6.2.3 Procedure

Prior to testing, the NNS group completed via email an extensive biographical questionnaire adapted from Marian et al. (2007) in which they were asked to provide information pertaining to language use and history. Questions included an estimation of the proportion of time spent on a daily basis hearing and speaking each of the participants' languages; self-reported proficiency in speaking, reading, pronunciation and listening (on a scale of 1 – 10) in each of the participants' languages; biographical factors such as amount and level of education in the participants' languages; the age of first exposure to the L2; and length of residence in France.²⁰ (See Appendix II for complete questionnaire.)

Participants were tested individually in a quiet room. Each participant first read and signed a consent form for their participation in the experiment. A French version of the consent form was given to native French speakers, while an English version was given to native English speakers (see Appendices III and IV). The native French-speaking participants also filled out a very brief questionnaire about language history (see Appendix V). Based on a body of methodological research which has suggested that it is important for bilingual participants to be in the appropriate language 'mode' while in the experimental environment (e.g. Grosjean 1998), oral communication at the time of testing was conducted solely in French with both participant groups. Written instructions concerning the experimental tasks to be performed were also presented in French (see Appendix VI).

The experimental protocol was created using E-Prime experimental software (Schneider, Eschman & Zuccolotto 2002) and presented on a Dell Inspiron 600m laptop computer. Stimuli were presented binaurally through Koss UR 20 headphones. Participants were instructed that they would hear pairs of phrases in French and to indicate whether the two phrases were identical or different by pressing on the keyboard either 1 or 2 respectively. No direction was offered to participants as to what parameters

²⁰ We will return to discussion of the NNS group's biographical information, including correlations with behavioral data in Chapter Eight.

responses should be based on. Participants were asked to respond quickly, but not so quickly as to sacrifice accuracy. Before beginning the experiment, participants completed a training portion consisting of 14 trials in order to familiarize them with the procedure. Items included in the training portion were not included in the experimental portion. Each experimental trial consisted of one pair of manipulated stimuli separated by an inter-stimulus interval of 250 ms²¹. Individual trials were separated by a 2000 ms pause. A limit for response time was set at 6000 ms. Following Diehl (personal communication), each of the 9 pairs of 12 stimuli was presented 6 times in random order, resulting in a total of 648 trials. There were no visual stimuli to accompany the auditory stimuli. No feedback as to the accuracy of responses was given in either the training or the experimental portion. Testing lasted approximately 50 minutes.

6.2.4 Results: Experiment 1

One NNS participant (NNS 5) was removed from analysis due to the fact that he responded *same* to all 648 trials of the experiment. The analyses that follow therefore include 17 NNS participants and 18 NS participants.

Responses for the three same pairs (1_1, 2_2, 3_3) were not included in analysis, therefore we report here only responses for the six different pairs (1_2, 1_3, 2_1, 2_3, 3_1, and 3_2). All analyses are by subject.

6.2.4.1 Mean Accuracy

We first calculated mean accuracy rates for each NS participant for each of the six different pairs, which are presented along with range and standard deviations in Table 6-3 below. We then determined whether mean accuracy for each pair was significantly above or below chance using single sample t-tests (two-tailed) with a 95% confidence interval. In a categorical AX task, in which participants must choose either *same* or *different* in each trial, chance performance is set at 50%. T-test results are also given at the bottom of Table 6-3.

We then calculated mean accuracy rates for each NNS participant for each of the six different pairs, which are presented along with range and standard deviations in Table 6-3.5 below. T-test results are also given at the bottom of Table 6-3.5.

²¹ Given that the objective of the current task was to tap the acoustic saliency of these durational differences, an ISI of 250 ms was used to ensure that processing was taking place in ‘phonetic mode’ as opposed to ‘phonological mode’ (Werker & Tees 1984). A relatively shorter ISI is thought to allow listeners to compare stimuli while they remain in auditory sensory memory and thus gives the listener an advantage in discriminating acoustic variation.

Table 6-3: Mean accuracy rates on AX discrimination task employing real-word stimuli for NS participants by pair. Given at the bottom of the table are mean accuracy and standard deviations across participants and the results of single sample t-tests comparing performance to chance (50%).

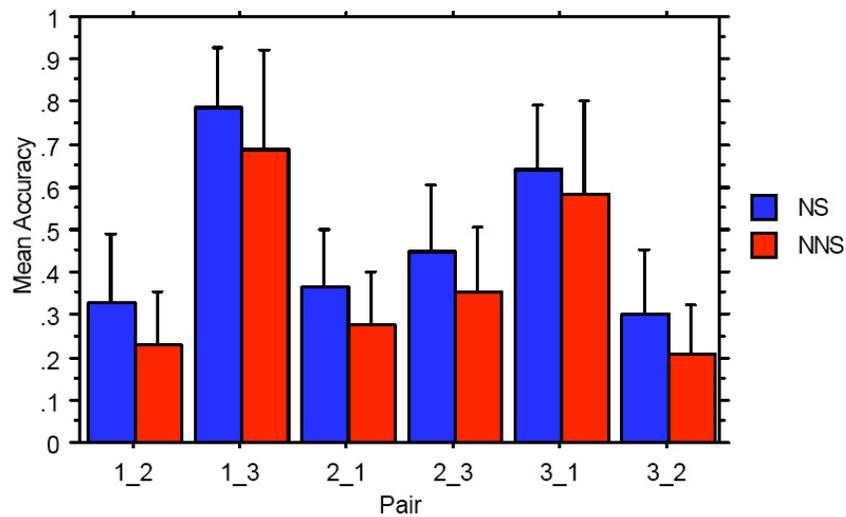
	<u>1_2</u>	<u>1_3</u>	<u>2_1</u>	<u>2_3</u>	<u>3_1</u>	<u>3_2</u>
NS 1	51	88	53	63	81	49
NS 2	51	71	58	51	76	44
NS 3	28	82	43	50	75	58
NS 4	33	97	47	42	78	28
NS 5	10	72	26	22	56	8
NS 6	67	88	50	76	71	56
NS 7	24	57	26	43	49	26
NS 8	33	79	38	44	56	35
NS 9	28	87	23	42	58	13
NS 10	32	60	18	42	38	30
NS 11	7	75	30	27	70	20
NS 12	27	73	43	25	57	18
NS 13	48	93	53	60	83	47
NS 14	30	47	12	37	32	18
NS 15	27	77	25	45	57	28
NS 16	55	98	37	73	80	28
NS 17	33	97	47	42	78	28
NS 18	10	72	26	22	56	8
MEAN ACC	33.0	78.5	36.4	44.8	63.9	30.1
RANGE	7-67	47-98	12-58	22-76	32-83	8-58
(SD)	(16.20)	(14.40)	(13.56)	(15.79)	(15.14)	(15.35)
DIFF FROM CHANCE	P = .0003	P < .0001	P = .0005	NS	P = .0011	P < .0001

Table 6-3.5: Mean accuracy rates on AX discrimination task employing real-word stimuli for NNS participant by pair. Given at the bottom of the table are mean accuracy and standard deviations across participants and the results of single sample t-tests comparing performance to chance (50%).

	<u>1 2</u>	<u>1 3</u>	<u>2 1</u>	<u>2 3</u>	<u>3 1</u>	<u>3 2</u>
NNS 1	1	13	10	7	17	6
NNS 2	10	21	17	8	25	8
NNS 3	18	67	21	24	71	19
NNS 4	8	35	4	18	17	6
NNS 5	removed from analysis					
NNS 6	31	88	49	38	92	40
NNS 7	19	83	32	33	82	32
NNS 8	19	76	25	25	57	17
NNS 9	45	95	33	47	72	20
NNS 10	24	69	33	50	40	22
NNS 11	10	73	20	43	73	20
NNS 12	13	60	25	35	58	18
NNS 13	33	87	30	38	72	13
NNS 14	18	83	15	32	70	10
NNS 15	35	82	45	58	62	42
NNS 16	27	77	37	48	57	13
NNS 17	28	77	23	40	73	28
NNS 18	47	80	47	57	57	38
MEAN ACC	<u>22.7</u>	<u>68.6</u>	<u>27.4</u>	<u>35.4</u>	<u>58.5</u>	<u>20.7</u>
RANGE	<u>1-47</u>	<u>13-95</u>	<u>4-49</u>	<u>7-57</u>	<u>17-92</u>	<u>6-42</u>
(SD)	<u>(12.84)</u>	<u>(23.60)</u>	<u>(12.70)</u>	<u>(15.19)</u>	<u>(21.90)</u>	<u>(11.62)</u>
DIFF FROM CHANCE	<u>P <.0001</u>	<u>P =.0051</u>	<u>P <.0001</u>	<u>P =.0011</u>	<u>NS</u>	<u>P <.0001</u>

Mean accuracy rates across participants were subsequently submitted as the dependant variable in a two-way ANOVA examining the factors Pair (six levels: 1_2, 1_3, 2_1, 2_3, 3_1, 3_2) and Participant Group (two levels: NS and NNS). This analysis revealed a main effect of Pair ($F(5, 198) = 52.64, p < .0001$), indicating a significant difference in discrimination among the six pairs for both groups. The analysis also revealed a main effect of Participant Group ($F(1, 198) = 16.18, p < .0001$) and no interaction ($F(5, 198) = .11, NS$). The NS group discriminated stimuli significantly better than the NNS group. These results are displayed in Figure 6-1 below.

Figure 6-1: Mean accuracy on AX discrimination task employing real-word stimuli by Pair and Participant Group.



A series of Scheffe post-hoc tests was then performed to establish which pairs were discriminated better than others for each participant group, the results of which are presented below in Table 6-4. As can be seen, the two pairs representing a two-degree separation on the continuum (3_1 and 1_3) were discriminated significantly better than those pairs separated by one degree on the continuum. Furthermore, NS and NNS participant groups behaved similarly in the discrimination of eight out of 15 pair-by-pair comparisons.

Table 6-4: Results of Scheffe post-hoc tests showing discrimination among different pairs of real-word stimuli according to mean accuracy rates for NS and NNS participants.

Pair-by-pair comparison	Native Speakers: Value of P	Non-native Speakers: Value of P
1_2, 1_3	< .0001	< .0001
1_2, 2_1	NS	NS
1_2, 2_3	NS	.0323
1_2, 3_1	< .0001	< .0001
1_2, 3_2	NS	NS
1_3, 2_1	< .0001	< .0001
1_3, 2_3	< .0001	< .0001
1_3, 3_1	NS	(.0873)
1_3, 3_2	< .0001	< .0001
2_1, 2_3	NS	NS
2_1, 3_1	< .0001	< .0001
2_1, 3_2	.NS	NS
2_3, 3_1	.0172	.0001
2_3, 3_2	NS	.0135
3_1, 3_2	< .0001	< .0001

6.2.4.2 D-prime

In addition to mean accuracy, d-prime scores were also calculated. D-prime, a measure used in signal detection, reflects not only accuracy (as measured by correct responses and correct rejections), but also factors out any possible response bias (as measured by incorrect responses and false alarms). This measure is therefore thought to provide a more accurate measure of discrimination than raw accuracy scores. D-prime scores generally range from 0 – 4, though negative scores are possible (and were observed in the current experiment). Good discrimination is generally thought to be reflected by a d-prime score of 3 or above, while a score of below 1 is taken to be indicative of very low (or a lack of) discrimination. Scores were calculated for each participant for each of six pairs across items and are presented in Table 6-5 below.

Table 6-5: *D-prime scores on AX discrimination task employing real-word stimuli for each participant by pair. Mean scores, standard deviations and range across participants are given at the bottom of the table.*

	1_2	1_3	2_1	2_3	3_1	3_2		1_2	1_3	2_1	2_3	3_1	3_2
NS 1	0.58	1.73	0.54	0.80	1.43	0.53	NNS 1	0.68	0.52	0.19	0.00	0.69	0.09
NS 2	0.33	0.86	0.56	0.38	1.23	0.37	NNS 2	0.36	0.84	0.45	0.00	0.73	0.00
NS 3	-0.14	1.36	0.29	0.47	0.98	0.51	NNS 3	0.43	1.78	0.53	0.63	1.40	-0.04
NS 4	1.89	4.21	2.25	2.12	2.42	1.06	NNS 4	0.24	1.26	-0.41	0.43	0.17	-0.43
NS 5	0.47	2.33	1.11	0.98	1.80	0.24	NNS 5	removed from analysis					
NS 6	0.96	1.70	0.33	1.04	1.11	0.70	NNS 6	0.50	2.17	0.89	0.61	2.81	1.15
NS 7	-0.24	0.64	0.20	0.67	0.31	-0.31	NNS 7	0.53	2.36	0.94	0.97	2.14	0.76
NS 8	0.69	1.93	0.57	0.73	0.99	0.46	NNS 8	1.45	3.03	0.97	0.97	2.23	1.10
NS 9	1.30	3.01	0.39	0.92	1.54	0.21	NNS 9	1.00	2.77	0.90	1.27	1.99	0.56
NS 10	0.94	1.66	0.08	0.79	0.40	0.18	NNS 10	0.14	1.34	0.30	0.74	0.42	-0.10
NS 11	0.85	3.00	0.88	0.79	2.08	0.71	NNS 11	0.27	2.17	1.04	1.70	2.49	1.04
NS 12	0.79	2.02	1.16	0.67	1.93	0.84	NNS 12	0.62	2.00	0.41	0.69	1.28	0.16
NS 13	0.26	1.78	0.52	0.69	1.18	0.15	NNS 13	0.55	2.12	0.70	0.92	1.58	-0.13
NS 14	0.56	1.01	-0.09	0.75	0.76	0.31	NNS 14	0.97	2.83	1.29	1.86	2.41	0.60
NS 15	0.47	1.82	0.20	0.75	1.05	0.30	NNS 15	0.53	1.83	0.95	1.28	0.80	0.29
NS 16	1.68	3.61	1.22	2.17	1.58	0.16	NNS 16	1.14	2.49	1.01	1.29	1.73	0.43
NS 17	1.80	4.38	2.28	2.18	2.49	1.20	NNS 17	0.76	2.08	0.60	1.09	1.42	0.22
NS 18	0.47	2.28	1.11	1.04	1.80	0.24	NNS 18	0.63	1.55	0.34	0.59	0.40	-0.08
MEAN	0.76	2.18	0.76	1.00	1.39	0.44	MEAN	0.52	1.87	0.62	0.84	1.37	0.31
	-0.24	0.64	-0.09	0.38	0.31	-0.31		0.14	0.52	-0.41	0.00	0.17	-0.43
RANGE	-	-	-	-	-	-	RANGE	-	-	-	-	-	-
	1.89	4.38	2.25	2.18	2.49	1.20		1.45	3.03	1.29	1.86	2.81	1.15
SD	0.60	1.07	0.67	0.56	0.62	0.36	SD	0.34	0.69	0.42	0.51	0.82	0.47

Mean d-prime scores were then compared by Pair and Participant Group in a two-way ANOVA, the results of which are shown below in Figure 6-2. This analysis revealed a main effect of Pair ($F(5, 198) = 33.32, p < .0001$) but no effect of Participant Group ($F(1, 198) = 1.67, NS$) and no interaction between the two factors ($F(5, 198) = .237, NS$). While an analysis of mean accuracy suggested that the NS group discriminated the stimuli significantly better than the NNS, this difference in performance was no longer significant in d-prime analysis. However, discrimination among the pairs remained significant for both groups.

Results of Scheffe post-hoc pair-by-pair analyses for both participant groups are given in Table 6-6 below. As observed in the analysis of mean accuracy, the two pairs representing a separation of two degrees on the continuum (3_1 and 1_3) were discriminated significantly better than those pairs representing a separation of one degree on the continuum.

Figure 6-2: *D-prime scores for AX discrimination task employing real-word stimuli by Pair and Participant Group.*

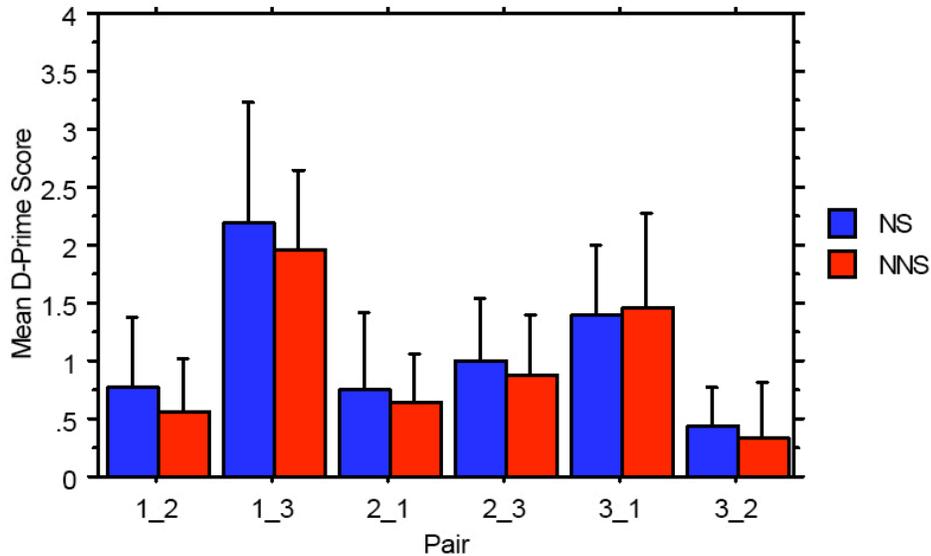


Table 6-6: *Results of Scheffe post-hoc tests showing discrimination among different pairs of real-word stimuli according to d-prime scores for NS and NNS participants.*

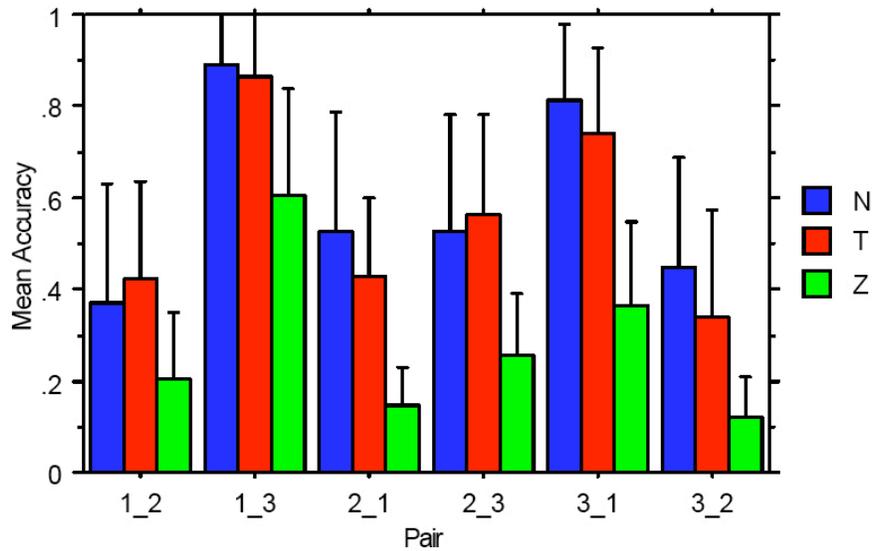
Pair-by-pair comparison	Native Speakers: Value of P	Non-native Speakers: Value of P
1_2, 1_3	< .0001	< .0001
1_2, 2_1	NS	NS
1_2, 2_3	NS	NS
1_2, 3_1	NS	< .0001
1_2, 3_2	NS	NS
1_3, 2_1	< .0001	< .0001
1_3, 2_3	.0001	< .0001
1_3, 3_1	.0382	.0141
1_3, 3_2	< .0001	< .0001
2_1, 2_3	NS	NS
2_1, 3_1	< .0001	.0001
2_1, 3_2	NS	NS
2_3, 3_1	NS	.0052
2_3, 3_2	NS	.0063
3_1, 3_2	.0046	< .0001

6.2.4.3 Analyses by Consonant

Some research on both production and perception has suggested that there may be differences in the saliency of acoustic differences among the various consonants that surface in liaison (Nguyen et al. 2007; Shoemaker 2006; Wauquier-Gravelines 1996). Therefore we investigate here whether participant performance differed significantly in sensitivity among the three consonants tested in the current experiment.

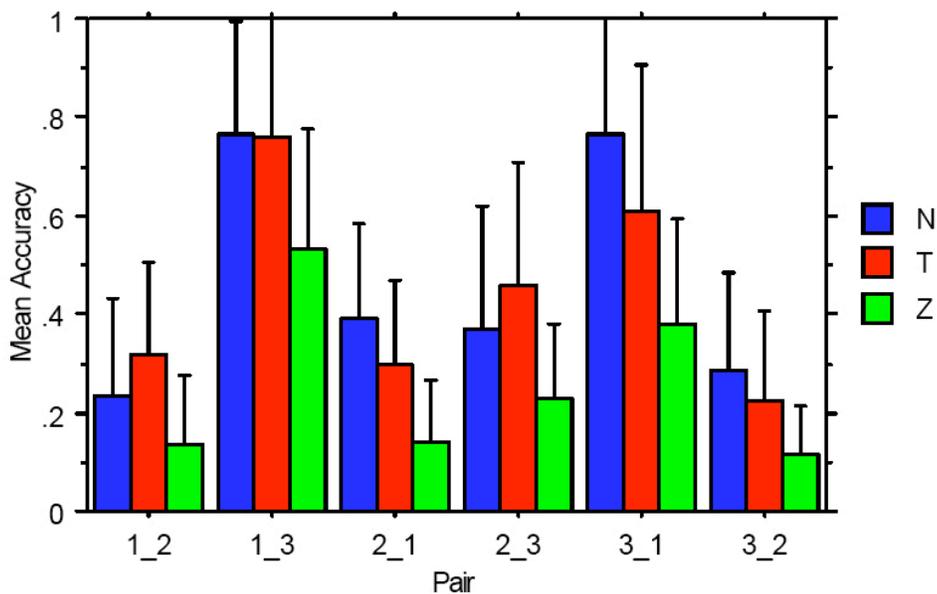
Mean accuracy rates for the NS group for each of the three consonants were subsequently compared in a two-way ANOVA analyzing the factors Consonant (three levels: /n/, /t/, /z/) and Pair (six levels). This analysis revealed a main effect of Consonant ($F(2, 306) = 84.34, p < .0001$) as well as a main effect of Pair ($F(5, 306) = 53.14, p < .0001$). No interaction was observed ($F(10, 306) = .08, NS$). Scheffe post-hoc analyses revealed that the NS group discriminated stimuli containing /z/ significantly worse than stimuli containing /n/ ($p < .0001$) and /t/ ($p < .0001$). There was no significant difference between /n/ and /t/. NS mean accuracy rates for each consonant are given below in Figure 6-3.

Figure 6-3: Native speaker mean accuracy on AX discrimination task employing real-word stimuli by Pair and Consonant.



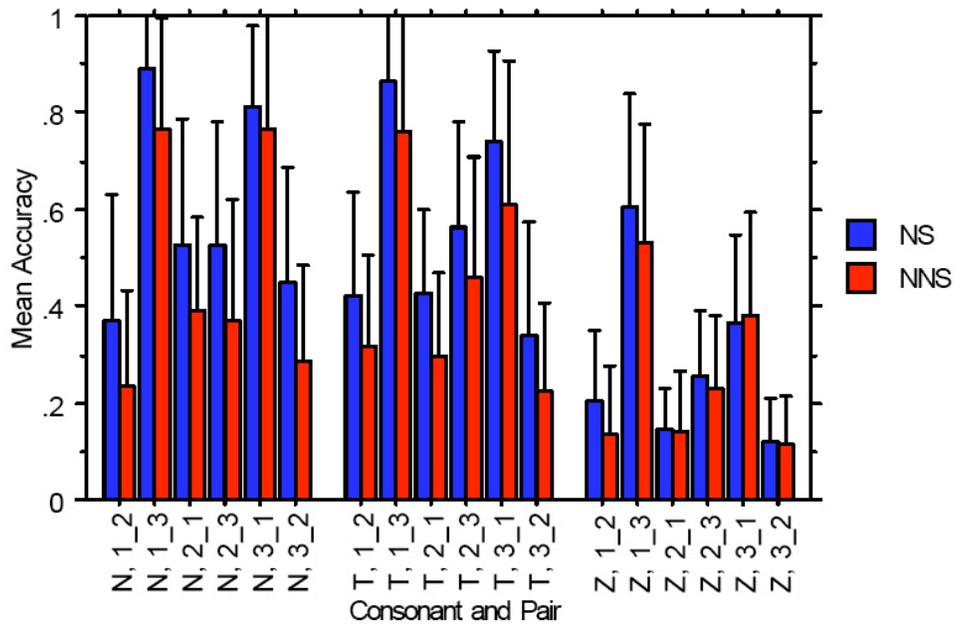
Accuracy rates for the NNS group for each consonant were then submitted to ANOVAs, which also revealed a main effect of Consonant ($F(2, 288) = 29.26, p < .0001$) and of Pair ($F(5, 288) = 43.47, p < .0001$), but no interaction ($F(10, 288) = 1.34, NS$). Of particular note is the fact that differences in accuracy among the three consonants for the NNS group were in the same direction as the NS group; the NNS group discriminated stimuli containing /z/ significantly worse than stimuli containing /n/ ($p < .0001$) and /t/ ($p < .0001$) according to Scheffe post-hoc tests. There was no significant difference between /n/ and /t/. Figure 6-4 shows these results.

Figure 6-4: *Non-native speaker mean accuracy on AX discrimination task employing real-word stimuli by Pair and Consonant.*



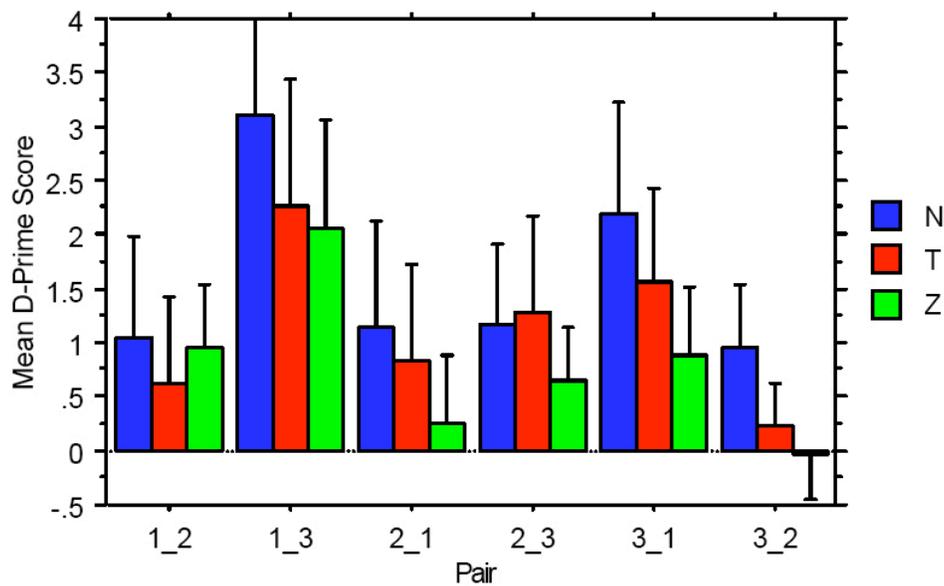
Mean accuracy rates for both participant groups were then compared for the same factors in a three-way ANOVA. This combined analysis revealed a significant effect of Participant Group ($F(1, 594) = 29.754, p < .0001$), a main effect of Pair ($F(5, 594) = 95.717, p < .0001$) as well as a main effect of Consonant ($F(2, 594) = 103.037, p < .0001$). This analysis also revealed two significant interactions: between the factors Pair and Consonant ($F(10, 594) = 2,510, p = .0058$) and the factors Consonant and Participant Group ($F(2, 594) = 3.951, p = .0197$). Both interactions indicate that the degree of discrimination for both participant groups varies as a function of the particular segment. Figure 6-5 summarizes this analysis.

Figure 6-5: Mean accuracy rates for AX discrimination task employing real-word stimuli by Pair, Consonant and Participant Group.



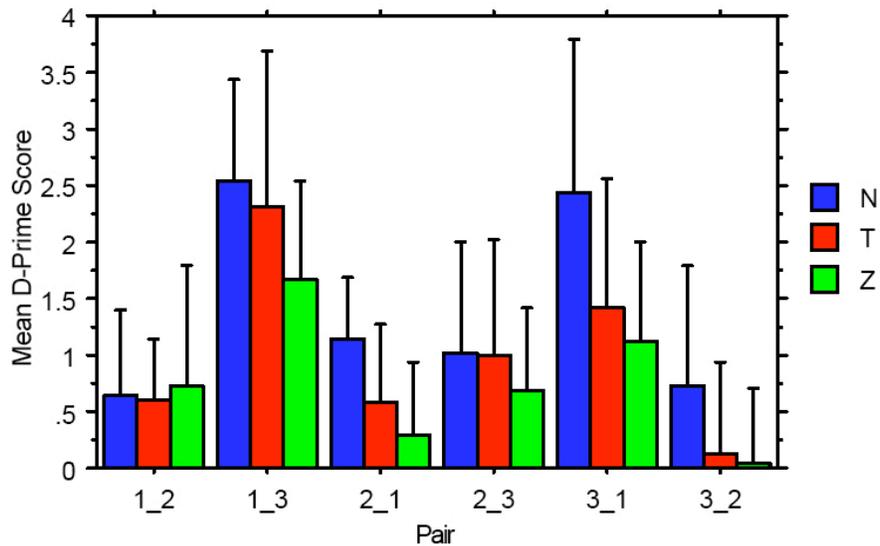
NS d-prime scores on real-word stimuli were then analyzed for each individual consonant and are presented below in Figure 6-6. D-prime scores for the NS group were submitted to ANOVAs with the factors Consonant and Pair, which indicated a main effect of Consonant ($F(2, 306) = 25.75, p < .0001$) as well as a main effect of Pair ($F(5, 306) = 42.63, p < .0001$). A significant interaction was also found ($F(10, 306) = 1.597, p = .0377$), which further underscores the differences in perceptual saliency among the three consonants. Scheffe post-hoc tests showed that the NS group discriminated /n/ significantly better than both /t/ ($p = .0002$) and /z/ ($p < .0001$) and /t/ better than /z/ ($p = .0122$).

Figure 6-6: Native speaker d-prime scores for AX discrimination task employing real-word stimuli by Pair and Consonant.



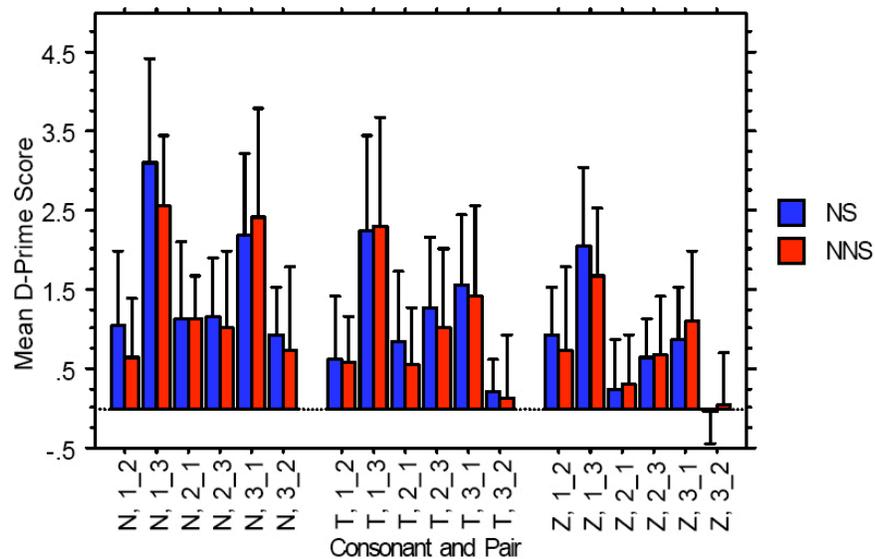
D-prime scores for the NNS group on real-word stimuli were then analyzed for each individual consonant in a factorial ANOVA analyzing the factors Consonant and Pair. This analysis revealed a main effect of Consonant ($F(2, 288) = 13.618, p < .0001$) as well as a main effect of Pair ($F(5, 288) = 30.186, p < .0001$), but no significant interaction ($F(10, 288) = 1.50, NS$). Post-hoc analyses of NNS d-prime scores showed that the group discriminated /n/ significantly better than both /t/ ($p = .0055$) and /z/ ($p < .0001$), but there was no difference between the discrimination of /t/ and /z/. These results are displayed below in Figure 6-7.

Figure 6-7: Non-native speaker d-prime scores for AX discrimination task employing real-word stimuli by Pair and Consonant.



D-prime scores for both participant groups were then compared for the same factors in a three-way ANOVA. This combined analysis revealed no significant effect of Participant Group ($F(1, 594) = 2.548$, NS), but did reveal an effect of Pair ($F(5, 594) = 71.041$, $p < .0001$) as well as a main effect of Consonant ($F(2, 594) = 37.704$, $p < .0001$). This analysis also showed one significant interaction between Pair and Consonant ($F(10, 594) = 2.916$, $p = .0014$), which again is attributable to the fact that the discrimination performance for both participant groups depends on the particular segment. Figure 6-8 below summarizes this analysis.

Figure 6-8: *D-prime scores for AX discrimination task employing real-word stimuli by Pair, Consonant and Participant Group.*



6.2.4.4 Reaction Time

Reaction times on the AX discrimination task were also analyzed for both participant groups. Response latencies are assumed to reflect the difficulty involved in the execution of a behavioral task in that the time incurred in the execution of the task is assumed to reflect cognitive load and/ or working memory resources — the less taxing the task, the less time is incurred in its execution. (See Lachman, Lachman & Butterfield 1979 for a review of reaction time theory.) In the case of an AX discrimination task, reaction time is taken indicate the degree of discrimination between stimuli. Studies investigating categorical perception (e.g. Pisoni & Tash 1974) have shown that participants are faster to decide that two stimuli are the *same* that are acoustically identical and from the same phonetic category than two pairs that are acoustically different, but yet still represent the same phonetic category. This finding predicts that in a study employing allophonic (within-category) variation such as the current study, the degree of acoustic difference between allophonic pairs should predict response latencies, i.e. longer response latencies should be observed for pairs separated by one degree on the

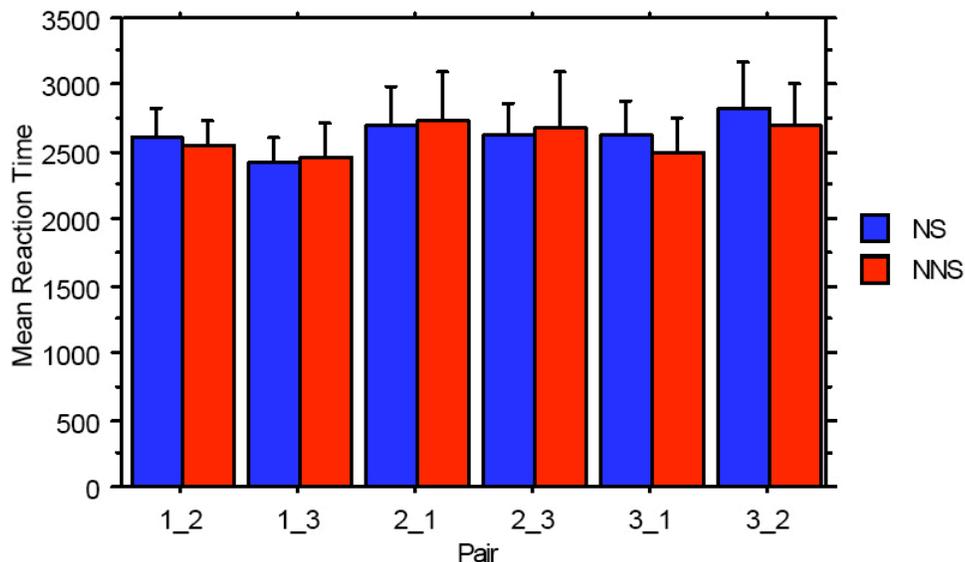
durational continuum in the current study, and shorter response latencies should be observed for pairs that are separated by two degrees on the durational continuum.

Differences in response latencies between the two participant groups are harder to predict. If a difference is observed between the NS and the NNS groups, this difference is predicted to be in line with previous research on the perception of liaison by native and non-native speakers of French which showed increased reaction times on the part of the NNS relative to the NS (Dejean de la Bâtie & Bradley 1995; Matter 1989). In general, longer reaction times in NNS participants relative to NS controls in behavioral tasks have been attributed to heavier cognitive load, possibly due to an increased demand for working memory resources.

Reaction times in the current study were measured as the time between the onset of the auditory stimulus to the participant's pressing of a response key. Mean reaction times across participants were calculated for each of the six different pairs. Following Ulrich and Miller (1994), cut-off values were established so as to remove outlying reaction times that could skew analyses. Reaction times below 200 ms and above 4000 ms were therefore discarded. Only correct responses for different pairs were analyzed.

A factorial ANOVA comparing mean reaction times across pairs and participant groups indicated a main effect of Pair ($F(5, 197) = 5.270, p = .0001$), but no effect of Participant Group ($F(1, 197) = .436, NS$) and no interaction ($F(5, 197) = .789, NS$). Post-hoc tests revealed that only the 1_3 pair was discriminated significantly faster than the other pairings as can be seen below in Figure 6-9.

Figure 6-9: Mean reaction times in milliseconds on AX discrimination task employing real-word stimuli by Pair and Participant Group.



The lack of a significant difference between the participant groups suggests that the non-native speakers did not experience a heavier cognitive load or increased demand on working memory than the native speakers in the processing of these ambiguous pairs.

6.2.5 Discussion: Experiment 1

The results of Experiment 1 suggest that, though both the physical differences and proportional change in duration between manipulated stimuli are substantial compared to what are considered to be thresholds of noticeability, sensitivity to segmental duration in environments of possible liaison is mitigated by other factors. The current results suggest that only manipulated stimuli separated by two degrees on the durational continuum are sufficiently different acoustically to be systematically distinguished by both participant groups. Mean accuracy rates for pairs separated by one degree were significantly below chance, indicating a significantly higher proportion of *same* responses for these different pairs and therefore a lack of consistent discrimination. In addition, d-prime scores, which are considered to be a more informative measure of discrimination than accuracy, largely followed the same pattern. For both participant groups, pairs separated by two degrees on the durational continuum showed significantly higher discrimination than one-degree differences according to d-prime analysis. Pairs separated by one degree were rarely associated with a d-prime score superior to 1, indicating a lack of discrimination. In addition, there was a great deal of variation across participants in both the NS and NNS groups as evidenced by the substantial range of both accuracy rates and d-prime scores for both groups, which we interpret as an indication of the difficulty of the task in that difficult tasks tend to generate a higher degree of individual variation.

At the beginning of the chapter, we proposed factors that could possibly render the discrimination of these durational differences more difficult than established JND thresholds would suggest. First, the durational differences in the current stimulus sample represent within-category variation, which has been shown to be less perceptually salient than variation which crosses categorical boundaries (e.g. Peperkamp et al. 2003). The durations of the manipulated segments used in this task fall within the range and distribution of the production sample discussed in Chapter Five, and therefore represent reasonable, though exaggerated, tokens of allophonic variation in spoken French. This fact may have rendered discrimination more challenging.

Second, we noted that variation in segmental duration in this task occurs phrase-medially in sequences of 2 – 4 syllables incorporating 4 – 9 phonemes. Before beginning the experiment, participants were instructed to indicate whether the two stimuli in each experimental trial were *same* or *different*. However, they were not given any indication as to what would constitute the basis for this qualification. This necessitated that participants attend to the entire signal in order to determine not only *where* differences lay, but along *which physical dimension* the variation occurred. Indeed, debriefing of several participants after the experiment revealed that even after hearing 648 trials over a 50-minute period, they were not able to pinpoint on which factor they had based their responses, suggesting a lack of saliency of these durational differences. Unfortunately, the current results do not allow us to pinpoint whether the relatively low sensitivity exhibited by both groups is due to difficulties associated with the discrimination of

allophonic variation or to the complexity of the signal, or to another as yet unidentified factor.

Regarding performance comparisons between the two participant groups, the current results fail to provide conclusive evidence of a difference in sensitivity between NS and NNS participants. An analysis of mean accuracy suggested that the NS group was significantly better at discriminating the durational differences than the NNS group, however this difference was no longer significant in an analysis of d-prime scores. Given that d-prime is considered to be a more sensitive measure of discrimination, the results suggest that the two groups are behaving similarly. Further supporting this conclusion, an analysis of response latencies suggests that NNS participants did not experience a heavier processing load than NS participants. Nonetheless, the amount of variation observed in both groups makes it difficult to ascertain whether the lack of significant difference between the two group's performance is attributable to nativelike behavior on the part of the NNS group or to the relatively noisy data obtained from both groups. Our tentative conclusion is that the NNS are performing in a nativelike manner, however we will return to this issue for further discussion below.

As predicted, both groups also showed differences in sensitivity to durational differences among the three consonants. Furthermore, the pattern of differences is in the same direction for both groups ($/n/ > /t/ > /z/$), suggesting that the two groups are sensitive to the same acoustic cues in the speech signal.

As we noted earlier, differences in discrimination among the three consonants could be attributable to two factors. The first possibility is that sensitivity to durational differences in liaison consonants is conditioned by language usage and reflects the respective frequency of occurrence of each consonant in liaison environments. However, data from corpus work has shown that $/z/$ occurs most frequently in environments of liaison followed by $/n/$, $/t/$, $/ʁ/$ and $/p/$ (Durand & Lyche 2008; Léon 1992), while the current results show that sensitivity to $/z/$ was significantly lower than $/n/$ and $/t/$. A statistical frequency account can therefore be rejected.

A more likely explanation for the observed differences among the consonants is that they represent an artifact of the particular production sample collected for this study. Given that perceptual saliency in this task follows the same pattern as the robustness of duration variation in the production sample (i.e. $/n/ > /t/ > /z/$; see §5.2.1), it appears that both participant groups are sensitive to this differentiation. While we maintain that the current production sample is generalizable to a larger population in terms of durational differences between LCs and ICs, further research would be needed to establish that the differences in robustness observed among the three consonants in the current sample are indicative of a general phenomenon.

We now discuss briefly an unexpected result that emerged from the current data, namely an effect of stimulus order. We observed that pairs in which the order of presentation was a shorter token followed by a longer token (e.g. 1_2, 1_3, 2_3) were discriminated better than pairs in which a longer token was followed by a shorter token (e.g. 2_1, 3_1, 3_2). The pair 1_3 which was discriminated better than 3_1 by both participant groups in mean accuracy rates ($p = .0609$) as well as d-prime scores ($p = .0033$). Similarly, the pair 2_3 was discriminated significantly better than 3_2 by both

participant groups (mean accuracy: $p = .0143$; d-prime scores: $p = .0207$). The same effect was not, however, observed for the pairs 1_2 and 2_1, where there was no significant difference in discrimination between the two pairs.

This result is puzzling if we consider the fact that these pairs are differentiated by the same absolute duration along the three-step durational continuum. In other words, the physical difference is the same between the two stimuli in a 1_3 pair as in a 3_1 pair. However, an analysis of the percentage change between the two stimuli in each pairing (see Table 6-2 above) can shed light on this effect. For example, the mean duration of all shortened consonants (1) in the current stimulus set is 63.21 ms, while the mean duration of all lengthened consonants (3) is 124.70 ms. Therefore, in the case of the stimulus pairs 1_3 and 3_1, the percentage change in the pairing 1_3, i.e. the proportion of the absolute durational difference between 1 and 3 relative to the absolute duration of the first member of the pair, 1, is an average increase of 104.16%; however in the pairing 3_1, i.e. the absolute durational difference between 1 and 3 relative to the absolute duration of the first member of the pair in this case, 3, the percentage change is an average decrease of only 49.58%. Given that a large body of research has established that the perceptual saliency of acoustic variation can be framed in proportions in addition to absolute values, an effect of stimulus order based on the relative change between the two tokens in each experimental trial proportion of change is plausible.

In order to test this hypothesis, we performed both simple regression and Spearman rank-order analyses between percentage change and both mean accuracy rates and d-prime scores for each pair, all of which revealed either significant effects or effects that just missed significance. These results are summarized in Table 6-7 below. Across all six different pairs, the percentage of durational difference from the first member of the pair to the second member of the pair predicted both mean accuracy rates and d-prime scores for both the NS and NNS groups. This finding lends further credence to the notion that the perception of duration is better represented by differential thresholds than by absolute values.

Table 6-7: *Correlation matrix: mean accuracy rates and d-prime scores on AX discrimination task employing real-word stimuli for NS and NNS participant groups as a function of percentage change within each token pair (e.g. 1_3, 3_1, 2_1, etc).*

	Value of r and p (simple regression)	Value of Rho and p (Spearman rank order)
NS accuracy * percent change	$r = .852, p = .0313$	$Rho = .829, p = .0639$
NS d-prime scores * percent change	$r = .930, p = .0071$	$Rho = .900, p = .0442$
NNS accuracy * percent change	$r = .811, p = .0500$	$Rho = .829, p = .0639$
NNS d-prime scores* percent change	$r = .870, p = .0241$	$Rho = .829, p = .0639$

In sum, Experiment 1 has demonstrated that durational differences that surface between LCs and ICs are perceptually salient only when these differences are greatly exaggerated; when differences are more in line with natural production values,

discrimination is significantly less acute. The two extremes of our durational continuum, which were discriminated relatively well, represent a difference of *more than* two standard deviations in our production sample. Recall that the target durations used in the manipulation of these stimuli were based not on the production means of the entire sample, but on the production means of each individual condition (i.e. liaison environments and C-initial environments). In other words, the target duration used for the shortened (liaison) version of each token was the mean duration of all instances of that consonant in the *liaison* environment minus one SD from that mean, while the target duration for the lengthened (word-initial) version of the consonant was the mean duration for that consonant in *word-initial* position plus one SD from that mean (see §5.4 above). Therefore the ends of the durational continuum are more than one standard deviation both above and below the mean production values across the entire production sample.

In our current results, stimuli differentiated by one degree on the three-step durational continuum were not distinguished well or at all. In a normal distribution of natural speech, the difference represented by one degree on the continuum would presumably fall *within* two standard deviations of mean production values. This suggests that these differences as they occur in natural speech may not represent a very robust processing cue. Furthermore, the large amount of variation among participants in both groups suggests that individual listeners may not make use of this variation in a systematic fashion.

In the next experiment, which employs non-word manipulated stimuli, we will investigate whether the removal of lexical information renders the perception of these durational differences more robust.

6.3 Experiment 1.1: AX Discrimination Task employing Non-Word Stimuli

Experiment 1.1 also consisted of an AX discrimination task employing stimuli in which the duration of pivotal consonants has been instrumentally manipulated, the only difference from Experiment 1 being that Experiment 1.1 employed non-word stimuli.

6.3.1 Predictions

As discussed above in §5.1.2, the use of non-words in addition to real words in the current experiment is motivated in order to investigate whether the removal of potential effects of lexical frequency and/or syntactic and semantic acceptability has an effect on the processing of ambiguous minimal pairs in spoken French. We foresee two possible outcomes in the processing of non-word stimuli relative to real-word stimuli. Differences in discrimination favoring non-word stimuli would suggest that processing occurs primarily at the auditory level plus an inhibitory effect of lexis on the processing of these pairs. Differences in discrimination favoring real-word stimuli would be consistent with a *word-superiority* view of processing. *Word-superiority* refers to a phenomenon by which recognition of phonemes is facilitated in real-word contexts relative to non-word contexts (e.g. Cutler, Mehler, Norris & Segui 1987; Eimas, Hornstein & Payton 1990; Frauenfelder, Segui & Dijkstra 1990; Pitt & Samuel 1995; Reicher 1969; Rubin, Turvey & van Gelder 1976). Lexical knowledge has also been shown to facilitate the categorization of ambiguous phonemes (e.g. Norris et al. 2003).

Word-superiority effects are usually framed within an interactive view of speech recognition (see §1.2 above), in which lexical information facilitates categorical decisions about the individual segments that make up the word. In the current study, an effect of lexical knowledge could plausibly render the durational differences more salient in real-word stimuli than in non-word stimuli.

6.3.2 Participants

Thirty-six participants who had not participated in Experiment 1 participated in Experiment 1.1. The control group was comprised of 18 NS of French (14 female, 4 male; mean age: 36.9 yrs; range: 20-60 yrs). All NS had studied languages other than their native French to varying degrees of proficiency, however no NS participant was bilingual from birth.

The experimental group consisted of 18 native speakers of English (13 female, 5 male; mean age: 41.6 yrs; range: 26-57 yrs). All NNS participants lived in France or a French-speaking country for a minimum of five years at the time of testing (mean: 9.8 yrs; range: 5 – 21 yrs). Average age of arrival in France was 31.9 years (range: 20 – 47.7 yrs). Average age of first exposure to French was 14.9 years (range: 9 – 30 yrs). The variety of English spoken by each participant was not controlled for and included 11 speakers of American English, 4 speakers of British English and 3 speakers of Australian English.

As in Experiment 1, participants were paid eight Euros to take part in the experiment. All had normal or corrected vision and none reported any hearing impairment. Handedness was not controlled for.

6.3.3 Stimuli

Stimuli consisted of non-word pairs drawn from the three-step continuum of manipulated phrases described in §5.4 above. As in Experiment 1, each token on the three-step continuum was paired with a duplicate version of the same token as well as with the other two manipulated versions of that token. This resulted in nine pairings for each of the 12 stimuli. Of the nine pairs, three were identical (1_1, 2_2, 3_3) and six were different (1_2, 1_3, 2_1, 2_3, 3_1, 3_2). Of the six different pairs, two pairs were separated by two degrees on the durational continuum (1_3, 3_1) and four were separated by one degree (1_2, 2_1, 2_3, 3_2).

6.3.4 Procedure

The procedure was identical to that of Experiment 1, though Experiment 1.1 employed pairs of manipulated non-word stimuli. In addition, participants were informed in the experimental instructions that the stimuli would consist of words that do not exist in French. (See Appendix VII.)

6.3.5 Results: Experiment 1.1

Two NS participants (NS 32 and NS 33) and one NNS participant (NNS 29) were removed from analysis in Experiment 1.1 due to the fact that they responded *same* to all

648 trials of the experiment. The following analyses therefore include 16 NS participants and 17 NNS participants.

As in Experiment 1, only responses for the six different pairs were included in analysis. All analyses are by subject.

6.3.5.1 Mean Accuracy

First, mean accuracy rates for NS participants for different stimulus pairs were calculated and are summarized below in Table 6-8. The table also includes the range and standard deviations of accuracy rates across participants, as well as the results of single sample t-tests indicating whether performance of each participant group is significantly different from chance (50%).

Table 6-8: Mean accuracy rates on AX discrimination task with non-word stimuli for each participant for each pair. Given at the bottom of the table are mean accuracy, range and standard deviations across participants as well as single sample t-test results comparing performance to chance (50%).

	1 2	1 3	2 1	2 3	3 1	3 2
NS 19	22	72	20	53	25	12
NS 20	2	52	13	20	60	33
NS 21	0	3	0	3	0	0
NS 22	88	91	89	89	89	86
NS 23	18	75	23	65	68	20
NS 24	10	53	7	42	30	8
NS 25	2	23	0	7	18	3
NS 26	35	93	35	73	70	48
NS 27	3	7	5	0	7	2
NS 28	10	57	30	28	70	32
NS 29	0	33	13	20	37	27
NS 30	3	80	35	42	80	35
NS 31	22	72	20	53	25	12
NS 32	removed from analysis					
NS 33	removed from analysis					
NS 34	25	75	27	40	52	13
NS 35	17	78	10	52	22	8
NS 36	50	90	40	77	77	43
RANGE	0 - 88	3 - 93	0 - 89	0 - 89	0 - 89	0 - 86
MEAN ACC	19.19	59.63	22.94	41.50	45.63	23.88
(SD)	(23.11)	(29.08)	(21.64)	(27.01)	(28.34)	(22.35)
DIFF FROM CHANCE	P < .0001	NS	P = .0002	NS	NS	P = .0003

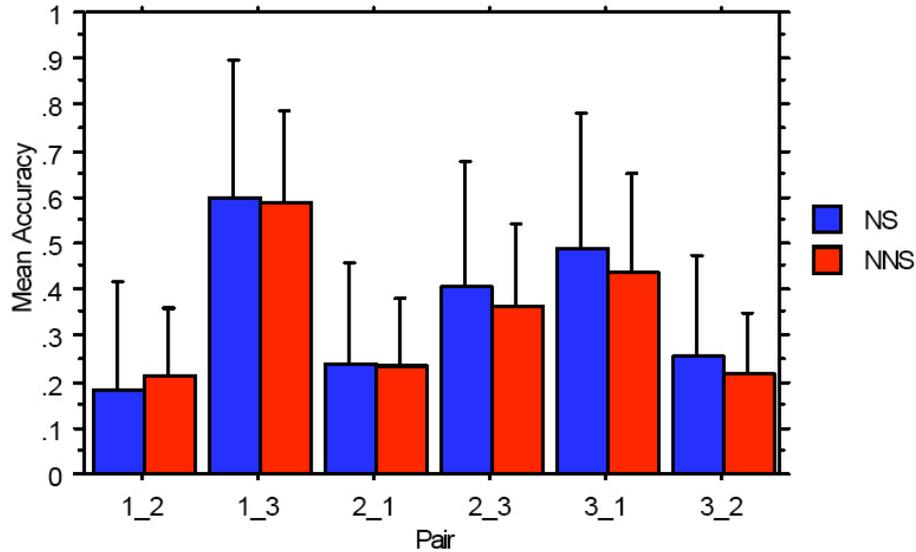
Then, mean accuracy rates for NNS participants for different stimulus pairs were calculated and are summarized below in Table 6-8.5. The table also includes the range and standard deviations of accuracy rates across participants, as well as the results of single sample t-tests indicating whether performance of each participant group is significantly different from chance (50%).

Table 6-8.5: Mean accuracy rates on AX discrimination task with non-word stimuli for each participant for each pair. Given at the bottom of the table are mean accuracy, range and standard deviations across participants as well as single sample t-test results comparing performance to chance (50%).

	1 2	1 3	2 1	2 3	3 1	3 2
NNS 19	43	75	38	62	70	35
NNS 20	23	43	13	25	42	25
NNS 21	2	27	13	5	27	3
NNS 22	12	75	18	37	57	20
NNS 23	25	87	30	48	70	37
NNS 24	2	22	7	10	15	2
NNS 25	28	70	43	47	57	27
NNS 26	2	50	5	22	25	3
NNS 27	42	93	47	62	78	27
NNS 28	36	58	33	56	53	35
NNS 29	removed from analysis					
NNS 30	20	57	12	28	32	27
NNS 31	23	58	43	35	70	35
NNS 32	42	82	40	55	58	43
NNS 33	17	47	15	33	27	18
NNS 34	3	48	5	17	17	3
NNS 35	32	63	18	50	23	15
NNS 36	12	43	20	27	27	13
RANGE	2 – 43	22– 93	5 – 47	5 – 62	15– 78	2 - 43
MEAN ACC	21.41	58.71	23.53	36.41	44.00	21.65
(SD)	(14.51)	(19.96)	(14.54)	(17.75)	(21.19)	(13.44)
DIFF FROM CHANCE	P < .0001	NS	P < .0001	P = .0061	NS	P < .0001

Mean accuracy rates for both groups were submitted to ANOVAs with the factors Pair (six levels) and Participant Group (two levels). A factorial ANOVA revealed a main effect of Pair ($F(5, 186) = 17.34, p < .0001$), but no effect of Participant Group ($F(1, 186) = .385, NS$) and no interaction ($F(5, 186) = .176, NS$). These results are summarized in Figure 6-10 below.

Figure 6-10: Mean accuracy on AX discrimination task employing non-words by Pair and Participant Group.



Again, as was observed in responses to real-word stimuli in Experiment 1, pairs separated by two degrees in the durational continuum show significantly higher accuracy rates. Table 6-9 below summarizes the results of Scheffe post-hoc comparisons of individual pairs.

Table 6-9: Results of Scheffe post-hoc tests showing discrimination among different pairs of non-word stimuli according to mean accuracy rates for NS and NNS participants.

Pair-by-pair comparison	Native Speakers: Value of P	Non-native Speakers: Value of P
1_2, 1_3	.0065	<.0001
1_2, 2_1	NS	NS
1_2, 2_3	NS	NS
1_2, 3_1	NS	.0161
1_2, 3_2	NS	NS
1_3, 2_1	.0218	<.0001
1_3, 2_3	NS	.0184
1_3, 3_1	NS	NS
1_3, 3_2	.0330	<.0001
2_1, 2_3	NS	NS
2_1, 3_1	.0091	.0410
2_1, 3_2	NS	NS
2_3, 3_1	NS	NS
2_3, 3_2	NS	NS
3_1, 3_2	NS	.0179

6.3.5.2 D-prime

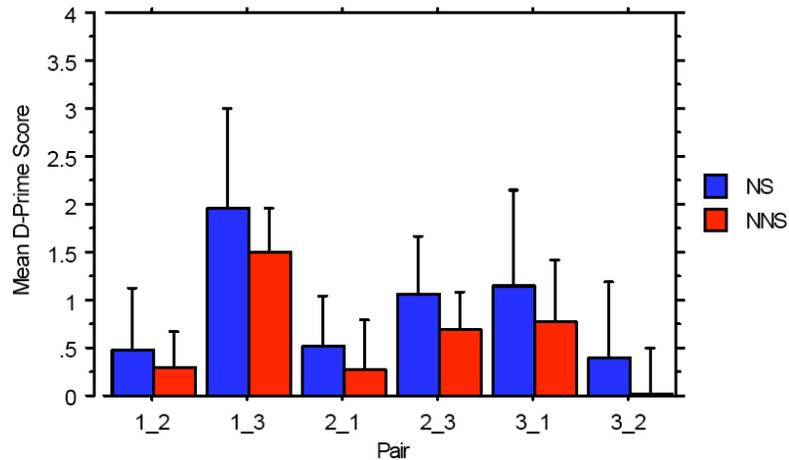
Next, d-prime scores for responses to non-word stimuli were calculated for each participant group for each of the six different pairs. See Table 6-10 below for a summary.

Table 6-10: *D-prime scores on AX discrimination task employing non-word stimuli for each participant for each pair. Mean scores, standard deviations and range across participants are given at the bottom of the table.*

	1 2	1 3	2 1	2 3	3 1	3 2		1 2	1 3	2 1	2 3	3 1	3 2
NS 19	-0.03	1.32	0.28	1.20	-0.12	-0.62	NNS 19	0.53	1.38	0.19	0.80	0.68	-0.23
NS 20	0.27	2.38	1.20	1.48	2.00	1.31	NNS 20	0.39	0.95	-0.48	-0.03	0.68	0.20
NS 21	-0.45	0.00	0.00	0.45	-0.58	-0.58	NNS 21	-0.30	1.14	0.62	0.11	0.94	-0.33
NS 22	0.84	2.43	0.60	1.73	1.17	-0.14	NNS 22	0.05	1.90	0.08	0.66	1.40	0.38
NS 23	1.04	2.40	0.85	2.12	1.03	0.15	NNS 23	0.55	2.35	0.28	0.76	1.17	0.31
NS 24	0.27	1.59	-0.45	0.41	1.41	0.45	NNS 24	0.27	1.55	0.85	1.04	1.29	0.27
NS 25	0.53	2.39	0.20	1.20	0.88	0.31	NNS 25	-0.03	1.08	0.56	0.66	1.05	0.27
NS 26	0.45	0.85	0.68	0.00	0.85	0.27	NNS 26	-0.30	1.75	0.11	0.98	0.73	-0.48
NS 27	-0.06	1.40	0.35	0.30	1.26	0.27	NNS 27	0.71	2.39	0.66	1.04	1.42	0.03
NS 28	-0.58	1.31	0.75	1.04	1.99	1.71	NNS 28	0.41	0.97	0.33	0.92	0.43	-0.03
NS 29	0.45	3.17	1.17	1.35	2.72	1.50	NNS 29	removed from analysis					
NS 30	removed from analysis						NNS 30	0.24	1.26	-0.44	0.16	0.18	0.03
NS 31	removed from analysis						NNS 31	1.01	1.95	1.70	1.50	2.28	1.37
NS 32	1.65	3.00	1.27	1.63	1.60	0.43	NNS 32	0.16	1.27	0.16	0.54	0.23	-0.15
NS 33	1.37	3.10	-0.16	1.18	-0.19	-0.82	NNS 33	0.13	1.01	-0.30	0.30	0.38	0.08
NS 34	1.41	2.69	0.33	1.32	1.73	0.82	NNS 34	-0.13	1.70	-0.09	0.60	0.39	-0.54
NS 35	0.14	0.30	0.19	0.19	-0.06	-0.20	NNS 35	0.94	1.74	0.21	1.13	-0.56	-0.86
NS 36	0.45	3.17	1.17	1.35	2.72	1.50	NNS 36	0.23	1.23	0.15	0.38	0.47	-0.05
MEAN	0.43	1.75	0.47	0.94	1.03	0.35	MEAN	0.29	1.51	0.27	0.68	0.77	0.02
	-0.58	0.00	-0.45	0.00	-0.58	-0.82		-0.30	0.95	-0.48	-0.03	-0.56	-0.86
RANGE	-	-	-	-	-	-	RANGE	-	-	-	-	-	-
	1.65	3.17	1.27	2.12	2.72	1.71		1.01	2.39	1.70	1.50	2.28	1.37
SD	0.64	1.02	0.52	0.61	1.00	0.79	SD	0.38	0.46	0.52	0.41	0.64	0.48

Contrary to the analysis of mean accuracy rates, analysis of d-prime scores did reveal a significant effect of Participant Group ($F(1,186) = 13.670, p = .0003$), indicating that the NS group showed significantly more sensitivity to durational differences among non-word stimuli than the NNS group. An effect of Pair was also observed ($F(5,186) = 24.591, p < .0001$), however no interaction was found ($F(5, 186) = .179, NS$). Figure 6-11 below summarizes this analysis.

Figure 6-11: *D*-prime scores for AX discrimination task employing non-word stimuli by Pair and Participant Group.



A series of Scheffe post-hoc analyses revealed that the two pairs which represented two degrees of separation in the durational continuum were discriminated better than most pairs representing one degree of separation. Table 6-11 below summarizes these results.

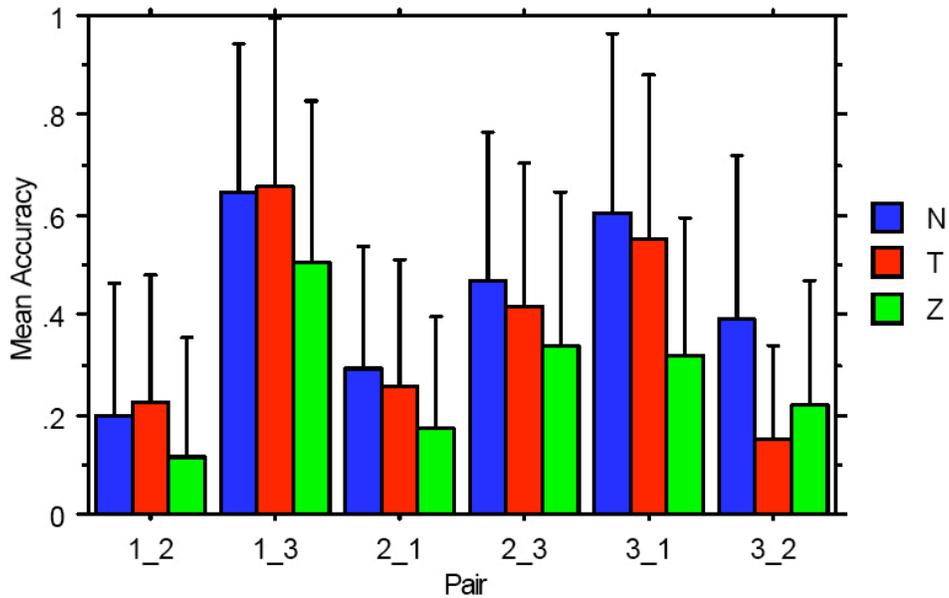
Table 6-11: *Results of Scheffe post-hoc tests showing discrimination among different pairs of non-word stimuli according to d-prime scores for NS and NNS participants.*

Pair-by-pair comparison	Native Speakers: Value of P	Non-native Speakers: Value of P
1_2, 1_3	.0001	<.0001
1_2, 2_1	NS	NS
1_2, 2_3	NS	NS
1_2, 3_1	NS	NS
1_2, 3_2	NS	NS
1_3, 2_1	.0002	<.0001
1_3, 2_3	(.0691)	.0005
1_3, 3_1	NS	.0034
1_3, 3_2	NS	<.0001
2_1, 2_3	NS	NS
2_1, 3_1	NS	NS
2_1, 3_2	NS	NS
2_3, 3_1	NS	NS
2_3, 3_2	NS	.0122
3_1, 3_2	NS	.0023

6.3.5.3 Results by Consonant

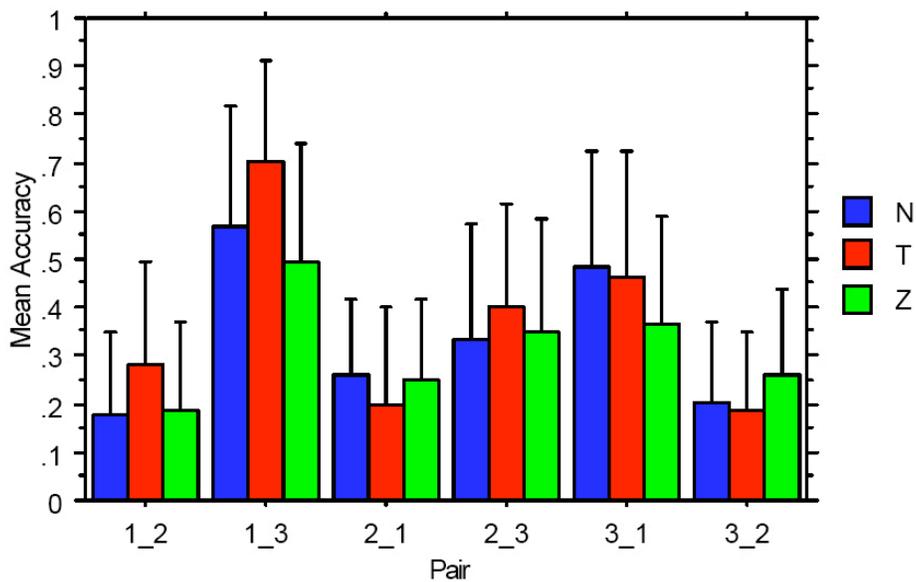
We also wanted to investigate whether sensitivity to durational variation differed among the three consonants in non-word stimuli. First, mean accuracy scores for each of the three segments for the NS group were submitted to ANOVAs which revealed a main effect of Consonant ($F(2, 270) = 7.28, p = .0008$) as well as a main effect of Pair ($F(5, 270) = 16.07, p < .0001$), but no interaction ($F(10, 270) = .781, NS$). As was observed in Experiment 1, Scheffe post-hoc tests showed that the NS group discriminated stimuli containing /z/ worse than stimuli containing both /n/ ($p = .0010$) and /t/ ($p = .0622$). There was no difference in discrimination between /n/ and /t/. These results are displayed graphically in Figure 6-12.

Figure 6-12: Native speaker mean accuracy on AX discrimination task employing non-word stimuli by Pair and Consonant.



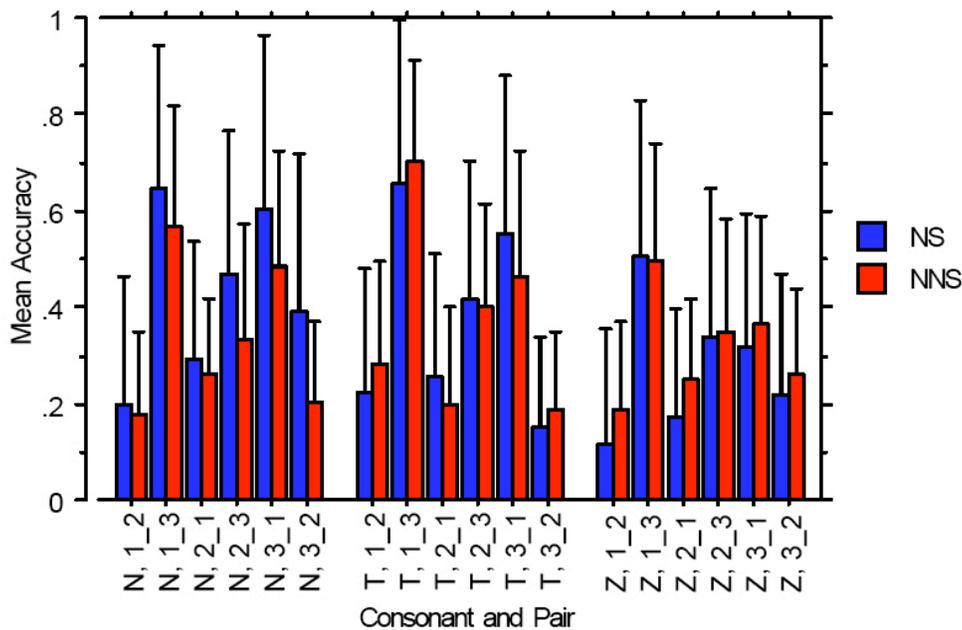
Contrary to the NS group, the NNS group showed no effect of Consonant for non-word stimuli ($F(2, 288) = 1.82$, NS), but did show a main effect of Pair ($F(5, 288) = 26.55$, $p < .0001$) and no interaction ($F(10, 288) = 1.40$, NS). The NNS group did not show significantly more sensitivity in the discrimination of any of the three consonants. These results are displayed below in Figure 6-13.

Figure 6-13: *Non-native speaker mean accuracy on AX discrimination task employing non-word stimuli by Pair and Consonant.*



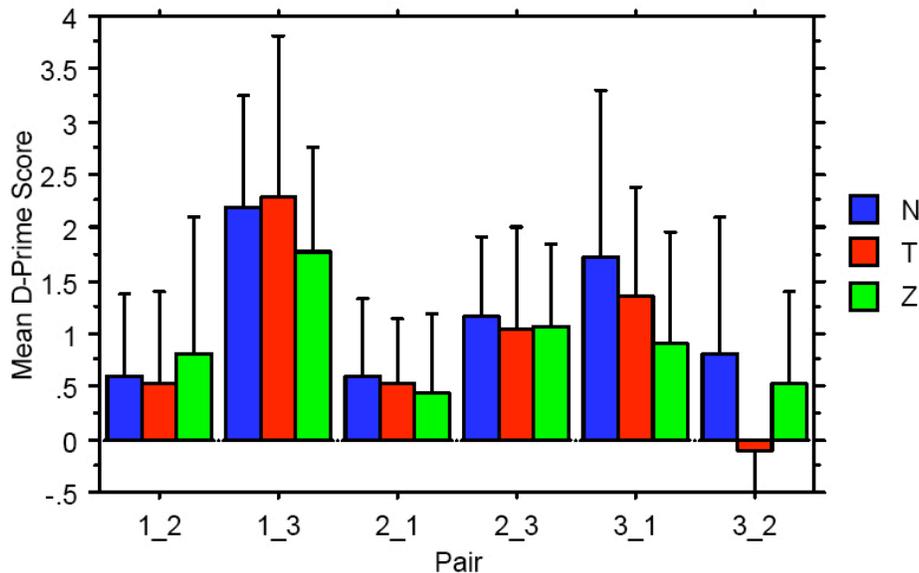
Mean accuracy rates for both participant groups were then compared for the same factors in a three-way ANOVA. This combined analysis revealed no effect of Participant Group ($F(1, 558) = .894$, NS), but did reveal a main effect of Pair ($F(5, 558) = 39.518$, $p < .0001$) as well as a main effect of Consonant ($F(2, 558) = 7.283$, $p = .0008$). This analysis also revealed a significant interaction between Consonant and Participant group ($F(2, 558) = 3.820$, $p = .0225$), indicating that the effect of Consonant was not equivalent between the two groups. See Figure 6-14 below.

Figure 6-14: Mean accuracy rates on AX discrimination task employing non-word stimuli by Pair, Consonant and Participant Group.



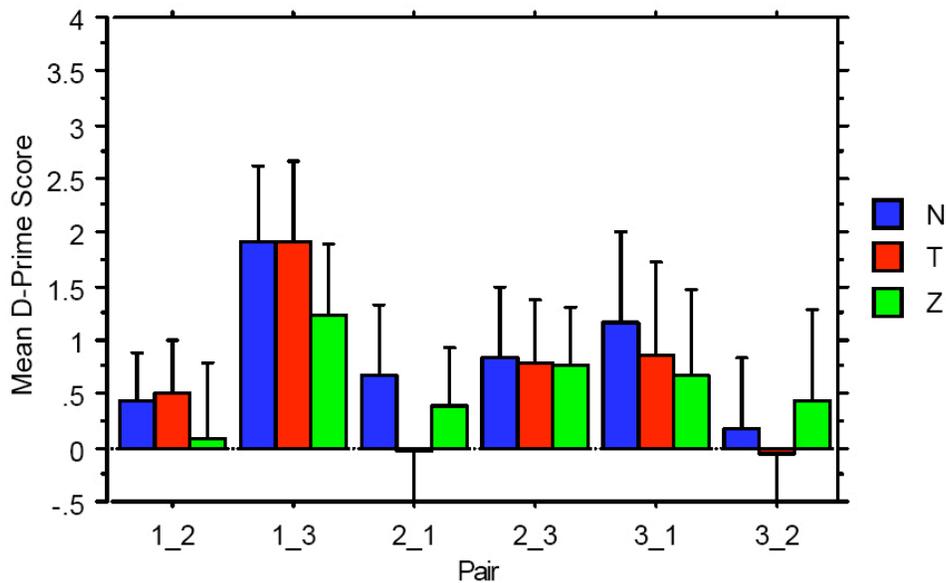
D-prime scores for non-word stimuli were also analyzed for each individual consonant for the NS group and were submitted to ANOVAs with the factors Consonant and Pair, which revealed no significant effect of Consonant ($F(2, 270) = 1.87, NS$), but did reveal a main effect of Pair ($F(5, 270) = 19.48, p < .0001$). There was no significant interaction ($F(10, 270) = 1.15, NS$). While a significant difference in discrimination among the three consonants was observed in an analysis of mean accuracy for the NS group, an analysis of d-prime scores showed no such effect. All three consonants were discriminated equally well, as seen below in Figure 6-15.

Figure 6-15: *Native speaker d-prime scores for AX discrimination task employing non-word stimuli by Pair and Consonant.*



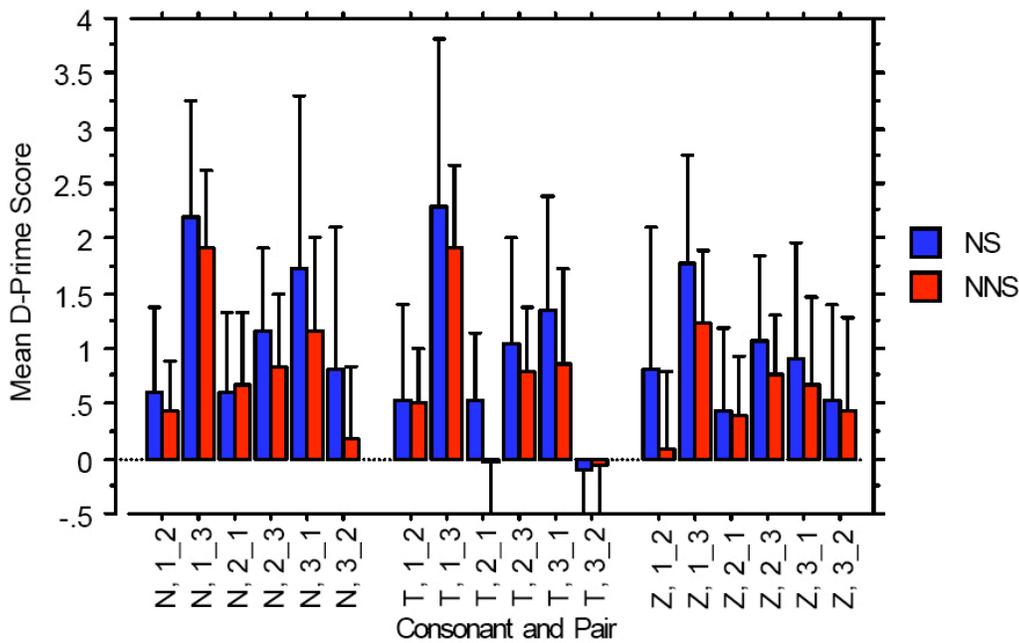
D-prime scores on non-word stimuli were also analyzed for each individual consonant for the NNS group in a factorial ANOVA analyzing the factors Consonant and Pair. This analysis revealed a main effect of Consonant ($F(2, 288) = 4.19, p = .016$) as well as a main effect of Pair ($F(5, 288) = 33.09, p < .0001$) and a significant interaction ($F(10, 288) = 2.34, p = .011$). A series of post-hoc Scheffe tests analyzing d-prime scores revealed that the NNS group discriminated /n/ significantly better than /z/ ($p = .0220$) but not better than /t/, nor was there a significant difference between /t/ and /z/. These results are presented below in Figure 6-16.

Figure 6-16: *Non-native speaker d-prime scores for AX discrimination task employing non-word stimuli by Pair and Consonant.*



D-prime scores for both participant groups were then compared for the same factors in a three-way ANOVA, which revealed main effects for all three factors: Participant Group ($F(1, 558) = 18.348, p < .0001$), Pair ($F(5, 558) = 46.537, p < .0001$) and Consonant ($F(2, 558) = 5.221, p = .0057$). This analysis also showed one significant interaction between Pair and Consonant ($F(10, 558) = 2.155, p = .0191$) which can be attributed to differences in sensitivity among the three segments. Figure 6-17 below presents this analysis graphically.

Figure 6-17: *D-prime* scores for AX discrimination task employing non-word stimuli by Pair, Consonant and Participant Group.



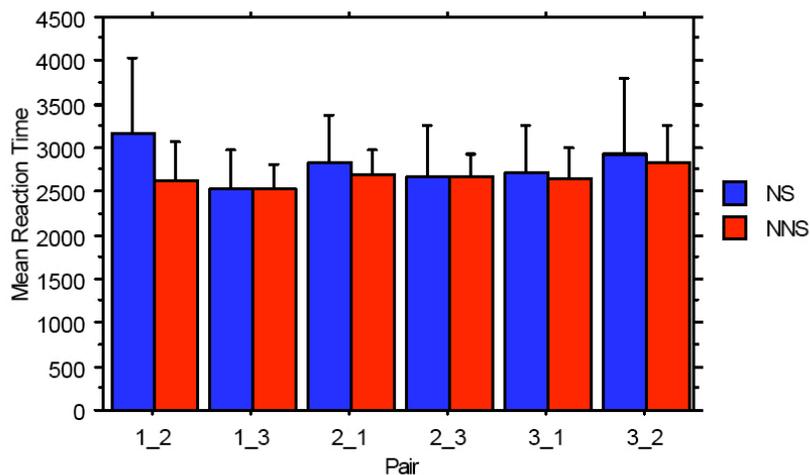
6.3.5.4 Reaction Time

Mean reaction times for correct responses to non-word stimuli for both participant groups were also analyzed, which revealed a significant effect of Pair ($F(5,171) = 2.333, p = .0444$), and an effect of Participant Group that fell just short of significance ($F(1,171) = 3.484, p = .0637$). No interaction was observed ($F(5,171) = 1.094, NS$). This

analysis indicates that the NS group had slightly longer reaction times (mean 2797.16, SD 662.94) than the NNS group (mean 2666.92, SD 350.46). See Figure 6-18 below.

As in Experiment 1, the lack of a robust difference in response latencies between the participant groups suggests that the NNS group did not experience a heavier cognitive load or increased demand on working memory than the NNS group in the processing of these ambiguous pairs.

Figure 6-18: Mean reaction times in milliseconds on AX discrimination task employing non-word stimuli by Pair and Participant Group.



6.3.6 Discussion: Experiment 1.1

The results of Experiment 1.1 are consistent with those of Experiment 1, however lower overall accuracy and d-prime scores were observed for both participant groups. Moreover, results show even more variation among participants as measured by both accuracy and d-prime. As predicted, discrimination was significantly better for both groups for those pairs representing a difference of two degrees on the durational continuum than for pairs representing a difference of one degree. Only the 1_3 stimulus pair was discriminated above chance levels by both participant groups in an analysis of accuracy rates. Similarly, this stimulus pair was the only to exhibit a d-prime score above 1 for both participant groups. The results of Experiment 1.1 suggest that, though both the physical differences and relative change in duration between these manipulated stimuli are substantial compared to what are considered to be thresholds of noticeability, sensitivity seems to be decreased even more in the non-word stimulus set.

Concerning differences between the two participant groups, results were again mixed. An analysis of mean accuracy rates revealed no significant difference between the two groups, while an analysis of d-prime scores did reveal a significant difference, with the NS group discriminating stimuli significantly better than the NNS group. Again, since d-prime scores are considered to be more indicative of discrimination than simple

accuracy, this would lead us to the conclusion that the NS group is more sensitive to durational variation in this environment in non-words.

Results pertaining to sensitivity to individual consonants were also mixed. In a mean accuracy analysis, the NS participant group discriminated non-word stimuli containing /n/ better than both /t/ and /z/. The NNS group, however, showed no such difference in mean accuracy among the three consonants. When d-prime scores were analyzed, however, the reverse pattern was observed. The NS group showed no difference in discrimination among the consonants, while the NNS group discriminated /n/ better than both /t/ and /z/.

Regarding the stimulus order effect that was observed in Experiment 1, a similar effect was observed in Experiment 1.1, though also to a somewhat lesser extent. Significant differences in discrimination between pairs in which the order of presentation was a shorter token followed by a longer token (e.g. 1_2, 1_3, 2_3) and pairs in which a longer token was followed by a shorter token (e.g. 2_1, 3_1, 3_2) were only observed in d-prime scores (see Table 6-11 above). Differences were not observed in an analysis of mean accuracy rates. To again test the hypothesis that the percentage change between stimulus pairs predicts discrimination sensitivity for the non-word stimuli, we performed both simple regression and Spearman rank-order analyses to investigate possible correlations between percentage change for each pair and both mean accuracy rates and d-prime scores. The percentage of durational difference between the two members of each stimulus pair tended to predict both mean accuracy rates and d-prime scores for both the NS and NNS groups, but to a lesser degree than was observed in Experiment 1. These results are summarized in Table 6-12 below.

Table 6-12: *Correlation matrix: mean accuracy rates and d-prime scores on AX discrimination task employing non-word stimuli for NS and NNS participant groups as a function of percentage change within each token pair (e.g. 1_3, 3_1, 2_1, etc).*

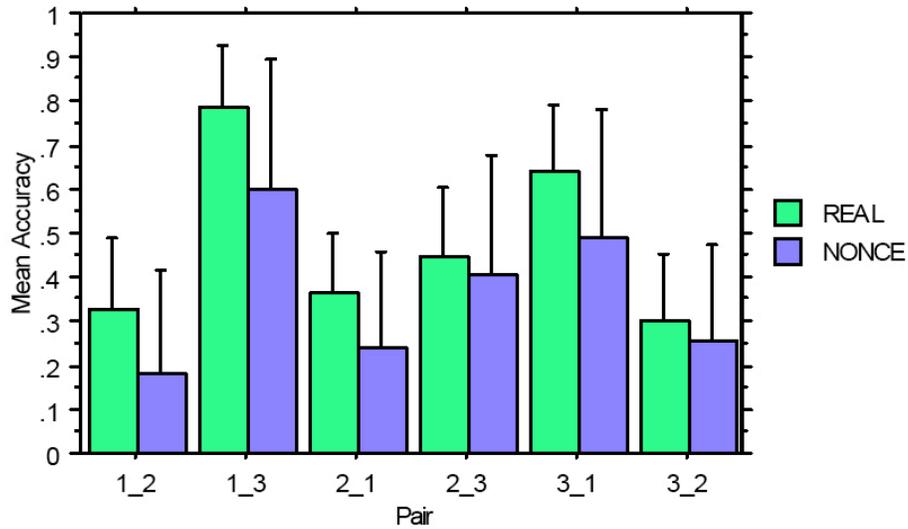
	Value of r and p (simple regression)	Value of Rho and p (Spearman rank order)
NS accuracy * percent change	r = .765, p = .0766	Rho = .600, NS
NS d-prime scores * percent change	r = .880, p = .0208	Rho = .829, p = .0639
NNS accuracy * percent change	r = .839, p = .0367	Rho = .657, NS
NNS d-prime scores* percent change	r = .906, p = .0128	Rho = .943, p = .0350

6.4 Discussion: Experiments 1 and 1.1 (real versus non-word stimuli)

As noted above, we observed that discrimination was lower overall for non-word items than for real lexical items for both participant groups. We therefore conducted further analyses to investigate whether differences between Experiments 1 (real-word stimuli) and 1.1 (non-word stimuli) were statistically significant.

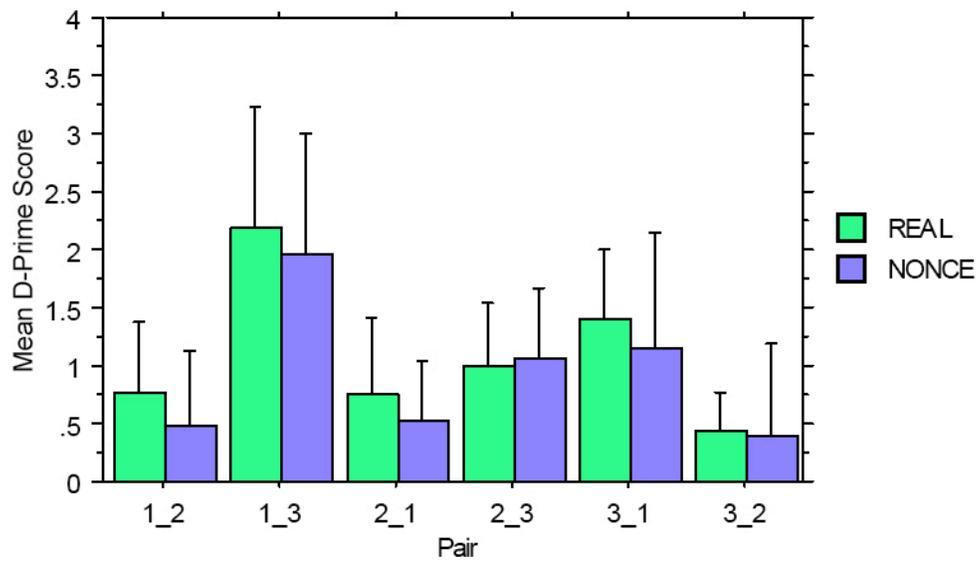
We first compared mean accuracy rates for the two NS groups. This analysis showed that the NS group tested on real-word stimuli in Experiment 1 performed significantly better than the NS group tested on non-word stimuli in Experiment 1.1 ($F(1, 192) = 15.896, p < .0001$) as shown below in Figure 6-19.

Figure 6-19: Mean accuracy rates for native speaker group tested on real-word stimuli in Experiment 1 and native speaker group tested on non-word stimuli in Experiment 1.1 by Pair and Word Type.



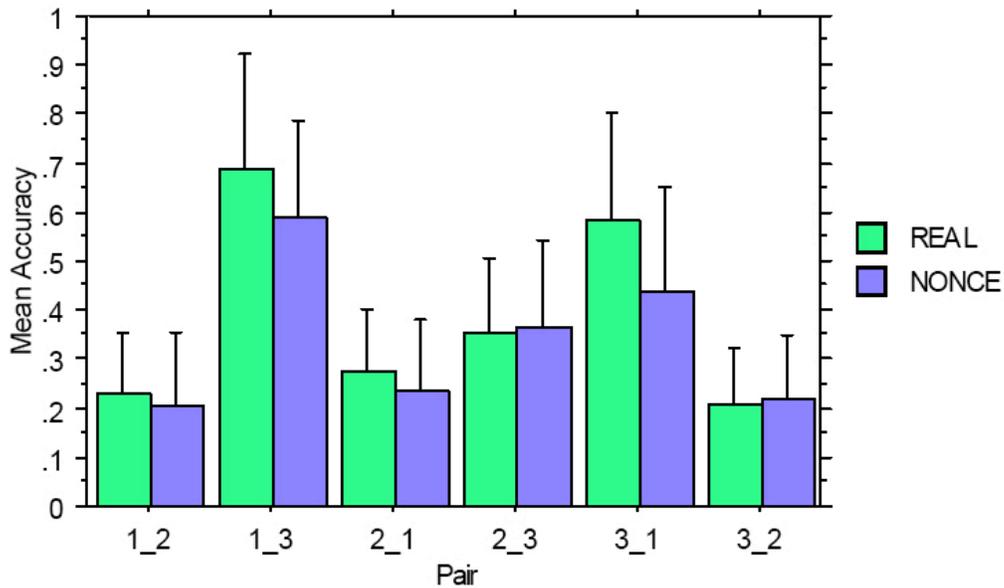
However, this difference was not significant when d-prime scores were analyzed for the two NS groups tested in Experiments 1 and 1.1 ($F(1,192) = 2.201$, NS) as shown below in Figure 6-20.

Figure 6-20: D-prime scores for native speaker group tested on real-word stimuli in Experiment 1 and native speaker group tested on non-word stimuli in Experiment 1.1 by Pair and Word Type.



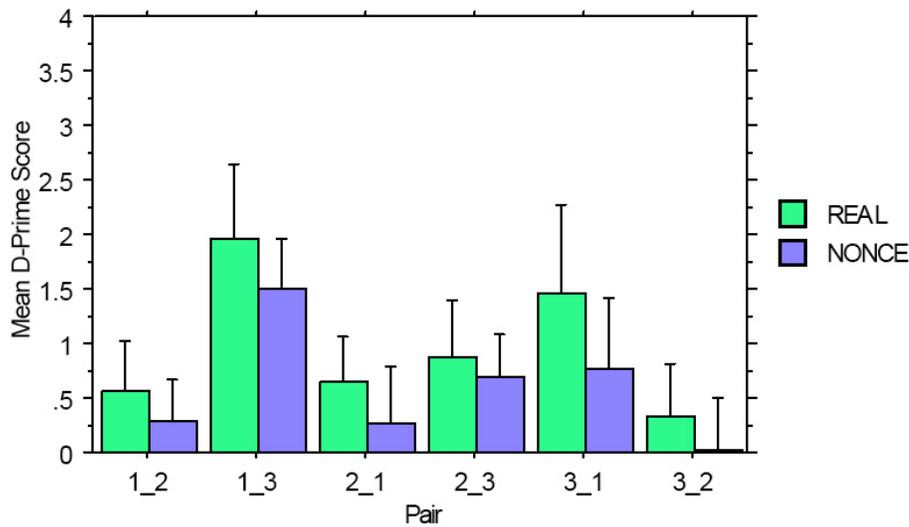
Regarding the two NNS groups, an analysis of mean accuracy rates revealed that NNS participants tested on real-word stimuli (Experiment 1) also performed significantly better than NNS participants tested on non-word stimuli (Experiment 1.1) ($F(1,193) = 4.038, p = .0459$) as Figure 6-21 shows below.

Figure 6-21: Mean accuracy rates for non-native speaker group tested on real-word stimuli in Experiment 1 and non-native speaker group tested on non-word stimuli in Experiment 1.1 by Pair and Word Type.



When d-prime scores were compared between the two NNS groups, the effect of word type was even stronger ($F(1,192) = 25.849, p < .0001$). The NNS participants tested on real-word stimuli in Experiment 1 discriminated stimuli significantly better than NNS participants tested on non-word stimuli in Experiment 1.1. See Figure 6-22 below.

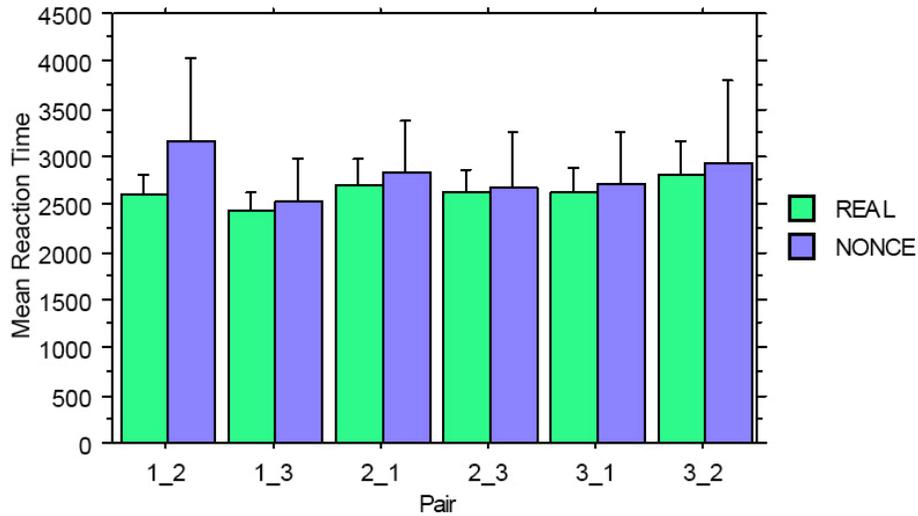
Figure 6-22: *D*-prime scores for non-native speaker group tested on real-word stimuli in Experiment 1 and non-native speaker group tested on non-word stimuli in Experiment 1.1 by Pair and Word Type.



6.4.1 Reaction Time

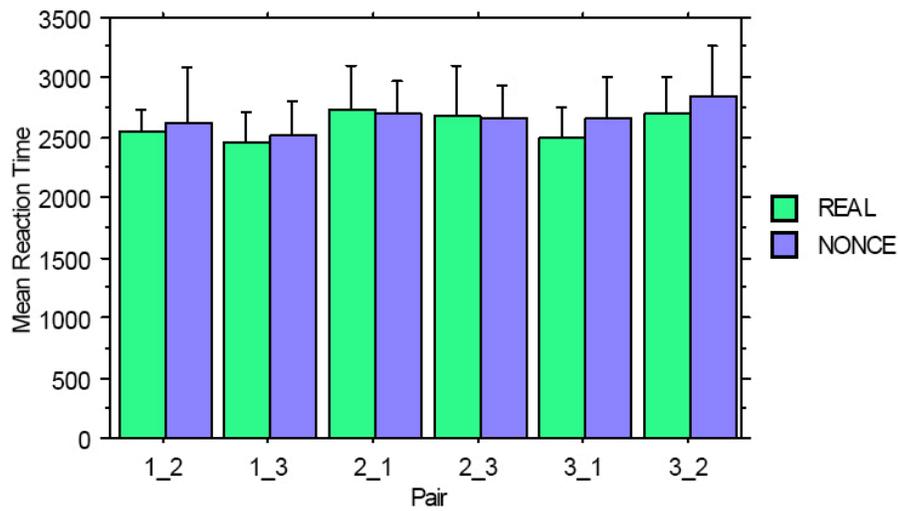
A comparison of mean reaction times in Experiments 1 and 1.1 was also performed. As in previous analyses, responses below 200 ms and above 4000 ms were discarded. An analysis of reaction times revealed a significant difference between the NS group tested on real-word stimuli in Experiment 1 and the NS group tested on non-word stimuli in Experiment 1.1 ($F(1,178) = 6.553, p = .0113$). Response latencies for NS participants tested on real-word stimuli were significantly shorter (mean 2632.41 ms) than for NS participants tested on non-word stimuli (mean 2797.16 ms). Note as well that much more variation in response latencies was observed for responses to non-word stimuli as evidenced by error bars in Figure 6-23 below.

Figure 6-23: Mean response latencies for native speaker group tested on real-word stimuli in Experiment 1 and native speaker group tested on non-word stimuli in Experiment 1.1 by Pair and Word Type.



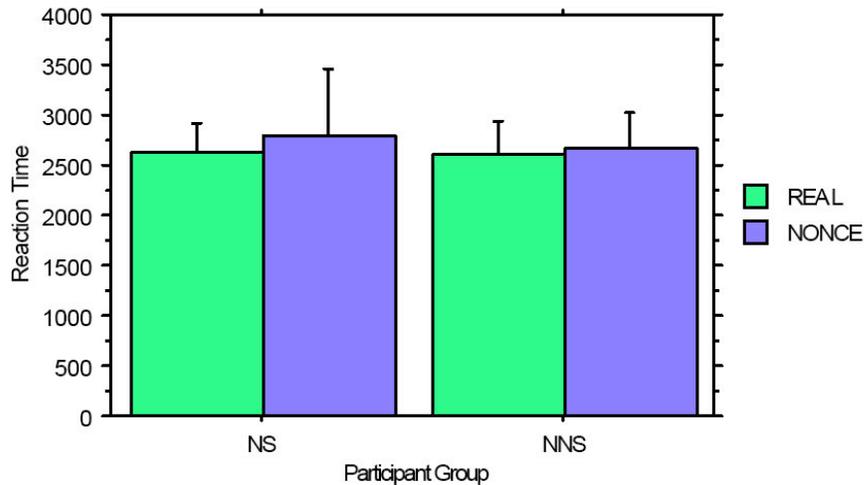
Mean reaction times were also compared between the NNS group tested on real lexical items in Experiment 1 and the NNS group tested on non-words in Experiment 1.1, which revealed no significant difference between the two Word Types ($F(1,190) = 1.978$, NS) as Figure 6-24 indicates below. Mean response latencies for NNS participants tested on real-word stimuli were 2603.48 ms while mean response latencies for NNS participants tested on non-word stimuli were 2666.92 ms.

Figure 6-24: Mean response latencies for non-native speaker group tested on real-word stimuli in Experiment 1 and non-native speaker group tested on non-word stimuli in Experiment 1.1 by Pair and Word Type.



Mean reaction times for all four participant groups were then analyzed, which revealed a main effect of Word Type ($F(1,388) = 7.313, p = .0071$) and an effect of Participant Group that fell just short of significance ($F(1,388) = 3.558, p = .0600$). This analysis revealed no significant interaction between the two factors ($F(1,388) = 1.442, NS$). This analysis is presented in Figure 6-25 below.

Figure 6-25: Mean response latencies for non-native speaker groups in Experiments 1 and 1.1 and native speaker groups in Experiments 1 and 1.1 by Word Type (real words and non-words).



An analysis of mean accuracy showed that discrimination of real-word stimuli was significantly better than non-word stimuli for both native and non-native speakers. An analysis of d-prime scores, however, showed a significant difference only for the two NNS groups. These results would lead us to reject the hypothesis that sensitivity to durational differences in environments of liaison is based purely on auditory processing. Manipulated consonants in both the real words and the non-words were based on the same production sample and the same target durations, and therefore the physical differences and proportions of change between items in each pair were identical for both word types. Therefore, if perceptual saliency were based purely on acoustic information, we would expect no differences in perceptual performance to arise between the two stimulus types for either NS or NNS participants.

The fact that differences favoring real-word stimuli did arise points to an effect of lexical knowledge on the saliency of durational differences in environments of possible liaison. In §6.3.1 above we discussed the possibility of *word-superiority* effects in the processing of these two word types, in which the categorization of phonemes is facilitated by lexical knowledge in real-word contexts relative to non-word contexts. Though the current task did not involve categorical decisions about phonemes, it could be argued that lexical knowledge increases sensitivity to within-category acoustic variation as well as categorical perception. As we will discuss in the following chapter, durational variation in environments of liaison influences categorical decisions about words in possible liaison environments, i.e. whether a lexical item is vowel- or consonant-initial in cases of ambiguity. Therefore it is plausible that lexical information associated with the real-word stimuli in the AX task improved sensitivity to the durational differences between segments.

Differences in response latency were also observed between the two NS groups; reaction times were significantly longer for the NS group tested on non-words than for

the NS group tested on real-words. This result is in line with identification tasks showing that response latencies are generally longer to reject non-words than to accept real words (e.g. Rubenstein et al. 1971). This suggests that the processing of non-word items incurs additional processing costs relative to existing real-word representations in the mental lexicon.

In sum, the differences in both discrimination sensitivity and reaction times between real- and non-word stimuli lead us to conclude that the processing of these durational differences is not occurring at a purely acoustic level, but rather that higher-level lexical processes are tapped in the perception of these durational differences.

6.4.2 NS versus NNS performance

The results of Experiments 1 and 1.1 taken as a whole fail to show a highly consistent pattern of differences between the NS and NNS participant groups. However, when mean accuracy and d-prime scores are compared across Experiments 1 and 1.1 a trend does emerge. Where there were significant differences between the two groups, they were consistently in the same direction, namely the NS group performed better. This would lead us to believe that perceptual saliency in the case of durational variation in possible liaison environments is indeed conditioned by language experience. As we observed in §6.2.5, however, the degree of variability across both participant groups makes it difficult to deduce whether the NNS group is behaving in a nativelike manner or whether the similarities between the two group's performance is attributable to the relatively noisy data obtained from both groups.

Nonetheless, we consider for the moment the possibility that sensitivity to durational variation in liaison environments is indeed conditioned by language experience. All participants in the NNS group had reached a minimum immersion requirement of five years in a French-speaking country. This period of time was chosen based on a body of research (Birdsong 2004; Johnson & Newport 1989) that has suggested that after five years of immersion in a L2 environment, a learner has reached an asymptotic state of attainment. An explanation of NNS behavior which is based on language exposure and use would therefore lead to the conclusion that the NNS groups have attained levels of nativelike sensitivity to allophonic variation in this particular phonological environment.

A second factor not related to language experience, however, could account for the NNS groups' sensitivity to these durational differences. Lengthening of word-initial consonants relative to both word-medial and word-final consonants has been shown cross-linguistically (see for example Cho & Keating 2001; Fougeron 2001). Specifically, the NNS in the current study are native speakers of English, a language that consistently shows word-initial lengthening (e.g. Klatt 1976; Cho 2007). Therefore, sensitivity to these durational differences could plausibly be a processing tool carried over from the L1 and hence not dependent on L2 exposure.

In light of the preceding considerations, an additional experiment was carried out. It was decided to test beginning learners of L2 French with very little exposure to French at the time of testing. If sensitivity to durational differences is indeed conditioned by language exposure we would expect the beginning group to show significantly lower

discrimination scores relative to both NS and NNS participants. If sensitivity to the durational differences is operating at a purely acoustic level or if listeners are using L1 perceptual strategies, then we would expect performance of the beginning group to be indistinguishable from that of the NS and NNS groups.

The next section presents an AX discrimination task testing beginning learners of L2 French. Beginners were tested uniquely on non-word stimuli (and not real-word stimuli) in order to control for lexical knowledge among the three participant groups (NS, NNS and beginners).

6.5 Experiment 1.2: AX Discrimination (beginning learners)

6.5.1 Participants

Eighteen beginning learners of French participated in Experiment 1.2. The beginning group (henceforth BEG) comprised 10 females and 8 males with an average age of 19.75 years (range: 18-21 yrs). All participants were native speakers of American English enrolled in first semester French at the University of Texas at Austin. At the time of testing participants had been exposed to approximately 20 hours of French instruction (five hours per week for approximately one month) with a native French-speaking instructor. Two participants had taken brief vacations in France as children, but other than this limited exposure, participants had no previous exposure to the French language. None of the participants was bilingual from birth.

Participants received extra credit in the French course for participation in the experiment. All had normal or corrected vision and none reported any hearing impairment. Handedness was not controlled for.

6.5.2 Stimuli

The stimuli used in Experiment 1.2 were identical to those used in Experiment 1.1.

6.5.3 Procedure

The procedure in Experiment 1.2 was identical to that of Experiment 1.1, the only alteration being that, given the BEG participants' low proficiency level in French, communication before testing was conducted in English and written instructions were presented in English (see Appendix VIII).

6.5.4 Results: Experiment 1.2

Two BEG participants (BEG 10 and 16) were removed from analysis in Experiment 1.2 due to the fact that they responded *same* to all 648 trials of the experiment. The following analyses thus include 16 BEG participants.

As in Experiments 1 and 1.1, only responses for the six different pairs (1_2, 1_3, 2_1, 2_3, 3_1, 3_2) are reported here. Responses for the three same pairs were excluded from analysis. All analyses are by subject.

6.5.4.1 Mean Accuracy

First mean accuracy rates for BEG group for the six different pairs were calculated and are reported below in Table 6-13. Mean accuracy rates, range and standard

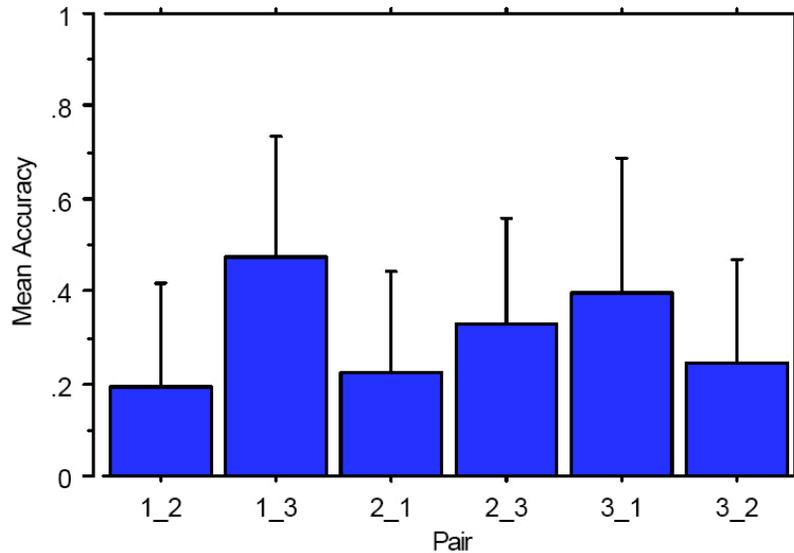
deviations for each condition across participants are also included at the bottom of the table in addition to results of single sample t-tests comparing BEG performance to chance (50%).

Table 6-13: Mean accuracy rates by pair on AX discrimination task employing non-word stimuli for beginning learners. The bottom rows of the table present mean accuracy, range, standard deviations across participants and single sample t-test results comparing performance to chance (50%).

	1 2	1 3	2 1	2 3	3 1	3 2
BEG 1	21	57	18	43	35	19
BEG 2	8	86	21	57	89	38
BEG 3	3	49	1	19	25	6
BEG 4	35	53	25	38	24	22
BEG 5	72	74	75	74	75	78
BEG 6	10	50	13	29	26	11
BEG 7	6	28	3	14	8	8
BEG 8	4	8	0	6	4	0
BEG 9	13	68	29	19	65	21
BEG 10			removed from analysis			
BEG 11	3	6	4	4	3	6
BEG 12	68	71	65	76	78	68
BEG 13	4	17	3	8	19	18
BEG 14	8	71	21	43	64	14
BEG 15	7	76	36	33	71	33
BEG 16			removed from analysis			
BEG 17	40	51	33	49	35	42
BEG 18	6	22	10	13	10	4
MEAN ACC	19.25	49.19	22.31	32.81	39.44	24.25
RANGE	3 – 72	6 – 86	1 – 75	4 – 76	4 – 89	0 – 78
(SD)	(22.68)	(25.64)	(21.98)	(23.02)	(29.42)	(22.62)
DIFF FROM CHANCE	P < .0001	NS	P = .0001	P = .0092	NS	P = .0004

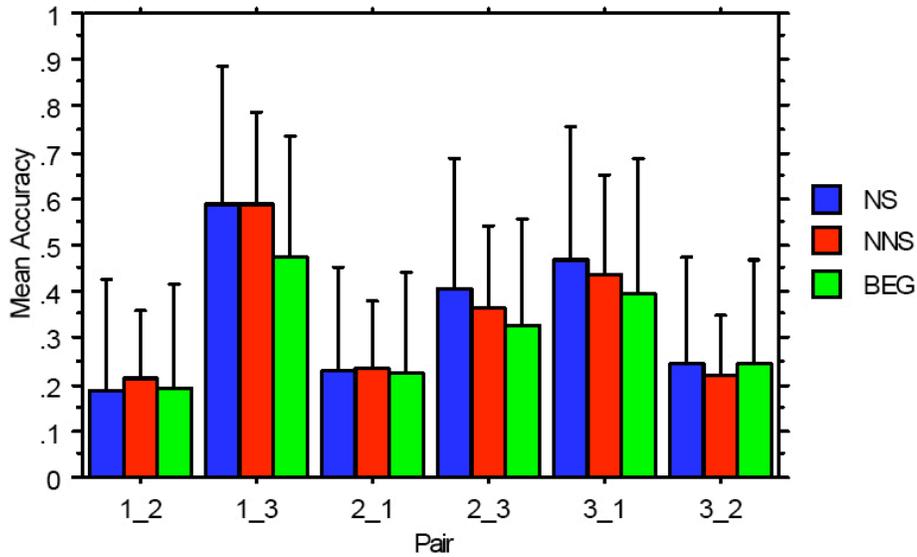
Mean accuracy rates for the BEG group were then submitted to a one-way ANOVA investigating the factor Pair (six levels), which revealed a significant main effect ($F(5, 102) = 3.367, p = .0078$). Though mean accuracy rates were at or below chance levels for each of the six pairs, there were significant differences in discrimination among the pairs. In addition these differences followed the same pattern as both the NS group and the advanced NNS group, with the 1_3 stimulus pair being differentiated better than other pairs. Scheffe post-hoc tests, however, revealed just one (marginally) significant difference for specific pair-by-pair comparisons, that of 1_2 and 1_3 ($p = .0583$). These results are displayed graphically in Figure 6-26 below.

Figure 6-26: Mean accuracy rates by Pair for beginning learners of French on AX discrimination task employing non-words.



Mean accuracy rates for the BEG group were then compared to those of both the NS group and the advanced NNS groups tested in Experiment 1.1 in a two-way factorial ANOVA, which revealed a main effect of Pair ($F(5, 271) = 18.164, p < .0001$), but no significant effect of Participant Group ($F(2, 271) = 1.045, NS$) and no interaction ($F(10, 271) = .267, NS$). However, Scheffe post-hoc tests revealed that the BEG group had significantly lower accuracy rates than both the NS group ($p = .0098$) and the advanced NNS group ($p = .0425$). There was no significant difference however between the NS and the NNS groups, which accounts for the lack of a significant overall main effect of Participant Group. See Figure 6-27 below.

Figure 6-27: Mean accuracy rates by Pair for three participant groups on AX discrimination task employing non-words.



6.5.4.2 D-prime

D-prime scores were also calculated for each BEG participant and are presented in Table 6-14 below. A significant effect of Pair was found ($F(5, 102) = 5.930, p < .0001$) as can be seen in Figure 6-28 below. Note, however, that only one stimulus pair, 1_3, showed a mean d-prime score above 1.

Figure 6-28: Beginner d-prime scores by Pair for AX discrimination task employing non-word stimuli.

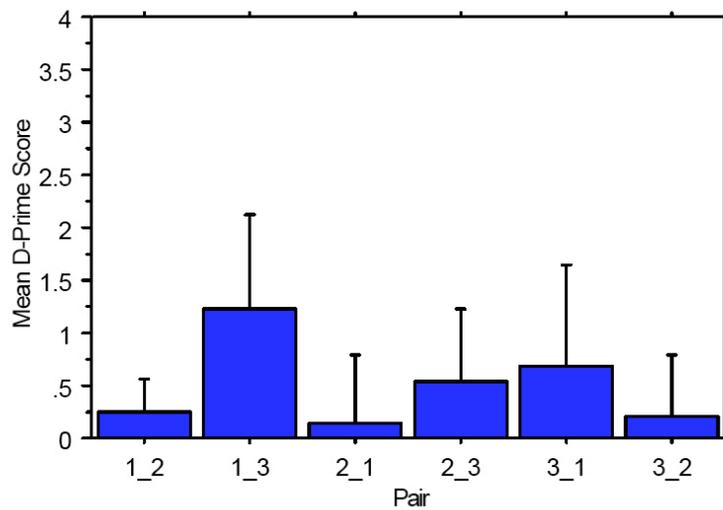


Table 6-14: *D-prime scores for beginning participants for each different pair in AX discrimination task employing non-words.*

	1 2	1 3	2 1	2 3	3 1	3 2
BEG 1	0.56	1.58	0.12	0.86	0.26	-0.23
BEG 2	0.07	2.60	1.25	2.23	2.51	1.17
BEG 3	-0.24	1.67	-0.58	0.87	1.62	0.77
BEG 4	0.26	0.77	0.13	0.50	-0.09	-0.16
BEG 5	-0.06	-0.03	-0.06	-0.10	0.25	0.44
BEG 6	0.41	1.78	-0.17	0.40	0.56	-0.05
BEG 7	0.41	1.50	-0.24	0.56	0.06	0.00
BEG 8	0.30	0.71	-0.27	0.50	-0.11	-0.68
BEG 9	0.58	2.22	1.20	0.87	2.03	0.84
BEG 10	removed from analysis					
BEG 11	0.45	0.92	-0.11	-0.11	-0.27	-0.08
BEG 12	0.08	0.22	-0.11	0.21	0.18	-0.03
BEG 13	-0.27	0.60	-0.54	-0.06	0.50	0.43
BEG 14	0.71	2.61	0.84	1.47	1.81	0.40
BEG 15	0.58	2.76	1.39	1.31	2.55	1.61
BEG 16	removed from analysis					
BEG 17	0.08	0.41	-0.11	0.31	-0.31	-0.15
BEG 18	0.11	1.04	0.77	0.93	0.42	-0.11
MEAN	0.25	1.23	0.15	0.55	0.69	0.22
RANGE	-0.24 –	-0.03 –	-0.58 –	-0.11 –	-0.31 –	-0.68 –
	0.71	2.76	1.39	2.23	2.55	1.61
(SD)	(0.30)	(0.89)	(0.65)	(0.63)	(1.00)	(0.59)

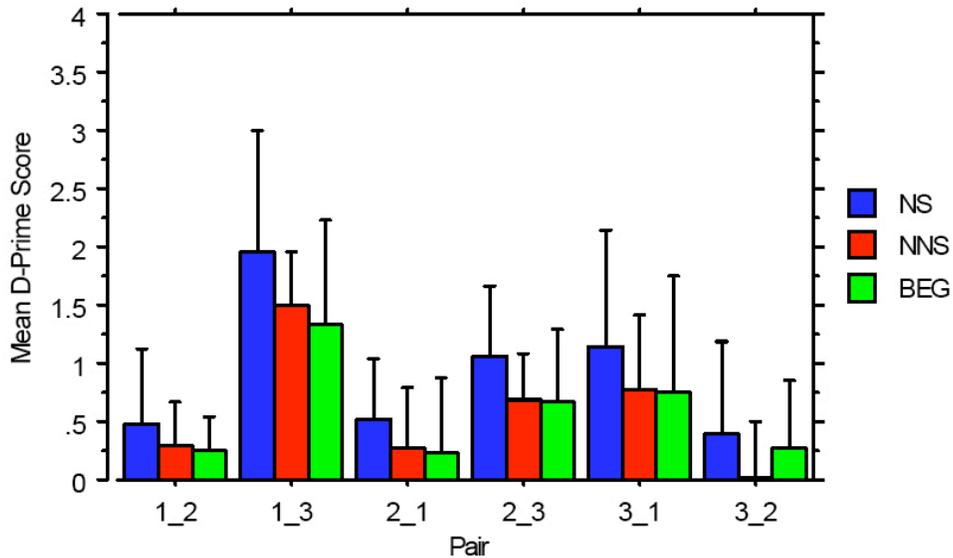
The BEG group was able to discriminate among the six pairs of manipulated stimuli, and in the same direction as both the NS and the NNS groups. The results of post-hoc specific comparisons for the BEG group are presented below in Table 6-15 along with results for the NS and NNS group reproduced from Table 6-11 below. As can be seen, the BEG group displayed significant discrimination only of the pair 1_3 as compared to other pairs.

Table 6-15: Results of Scheffe post-hoc tests showing discrimination among pairs of non-word stimuli according to d-prime scores for beginning learners, native speakers and non-native speakers.

Pair-by-pair comparison	Beginners: Value of P	Native Speakers: Value of P	Non-native Speakers: Value of P
1_2, 1_3	.0043	.0001	<.0001
1_2, 2_1	NS	NS	NS
1_2, 2_3	NS	NS	NS
1_2, 3_1	NS	NS	NS
1_2, 3_2	NS	NS	NS
1_3, 2_1	.0029	.0002	<.0001
1_3, 2_3	NS	(.0691)	.0005
1_3, 3_1	NS	NS	.0034
1_3, 3_2	.0048	NS	<.0001
2_1, 2_3	NS	NS	NS
2_1, 3_1	NS	NS	NS
2_1, 3_2	NS	NS	NS
2_3, 3_1	NS	NS	NS
2_3, 3_2	NS	NS	.0122
3_1, 3_2	NS	NS	.0023

D-prime scores for the BEG group were then compared to NS and NNS d-prime scores from Experiment 1.1. This analysis revealed a significant effect of Pair ($F(5,276) = 29.205, p < .0001$) as well as an effect of Participant Group ($F(1,276) = 8.589, p = .0002$) and no interaction ($F(5,276) = .340, NS$). A Scheffe post-hoc test showed that the NS group performed significantly better than both the advanced NNS group ($p = .0019$) and the BEG group ($p = .0017$). The difference between the NNS and the BEG groups was also marginally significant ($p = .0553$). This analysis is displayed in Figure 6-29 below.

Figure 6-29: Mean *d*-prime scores by Pair for three participant groups on AX discrimination task employing non-word stimuli.



6.5.4.3 Results by Consonant

Accuracy rates for the BEG group for each of the three consonants were then submitted to ANOVAs which revealed a main effect of Consonant ($F(2, 306) = 3.430, p = .0336$) as well as a main effect of Pair ($F(5, 306) = 7.980, p < .0001$) and no interaction ($F(10, 306) = .244, NS$). Post-hoc analyses showed that the BEG group discriminated stimuli containing /n/ better than stimuli containing /z/ ($p = .0366$), however there was no significant difference between /n/ and /t/ or between /t/ and /z/. These results are displayed graphically in Figure 6-30.

Mean accuracy rates for each of the three consonants for the BEG group were then compared to those of both the NS group and the advanced NNS group also using a factorial ANOVA, which revealed a main effect of Pair ($F(5, 864) = 45.206, p < .0001$), as well as an effect of Participant Group ($F(2, 864) = 9.215, p = .0001$) and an effect of Consonant ($F(2, 864) = 9.9125, p < .0001$). There were no significant interactions. The BEG group displayed significantly lower discrimination than both the NS group ($p = .0003$) and the advanced NNS group ($p = .0067$) according to Scheffe post-hoc specific comparisons. See Figure 6-31 below.

Figure 6-30: *Beginner mean accuracy on AX discrimination task employing non-word stimuli by Pair and Consonant.*

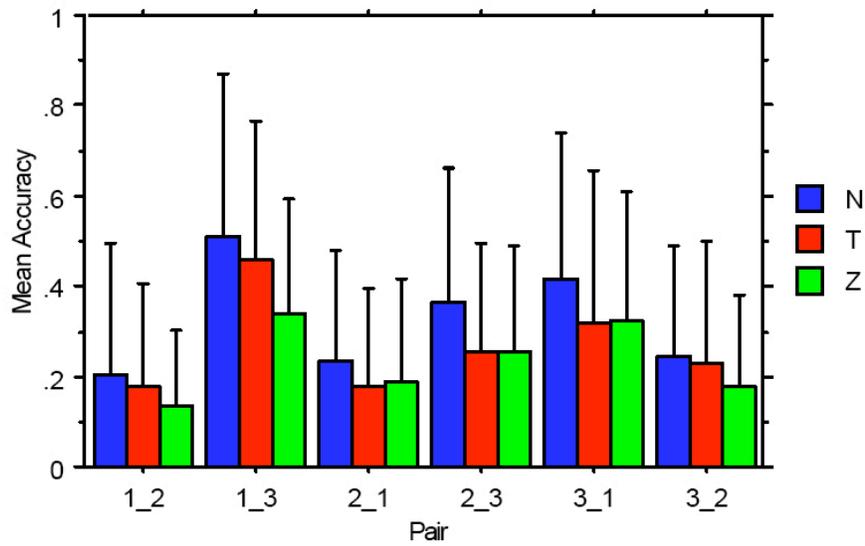
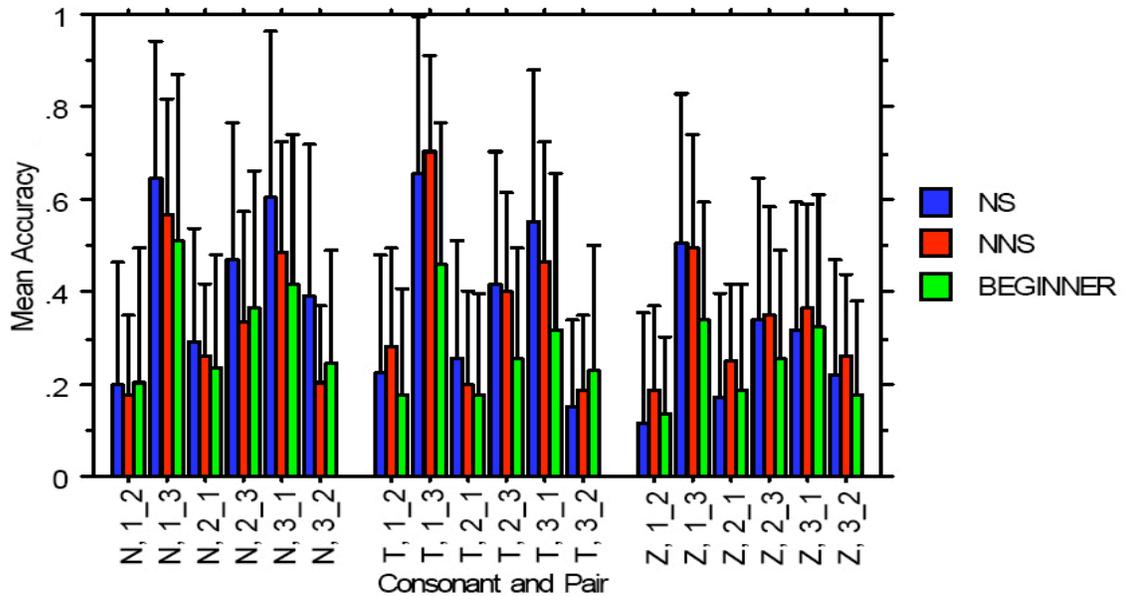
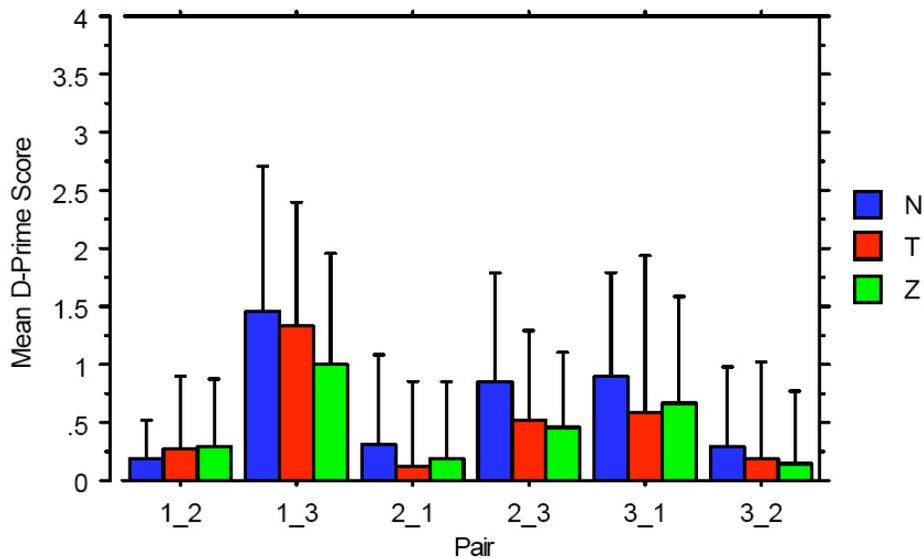


Figure 6-31: *Mean accuracy on AX discrimination task employing non-word stimuli for three participant groups by Pair and Consonant.*



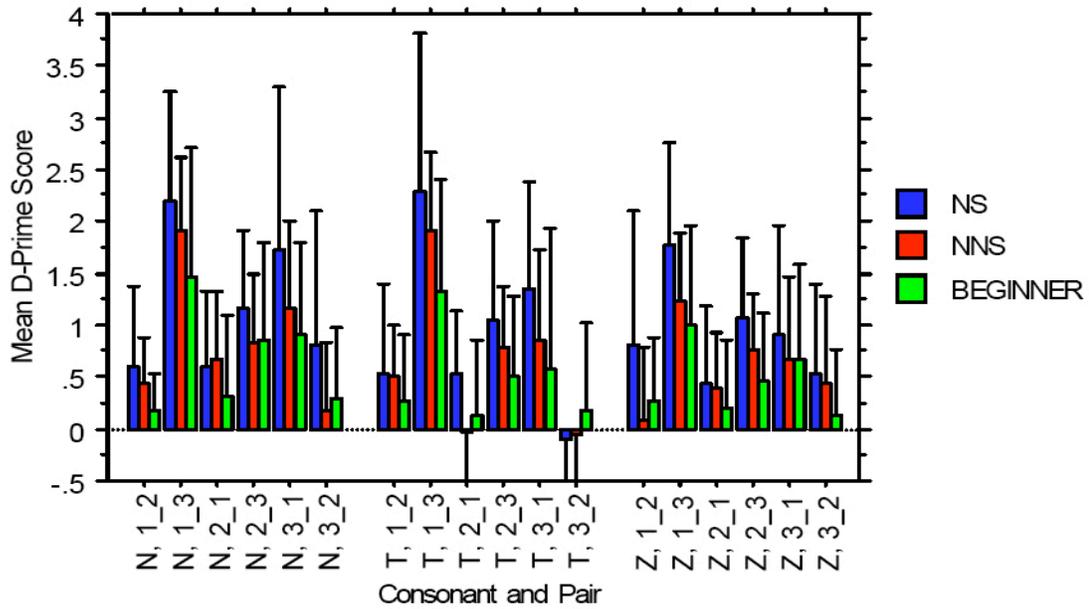
D-prime scores on non-word stimuli were also analyzed for each individual consonant for the beginning learners and were submitted to ANOVAs with the factors Consonant and Pair, which revealed no significant effect of Consonant ($F(2, 306) = 1.888$, NS), but a main effect of Pair ($F(5, 306) = 12.867$, $p < .0001$). There was no significant interaction ($F(10, 306) = .373$, NS). These results are summarized below in Figure 6-32.

Figure 6-32: *Beginner d-prime scores for AX discrimination task employing non-word stimuli by Pair and Consonant.*



D-prime scores for each of the three consonants for the BEG group were subsequently compared to those of both the NS group and the advanced NNS group, which revealed main effects for each of three factors: Pair ($F(5, 864) = 58.828$, $p < .0001$), Participant Group ($F(2, 864) = 7.103$, $p = .0009$) and Consonant ($F(2, 864) = 23.301$, $p < .0001$). There was one significant interaction between the factors Pair and Consonant ($F(10, 864) = 1.924$, $p = .0388$), again demonstrating that the effect of Pair is not equivalent among the three segments. The BEG group displayed significantly poorer discrimination than both the NS group ($p < .0001$) and the advanced NNS group ($p = .0531$) according to Scheffe post-hoc specific comparisons. These results are displayed in Figure 6-33 below.

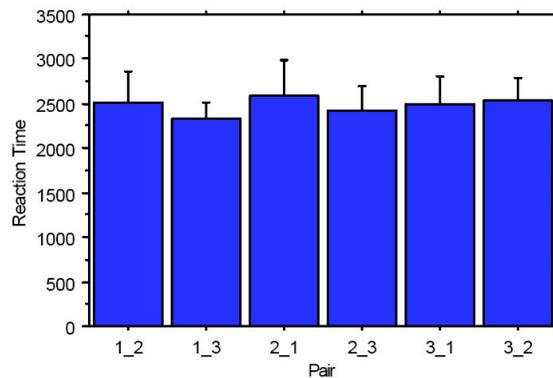
Figure 6-33: *D*-prime scores for AX discrimination task employing non-word stimuli for three participant groups by Pair and Consonant.



6.5.4.4 Reaction Time

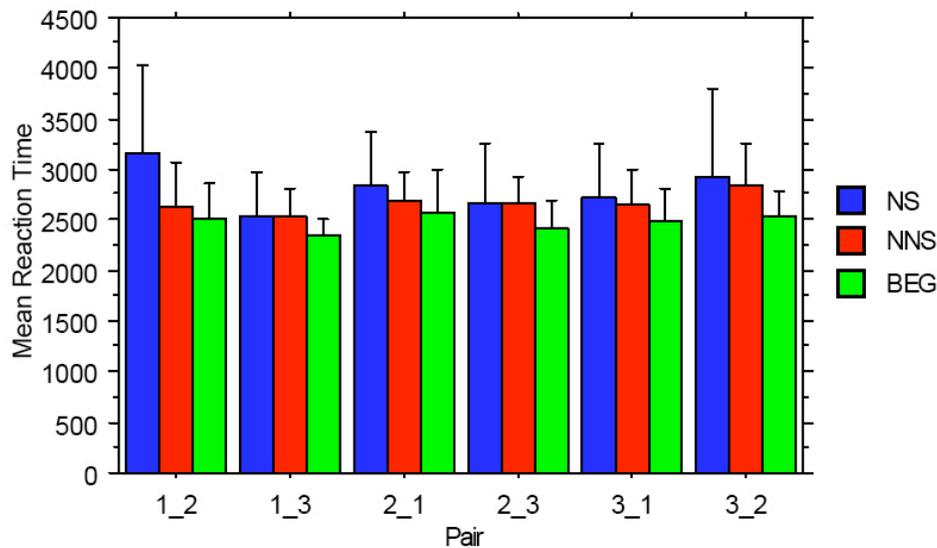
Mean reaction times for correct responses in the AX discrimination task were also analyzed for the BEG group, which showed no significant effect of Pair ($F(5, 88) = 1.303$, NS) as shown in Figure 6-34 below.

Figure 6-34: Mean reaction times in milliseconds on AX discrimination task employing non-word stimuli by Pair for beginning group.



BEG reaction times were then compared with the NS and NNS groups and are shown below in Figure 6-35. As in previous analyses, responses below 200 ms and above 4000 ms were discarded. A factorial ANOVA showed a significant effect of Pair ($F(5, 259) = 3.209, p = .0079$), as well as a significant effect of Participant Group ($F(2, 259) = 12.041, p < .0001$) and no interaction ($F(10, 259) = .860, NS$). Scheffe post-hoc tests showed no significant difference between the NS group and the advanced NNS speaker group. However, the BEG group had significantly faster reaction times than both the NS group ($p < .0001$) and the NNS group ($p = .0035$).

Figure 6-35: Mean reaction times in milliseconds on AX discrimination task employing non-word stimuli by Pair and Participant Group for three participant groups.



6.5.5 Discussion: Experiment 1.2

The BEG group was able to distinguish among manipulated stimuli from the three-degree durational continuum. In addition, this group exhibited the same pattern of results as the NS and NNS groups, namely pairs of stimuli that represented two degrees of physical difference on the continuum were discriminated significantly better than pairs separated by only one degree. However, when mean accuracy rates and d-prime scores across the three participant groups were compared, the BEG group showed significantly poorer discrimination than both the NS and NNS groups. Furthermore, the BEG group showed little sensitivity among the three consonants. An analysis of mean accuracy revealed a significant difference only between /n/ and /z/ for the BEG group, whereas d-prime scores revealed no differences among the consonants for beginning learners. Recall that both the NS and the NNS group showed significant discrimination differences among the three consonants and in the same direction (/n/ > /t/ > /z/) in Experiment 1.1.

6.6 General Discussion: AX Discrimination Task

The objective of the AX discrimination tasks presented in this chapter was to investigate the perceptual saliency of durational variation in consonants at word boundaries in possible environments of liaison. Stimuli were employed in which the duration of pivotal consonants was instrumentally manipulated according to a three-step durational continuum.

In light of the inconclusive comparisons between the NS and NNS participants in Experiments 1 and 1.1, it was decided to test beginning learners in Experiment 1.2. The results of Experiment 1.2 show that listeners with very little exposure to French are sensitive to variation in segmental duration in French, but to a significantly lesser extent than both native speakers and more experienced learners, suggesting that the perceptual saliency of durational differences between consonants that arise in liaison environments and initial consonants is indeed conditioned by language exposure.

Earlier, we noted three possibilities that could account for similarities in performance between the NS and the NNS groups observed in Experiments 1 and 1.1. We proposed that nativelike sensitivity to durational differences in environments of liaison on the part of the NNS group could be attributable to 1) extensive exposure to the L2 French 2) sensitivity to domain-initial strengthening carried over from the L1 or 3) general auditory processes. The results of Experiment 1.2 taken together with those of Experiments 1 and 1.1 suggest that the perceptual saliency of variation in segmental duration in spoken French is in fact largely conditioned by language exposure. While analyses in Experiments 1 and 1.1 were not consistent as to differences in the perceptual capacities of the NS and NNS participant groups, where significant differences did arise in analyses of both mean accuracy and d-prime scores they were in the same direction, namely the NS group performed better than the NNS group. The further results of Experiment 1.2 confirm this trend. The BEG group showed decreased discrimination relative to both the NS and the NNS groups.

However, the performance of the BEG group also indicates that general auditory processing plays a role in the current task. Though the BEG group displayed decreased discrimination relative to both the NS and NNS groups, BEG participants were able to discriminate pairs that were separated by two degrees on the continuum. The 1_3 stimulus pair was discriminated better than the 3_1 stimulus pair, which we again attribute to the relative proportion of change between the two pairs. This result is not surprising given that the percent change for the pair 1_3 is sixfold the assumed JND differential threshold of 20%.

In reference to what are considered to be JNDs for segmental duration, discrimination performance by all participants was lower than would be predicted. We noted at the beginning of this chapter, however, that we expected several factors to mitigate perception in the case of the current stimulus sample. First, we noted that the variation being investigated in the current study is not categorical, but represents a non-contrastive allophonic difference in segmental duration. In the stimulus sample employed in the current AX task, we have purposefully exaggerated the durational differences observed between LCs and ICs in order to investigate the effect of this single factor on SWR processes. Nonetheless, the use of standard deviations as the factor by which the

stimuli were manipulated ensured that the degree of durational variation fell within a reasonable production range and therefore represents ecologically valid instances of allophonic variation.

Following this line of thought, the question arises as to whether the durational differences would be discriminated better if the consonants were presented in isolation or in CV or VC syllables rather than in an environment that licenses the allophonic variation being studied here. We did not test this possibility, but we predict that participants would indeed show increased discrimination if these consonants were removed from their phonological environment and presented to participants in, for example, isolated CV syllables.

An analysis of reaction times showed an unexpected result. There were no significant differences found between the NS and NNS groups in Experiments 1 and 1.1, suggesting that these two groups experienced similar demand on working memory resources and processing load in the execution of this task. However, a comparison of response latencies between the BEG group and the NS and NNS groups showed that BEG participants had significantly shorter response times. As noted above, reaction time analysis is usually taken to be an indicator of the difficulty of a task as it is assumed to index an increased processing load. The decreased sensitivity to segmental duration on the part of the BEG group as shown by low accuracy rates and d-prime scores would lead us to believe that the task was more difficult for this group. However, the fact that the BEG group had shorter reaction times than both the NNS and NS groups suggests either a lack of attentional resources given to the task or shallower processing (i.e. processing not influenced by higher-level information) relative to the NS and the NNS group.

To sum up, two major conclusions emerge from the current results. First, this series of experiments has shown that, when durational differences between LCs and ICs are greatly exaggerated in relation to a normal distribution of production values discrimination is strong, however when differences are separated by only one degree of separation on our continuum, representing a more normal production range, discrimination of durational variation is inconsistent. While the current stimulus sample is based on the range and distribution of the production sample described above in Chapter Five, as noted above in §6.2.5, it reflects a degree of variation that would rarely occur in natural speech. Therefore, this acoustic component probably does not represent a cue that would be consistently exploited in the processing of natural speech.

Second, these results provide support for the hypothesis that the perceptual saliency of durational differences that occur in possible environments of liaison in spoken French is conditioned by exposure to the language. Due to the amount of variation in performance across participants from all three groups, however, this conclusion remains somewhat tentative and we will revisit it later in this dissertation.

On the surface, this second conclusion may not appear to be in line with our first conclusion. If duration is not a robust cue to processing in possible liaison environments, why would the perceptual system bother to encode it in the listener's grammar? Our response to this is that, though this cue may not be the most robust, it is still systematically present in the speech signal as evidenced by numerous production studies (Dejean de la Bâtie 1993; Gaskell et al. 2002; Nguyen et al. 2007; Shoemaker 2006;

Spinelli et al. 2003). The study of statistical learning has shown that the perceptual system is extremely sensitive to distributions in the speech signal (Clarke & Garrett 2004; Dupoux & Green 1997; Maye, Aslin & Tanenhaus 2008; Saffran, Newport & Aslin 1996), therefore it is not surprising that sensitivity to this durational cue, albeit relatively weak, can be modulated by language experience.

In the following chapter we examine whether durational differences between LCs and ICs are exploited in higher level processing, specifically lexical access and the segmentation of continuous speech.

CHAPTER SEVEN: Forced-Choice Identification Task

7.0 Introduction

The current chapter presents the second of two behavioral studies probing the processing of durational variation between consonants that surface in environments of liaison and initial consonants in spoken French. Experiment 2, which employs real-word stimuli, and Experiment 2.1, which employs non-word stimuli, each consist of a two-alternative forced-choice identification task employing globally ambiguous phrases in which the duration of the pivotal consonants has been instrumentally manipulated. While Experiments 1, 1.1 and 1.2 were designed to tap thresholds of acoustic saliency through the use of an AX discrimination task, a forced-choice identification task is employed to examine whether listeners exploit segmental duration in the lexical interpretation of phonemically identical sequences. Employing methodology used in previous research (e.g. Quené 1992; Shatzman & McQueen 2006; Warner et al. 2004), the current task isolates and exaggerates this single acoustic cue while holding all other acoustic information in the speech signal constant. In this way, we can directly test whether durational variation represents a sufficient acoustic cue for segmentation in cases of homophonic ambiguity in possible liaison environments.

7.1 Predictions

Variation in segmental duration at word boundaries been shown cross-linguistically. Moreover, differences are consistently in the same direction, i.e. segments tend to be longer word-initially than in other positions²² (e.g. Fougeron & Keating 1997; Fougeron 2001; Klatt 1976; Gow & Gordon 1995; Lehiste 1960, 1961; Oller 1973; Umeda 1975). In addition, segmental duration has been shown to serve as a perceptual cue to the location of both word and syllable boundaries in English (Davis et al. 2002; Salverda et al. 2003), Dutch (Quené 1992; Shatzman & McQueen 2006), Italian (Tabossi et al. 2000), and French (Banel & Bacri 1994).

Durational differences between LCs and ICs in spoken French are consistently produced (Dejean de la Bâtie 1993; Gaskell et al. 2002; Nguyen et al. 2007; Shoemaker 2006; Spinelli et al. 2003; Wauquier-Gravelines 1996). Consonants that surface in liaison environments (e.g. /n/ in *un air*) have been shown to be shorter than the same consonant in initial position (e.g. /n/ in *un nerf*). Several researchers have suggested that duration may be a cue to segmentation in environments of liaison (Gaskell et al. 2002; Spinelli et al. 2003), however, no study to date has directly demonstrated that listeners exploit this cue in the processing of continuous speech in French.

²² Segments in word-final position also exhibit lengthening, however such lengthening is usually limited to phrase-final or pre-pausal position (Klatt 1975, 1976).

Based on the above body of research, we predict that the lexical interpretation of ambiguous sequences such as [œ.nɛʁ] will be modulated by the duration of the pivotal consonant. Specifically, we predict that stimuli containing a shortened segment will elicit a significantly higher proportion of vowel-initial (liaison) responses, while stimuli containing a lengthened segment will elicit a higher proportion of consonant-initial responses. Furthermore, we predict that the baseline stimuli, which represent segmental durations intermediate to LCs and ICs will not offer participants enough acoustic information to guide responses and will therefore elicit a guessing strategy resulting in an equal number of V-initial (liaison) responses and C-initial responses.

The AX discrimination tasks in Experiments 1, 1.1, and 1.2 established that durational differences between LCs and ICs are salient to both native and advanced non-native speakers of French only when these differences are greatly exaggerated. In addition, we also observed a great deal of variation across both NS and NNS participants. The combination of these two factors suggests that this cue is not systematically exploited to the same extent by listeners in natural speech environments. Consequently, we predict that the current identification tasks will exhibit a comparable amount of variation.

The identification tasks in Experiments 2 (real-word stimuli) and 2.1 (non-word stimuli) test the same participant groups as Experiments 1 (real-word stimuli) and 1.1 (non-word stimuli), native speakers of French and highly advanced adult learners of French. Concerning the performance of the two participant groups in the current task, the results of Experiments 1, 1.1, and 1.2 suggest that language exposure plays a significant role in the perceptual saliency of durational contrasts in this particular phonological environment in spoken French. Following these results, we predict that the NS group will show increased sensitivity to the duration of the pivotal consonants in the identification task relative to the NNS group, resulting in a higher proportion of liaison responses for shortened consonants and a higher portion of consonant-initial responses for lengthened consonants.

One further prediction pertains to the three individual segments under investigation. As in the AX discrimination tasks presented in Chapter 6, the current forced-choice identification tasks test three consonants that surface in environments of liaison, /n/, /t/ and /z/. Experiments 1 and 1.1 showed that the robustness of durational differences between LCs and ICs for each individual consonant predicted participants' discrimination performance, i.e. sensitivity to durational differences followed the same pattern of differences observed in the production sample (/n/ > /t/ > /z/). We predict that the pattern of responses in Experiments 2 and 2.1 will mirror this trend.

7.2 Experiment 2: Forced-choice Identification Task (real-word stimuli)

7.2.1 Participants

Thirty-six participants (18 native speakers and 18 non-native speakers) who participated in Experiment 1 took part in Experiment 2. See §6.2.1 above for a full description of these two participant groups.

7.2.2 Stimuli

The stimuli used in Experiment 2 consist of manipulated tokens taken from the three-step durational continuum of real-word stimuli described in §5.4.

7.2.3 Procedure

After completing Experiment 1, participants were given the option to take a short break before beginning Experiment 2. As in Experiment 1, participants were tested individually in a quiet room. The experimental protocol was created using E-Prime software (Schneider, Eschman & Zuccolotto 2002) and was presented on a Dell Inspiron 600m laptop computer. Stimuli were presented binaurally through Koss UR 20 headphones. Each experimental trial had the following structure: Participants heard one of the manipulated phrases from the durational continuum presented aurally through headphones. Phrases were presented without a carrier frame, thus eliminating any potential priming effects from context. At the offset of the auditory stimulus, two words appeared on the computer screen. There was no delay between the offset of the auditory stimulus and the presentation of the visual targets²³. The two visual targets consisted of the V-initial and C-initial candidates representing the two possible interpretations of the final word in each ambiguous sequence described in Chapter Five, e.g. when auditory stimulus is a manipulated version of [œ.nɛʁ] *air* and *nerf* are visual targets. Participants were instructed to indicate which of the two words presented on the screen was present in the phrase they had heard by pressing on the computer keyboard either (1), corresponding to the word on the left of the screen, or (2), corresponding to the word on the right of the screen. Each of the 36 stimuli (i.e. 3 manipulated versions of each of 12 tokens; see §5.1.2) was presented randomly 6 times resulting in a total of 216 trials. Participants completed a training portion consisting of 14 trials before beginning the experimental portion in order to familiarize them with the procedure. Items included in the training portion were not included in the experimental portion. Individual trials were separated by 2000 ms. No response limit was set for Experiment 2. Visual targets were counter-balanced across participants in order to offset any possible bias toward the left-hand visual target that might occur from reading effects. Half of the participants were presented with the V-initial (liaison) target on the left of the screen and the other half were presented with the C-initial target on the left of the screen. Testing for Experiment 2 lasted approximately 25 minutes. See Appendices IX and X for experimental instructions and sample presentation screen.

7.2.4 Experiment 2: Results

The proportion of V-initial (i.e. ‘liaison’) responses was calculated for manipulated stimuli in each of the three continuum conditions: the shortened (LC)

²³ A version of the experiment in which auditory and visual stimuli were presented simultaneously was piloted on several subjects. Simultaneous presentation of auditory and visual stimuli was initially considered to avoid effects of memory decay on the auditory stimulus. However, piloting subjects reported it difficult to attend to the auditory stimulus while simultaneously reading the visual targets. Therefore, it was ultimately decided to present the auditory stimulus accompanied by a blank screen before presenting visual targets in order to reduce cognitive load.

version, the baseline version, and lengthened (IC) version. Table 7-1 below presents the proportion of ‘liaison’ responses for each participant in each of the three continuum conditions. Range, mean proportion of ‘liaison’ responses, and standard deviations across participants are also reported at the bottom of the table. Single-sample t-tests were also conducted in order to ascertain whether mean response proportions were significantly above or below chance performance (50%), the results of which are also summarized below. Note that all NS participants showed the predicted linear decline of proportion of liaison responses across the three conditions, i.e. proportions of V-initial responses were highest for shortened stimuli, followed by baseline stimuli and proportions were lowest for lengthened stimuli. All but two NNS participants (NNS 1 and 17) showed the predicted linear decline as well.

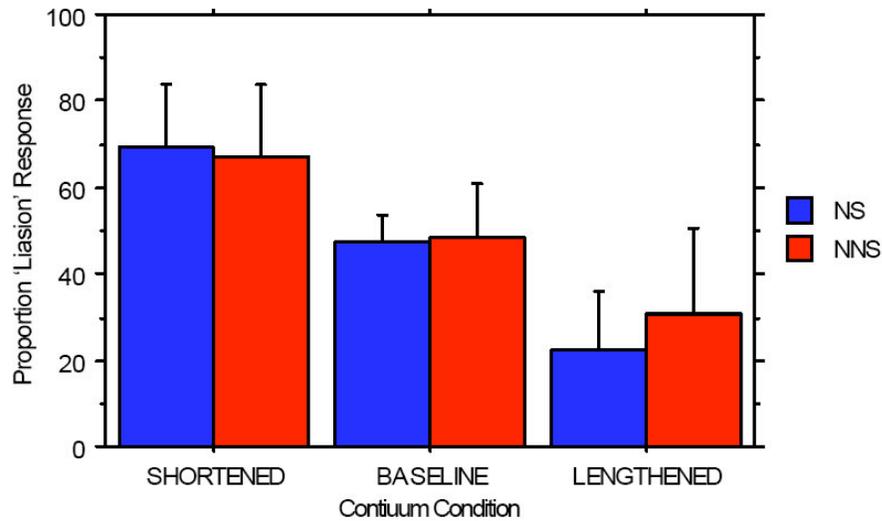
Table 7-1: *Proportion of ‘liaison’ responses for NS and NNS participants in forced-choice identification task employing real-word stimuli across three conditions of durational continuum.*

	Shortened	Baseline	Lengthened		Shortened	Baseline	Lengthened
NS 1	65.28	52.78	30.56	NNS 1	93.06	83.33	86.11
NS 2	72.22	40.28	4.17	NNS 2	58.33	44.44	37.50
NS 3	62.50	48.61	23.61	NNS 3	56.94	45.83	30.56
NS 4	94.44	44.44	12.50	NNS 4	34.72	29.17	23.61
NS 5	87.50	51.39	2.78	NNS 5	52.78	47.22	41.67
NS 6	83.33	55.56	16.67	NNS 6	88.89	50.00	18.06
NS 7	62.50	54.17	47.22	NNS 7	73.61	50.00	12.50
NS 8	56.94	51.39	37.50	NNS 8	59.72	41.67	25.00
NS 9	69.44	38.89	15.49	NNS 9	90.28	45.83	6.94
NS 10	56.94	38.89	20.83	NNS 10	70.83	40.91	31.94
NS 11	66.67	54.17	19.44	NNS 11	81.94	61.11	23.61
NS 12	68.06	37.50	15.28	NNS 12	65.28	31.94	8.33
NS 13	48.61	40.28	34.72	NNS 13	58.33	38.89	18.06
NS 14	50.00	48.61	38.89	NNS 14	80.56	63.89	25.00
NS 15	66.67	54.17	33.33	NNS 15	62.50	40.85	36.11
NS 16	55.56	43.06	37.50	NNS 16	76.39	61.11	33.33
NS 17	95.55	42.44	10.52	NNS 17	37.50	43.06	68.06
NS 18	88.30	51.77	3.68	NNS 18	66.67	50.00	29.17
RANGE	48.61- 95.55	37.50- 54.17	2.78- 47.22	RANGE	34.72- 93.06	29.17- 83.33	6.94- 86.11
MEAN	69.47	47.13	22.48	MEAN	67.13	48.29	30.86
(SD)	(14.63)	(6.33)	(13.57)	(SD)	(16.62)	(12.69)	(19.63)
DIFF				DIFF			
FROM	p < .0001	NS	p < .0001	FROM	p = .0004	NS	p = .0007
CHANCE				CHANCE			

A two-way factorial ANOVA compared participant groups and proportions of responses across the three continuum conditions. This analysis revealed a main effect of Continuum Condition ($F(2,102) = 74.30, p < .0001$), however no significant difference between the two Participant Groups was observed ($F(1,102) = .734, NS$) and there was no interaction between the two factors ($F(2,102) = 1.256, NS$). Mean proportions of

'liaison' responses for both participants groups across continuum conditions are shown below in Figure 7-1.

Figure 7-1: Mean proportion of 'liaison' responses in forced-choice identification task employing real-word stimuli across three conditions of durational continuum for NS and NNS participants.



The above results suggest that the duration of the pivotal consonant in liaison environments can indeed modulate the lexical interpretation of ambiguous sequences for both NS and NNS. Shortened consonants elicited significantly more V-initial responses, while lengthened consonants elicited significantly more C-initial responses. In addition, baseline consonants elicited roughly the same proportion of V-initial and C-initial responses. However, as was observed in the AX discrimination tasks presented in Chapter Six, there was a great deal of variation across participants in both groups as evidenced by the range of responses shown above in Table 7-1. This again brings into question the consistency with which this single acoustic cue is exploited in natural speech.

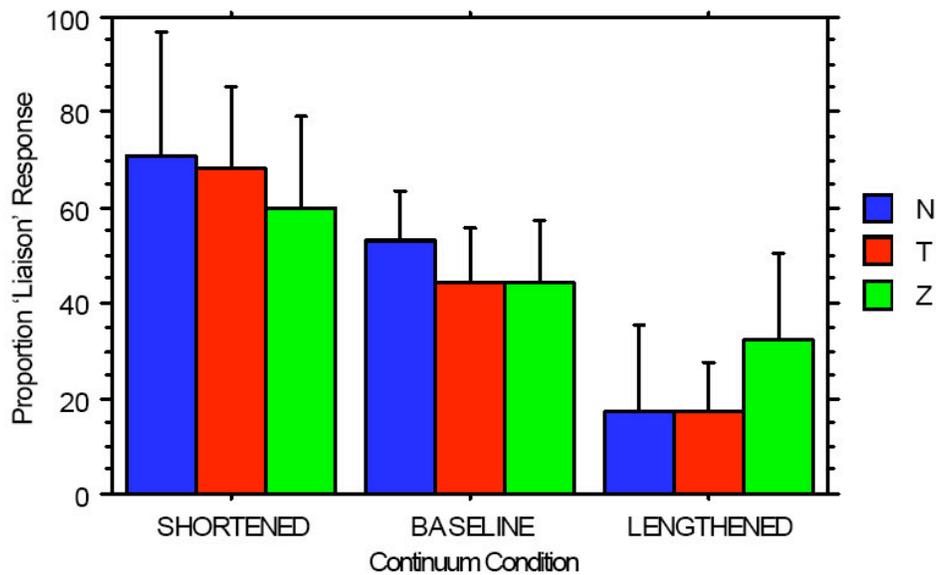
7.2.4.1 Analyses by Consonant

We have already noted that our production sample generated significant differences among the three consonants under investigation. Furthermore, performance on the AX discrimination tasks reviewed above differed significantly among the three segments for both NS and NNS. For these reasons, possible differences in perceptual sensitivity among the three segments /n, t, z/ were also considered in the identification task.

First, proportions of 'liaison' responses were compared across the three consonants for each continuum condition for the NS group. This analysis showed significant differences among Continuum Conditions ($F(2,153) = 95.368, p < .0001$), but

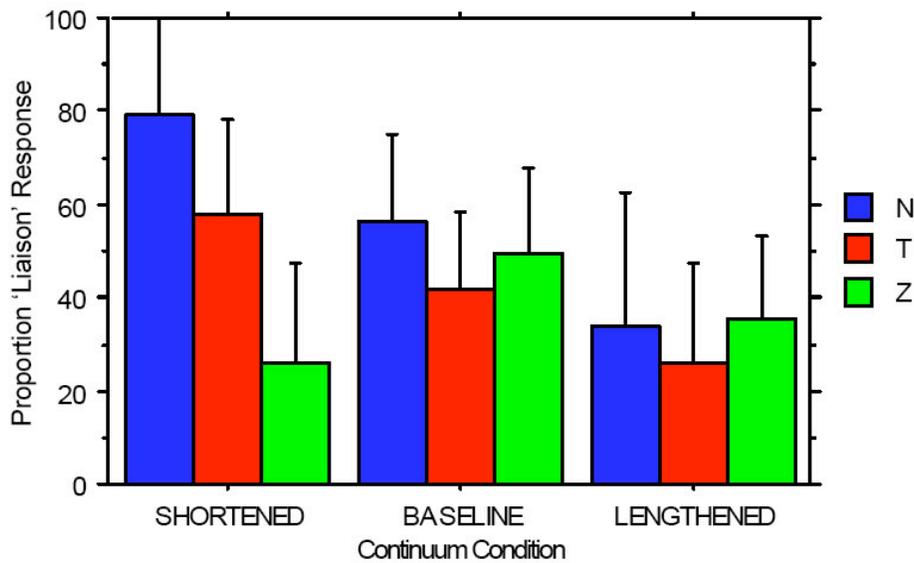
no significant difference among the three Consonants ($F(2,153) = .726$, NS). A significant interaction was also observed ($F(4,153) = 3.958$, $p = .0044$), indicating that the effect of Continuum Condition was not equivalent among the three segments. Proportions of ‘liaison’ responses across conditions and consonants for the NS group are shown below in Figure 7-2.

Figure 7-2: Mean proportion of ‘liaison’ responses in forced-choice identification task employing real-word stimuli across three continuum conditions and consonants for NS participants.



We then compared proportions of ‘liaison’ responses across the three consonants for each of the three continuum conditions for the NNS group. This analysis showed significant differences across Continuum Conditions ($F(2,153) = 17.242$, $p < .0001$) as well as among the Consonants ($F(2,153) = 13.062$, $p < .0001$). A significant interaction was also observed ($F(4,153) = 10.059$, $p < .0001$) indicating that the effect of Condition was not equivalent across the three segments. Scheffe post-hoc tests showed significant differences between /n/ and /t/ ($p = .0014$) and between /n/ and /z/ ($p < .0001$), but no difference between /t/ and /z/. Proportions of ‘liaison’ responses across conditions and consonants for the NNS group are shown below in Figure 7-3. Note that NNS responses for the segment /z/ did not follow the predicted response pattern in that the proportion of ‘liaison’ responses did not decrease linearly across the three continuum conditions. Shortened /z/ stimuli elicited the fewest V-initial responses of the three continuum conditions, while they would be predicted to elicit the highest proportion of V-initial response.

Figure 7-3: Mean proportion of 'liaison' responses in forced-choice identification task employing real-word stimuli across three continuum conditions and consonants for NNS group.



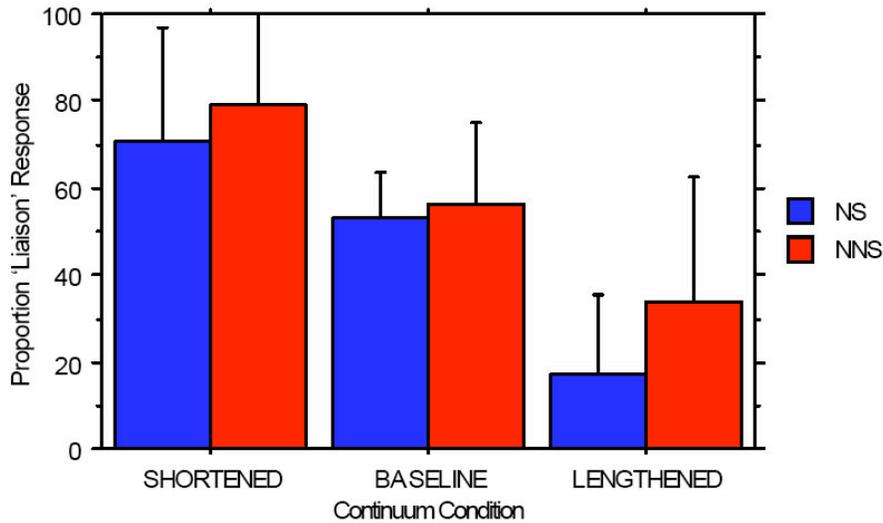
We then compared NS and NNS responses for each individual consonant separately. We first calculated the proportion of 'liaison' responses for only those stimuli containing the segment /n/ for each participant in each continuum condition. These proportions as well as the results of single-sample t-tests comparing these proportions to chance performance are presented below in Table 7-2. As can be seen, an analysis of only those stimuli containing the segment /n/ showed that six NS participants (NS 10, 12, 13, 14, 15, 16) and two NNS participants (NNS 15 and 17) did not show the predicted decrease of proportion responses across the three continuum conditions.

Table 7-2: Proportion of 'liaison' responses for NS and NNS participants in forced-choice identification task employing real-word stimuli across three conditions of durational continuum for stimuli containing /n/.

	Shortened /n/	Baseline /n/	Lengthened /n/		Shortened /n/	Baseline /n/	Lengthened /n/
NS1	83.33	58.33	37.50	NNS1	95.83	83.33	95.83
NS 2	91.67	54.17	0.00	NNS 2	62.50	45.83	37.50
NS 3	75.00	50.00	25.00	NNS 3	70.83	58.33	54.17
NS 4	100.00	37.50	0.00	NNS 4	54.17	37.50	45.83
NS 5	95.83	58.33	0.00	NNS 5	79.17	87.50	75.00
NS 6	100.00	45.83	0.00	NNS 6	100.00	50.00	0.00
NS 7	83.33	79.17	62.50	NNS 7	91.67	58.33	4.17
NS 8	70.83	62.50	37.50	NNS 8	83.33	45.83	12.50
NS 9	79.17	54.17	8.70	NNS 9	100.00	33.33	0.00
NS 10	37.50	45.83	8.70	NNS 10	79.17	54.17	12.50
NS 11	50.00	41.67	12.50	NNS 11	95.83	75.00	25.00
NS 12	50.00	58.33	12.50	NNS 12	100.00	58.33	8.33
NS 13	45.83	50.00	16.67	NNS 13	54.17	41.67	29.17
NS 14	45.83	45.83	39.13	NNS 14	83.33	79.17	20.83
NS 15	37.50	66.67	32.26	NNS 15	87.50	45.83	45.83
NS 16	29.17	50.00	15.00	NNS 16	100.00	83.33	20.83
NS 17	100.00	41.67	0.00	NNS 17	12.50	25.00	83.33
NS 18	100.00	58.33	0.00	NNS 18	75.00	54.17	41.67
RANGE	29.17 – 100.00	37.50 – 79.17	0.00 – 62.50	RANGE	12.50 – 100.00	25.00 – 87.50	0.00 – 95.83
MEAN	70.83	53.24	17.11	MEAN	79.17	56.48	34.03
(SD)	(25.36)	(10.16)	(18.17)	(SD)	(22.51)	(18.43)	(28.63)
DIFF				DIFF			
FROM	p = .0028	NS	p < .0001	FROM	p < .0001	NS	p = .0301
CHANCE				CHANCE			

A factorial ANOVA was conducted to compare response proportions for the segment /n/ across the three conditions between participant groups, which revealed a main effect of Continuum Condition ($F(2,102) = 48.498, p < .0001$), as well as an effect of Participant Group ($F(1,102) = 5.290, p = .0235$) and no interaction ($F(2,102) = .912, NS$). This analysis is shown graphically in Figure 7-4.

Figure 7-4: Mean proportion of 'liaison' responses in forced-choice identification task employing real-word stimuli across three conditions of durational continuum for stimuli containing the segment /n/ for NS and NNS participants.



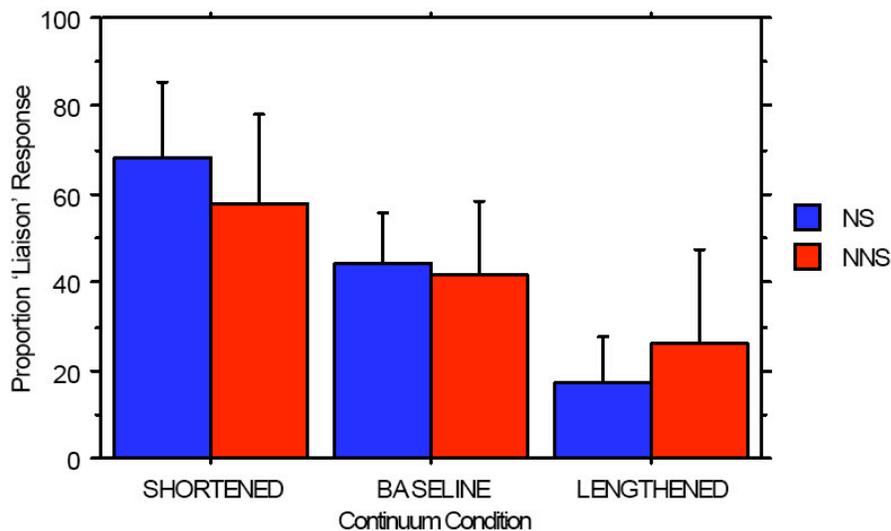
Responses to stimuli containing the segment /t/ were then examined. The proportion and range of 'liaison' responses and the results of single-sample t-tests comparing these proportions to chance performance are displayed below in Table 7-3. An analysis of /t/ responses showed that one NS participant (NS 14) and 11 NNS participants (NNS 3, 5, 7, 8, 10, 12, 14, 15, 16, 17, 18) did not exhibit response patterns in the predicted direction.

Table 7-3: Proportion of 'liaison' responses for NS and NNS participants in forced-choice identification task employing real-word stimuli across three conditions of durational continuum for stimuli containing /t/.

	Shortened /t/	Baseline /t/	Lengthened /t/		Shortened /t/	Baseline /t/	Lengthened /t/
NS 1	66.67	41.67	16.67	NNS 1	91.67	83.33	79.17
NS 2	79.17	45.83	4.17	NNS 2	50.00	45.83	37.50
NS 3	50.00	45.83	29.17	NNS 3	29.17	58.33	0.00
NS 4	83.33	37.50	8.33	NNS 4	50.00	37.50	50.00
NS 5	100.00	54.17	4.17	NNS 5	50.00	87.50	29.17
NS 6	83.33	70.83	16.67	NNS 6	87.50	50.00	25.00
NS 7	54.17	37.50	16.67	NNS 7	54.17	58.33	12.50
NS 8	58.33	54.17	33.33	NNS 8	45.83	45.83	25.00
NS 9	58.33	25.00	8.33	NNS 9	87.50	33.33	8.33
NS 10	75.00	45.83	20.83	NNS 10	54.17	54.17	25.00
NS 11	70.83	58.33	16.67	NNS 11	95.83	75.00	8.33
NS 12	58.33	29.17	20.83	NNS 12	50.00	58.33	0.00
NS 13	50.00	33.33	29.17	NNS 13	45.83	41.67	8.33
NS 14	41.67	45.83	33.33	NNS 14	75.00	79.17	25.00
NS 15	66.67	50.00	29.17	NNS 15	41.67	45.83	16.67
NS 16	50.00	37.50	16.67	NNS 16	45.83	83.33	37.50
NS 17	83.33	33.33	4.17	NNS 17	45.83	25.00	62.50
NS 18	100.00	50.00	4.17	NNS 18	37.50	54.17	16.67
RANGE	41.67 – 100.00	25.00 – 70.83	4.17 – 33.33	RANGE	29.17 – 95.83	25.00 – 87.50	0.00 – 79.17
MEAN (SD)	68.29 (17.28)	44.44 (11.43)	17.36 (10.33)	MEAN (SD)	57.64 (20.27)	41.51 (16.60)	25.93 (21.23)
DIFF FROM CHANCE	p = .0003	p = .0497	p < .0001	DIFF FROM CHANCE	NS	p = .0445	p = .0002

A two-way factorial ANOVA comparing proportions of ‘liaison’ responses between participant groups for the segment /t/ across the three conditions was subsequently performed. This analysis showed a main effect of Continuum Condition ($F(2,102) = 55.11, p < .0001$), but no significant difference between the Participant Groups ($F(1,102) = .222, NS$). Furthermore, the interaction between the two factors just missed significance ($F(2,102) = 3.006, p = .0509$). See Figure 7-5 below.

Figure 7-5: Mean proportion of ‘liaison’ responses in forced-choice identification task employing real-word stimuli across three conditions of durational continuum for stimuli containing the segment /t/ for NS and NNS participants.



For the segment /t/, no significant difference in performance between the two groups was observed. Of note, however, is the fact that response proportions for baseline stimuli for both groups are significantly below chance, pointing to a bias for C-initial words on the part of both participant groups for this segment. We will return to the issue of response bias below in §7.5.1 including an examination of possible factors contributing to this bias.

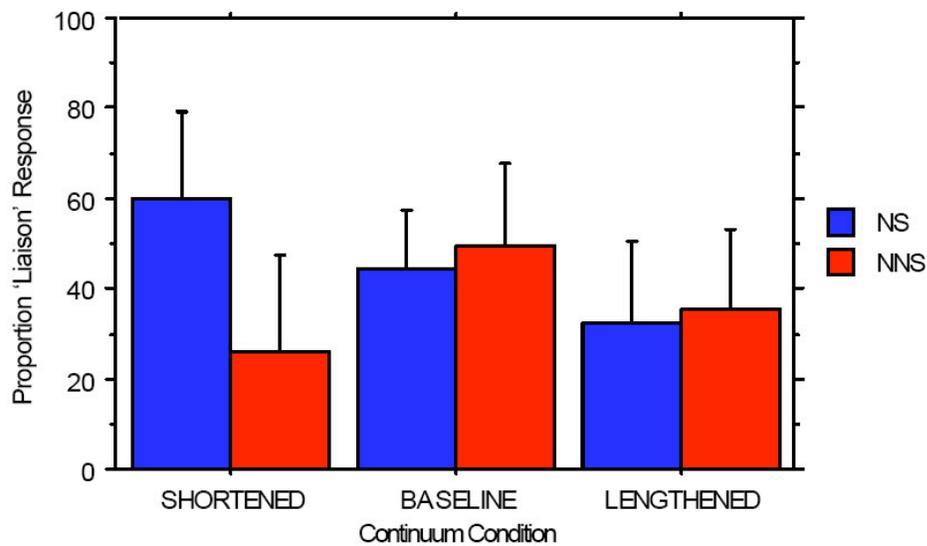
Finally, the proportion and range of ‘liaison’ responses for real-word stimuli containing the segment /z/ were examined. Eight NS participants (NS 1, 7, 8, 10, 13, 14, 15, 16) did not exhibit the predicted response pattern. In addition, surprisingly, each of 18 NNS participants failed to exhibit the predicted linear decline of proportion of liaison response across continuum conditions. These proportions as well as the results of single-sample t-tests comparing these proportions to chance performance are summarized below in Table 7-4.

Table 7-4: Proportion of ‘liaison’ responses for NS and NNS participants in forced-choice identification task employing real-word stimuli across three conditions of durational continuum for stimuli containing /z/.

	Shortened	Baseline	Lengthened		Shortened	Baseline	Lengthened
	/z/	/z/	/z/		/z/	/z/	/z/
NS1	45.83	58.33	37.50	NNS1	79.17	87.50	83.33
NS2	45.83	20.83	8.33	NNS2	37.50	54.17	37.50
NS3	62.50	50.00	16.67	NNS3	0.00	54.17	37.50
NS4	100.00	63.64	33.33	NNS4	50.00	50.00	25.00
NS5	62.50	41.67	4.17	NNS5	29.17	20.83	20.83
NS6	66.67	50.00	33.33	NNS6	25.00	41.67	29.17
NS7	50.00	45.83	62.50	NNS7	12.50	70.83	20.83
NS8	41.67	37.50	41.67	NNS8	25.00	33.33	37.50
NS9	70.83	37.50	29.17	NNS9	8.33	58.33	12.50
NS10	33.33	25.00	29.17	NNS10	25.00	41.67	58.33
NS11	79.17	62.50	29.17	NNS11	8.33	37.50	37.50
NS12	45.83	25.00	12.50	NNS12	0.00	12.50	16.67
NS13	41.67	37.50	37.50	NNS13	8.33	37.50	16.67
NS14	54.17	54.17	41.67	NNS14	25.00	62.50	29.17
NS15	70.83	45.83	58.33	NNS15	16.67	43.48	45.83
NS16	45.83	41.67	66.67	NNS16	37.50	66.67	41.67
NS17	100.00	62.50	41.67	NNS17	62.50	50.00	58.33
NS18	62.50	41.67	4.17	NNS18	16.67	66.67	29.17
RANGE	33.33 – 100.00	20.83 – 63.64	4.17 – 66.67	RANGE	0.00 – 79.17	12.50 – 87.50	12.50 – 83.33
MEAN	59.95	44.51	32.64	MEAN	25.93	49.41	35.42
(SD)	(19.08)	(12.96)	(18.65)	(SD)	(21.23)	(18.28)	(17.81)
DIFF				DIFF			
FROM	p = .0408	NS	p = .0008	FROM	p = .0002	NS	p = .0029
CHANCE				CHANCE			

A two-way factorial ANOVA comparing response proportions for the segment /z/ across the three conditions for each participant group revealed a main effect of Continuum Condition ($F(2,102) = 4.96, p = .0088$), as well as a main effect of Participant Group ($F(1,102) = 6.083, p = .0153$). A significant interaction between the two factors was also observed ($F(2,102) = 13.279, p < .0001$), which is attributable to the NNS pattern of responses for this segment. Proportions of ‘liaison’ responses for only those stimuli including /z/ are shown below in Figure 7-6.

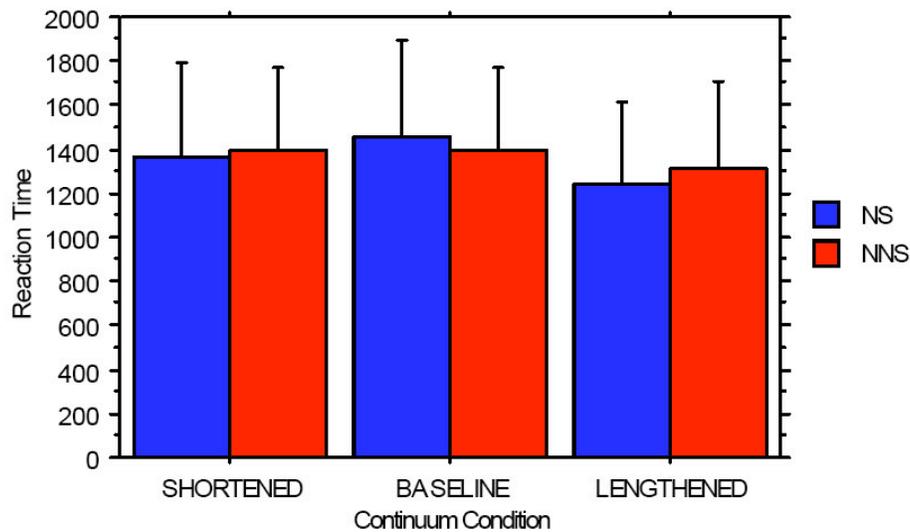
Figure 7-6: Mean proportion of 'liaison' responses in forced-choice identification task employing real-word stimuli across three conditions of durational continuum for stimuli containing the segment /z/ for NS and NNS participants.



7.2.4.2 Reaction Time

Reaction times were also analyzed across the three continuum conditions for each participant group in a two-way factorial ANOVA. As with the AX discrimination tasks presented in Chapter Six, cut-off points were set and reaction times below 200 ms and above 4000 ms were discarded from analysis. This analysis showed no significant effect of Condition ($F(2, 102) = 1.371$, NS), no effect of Participant Group ($F(1, 102) = .036$, NS), and no significant interaction ($F(2, 102) = .038$, NS). Response latencies across continuum conditions for the NS group (mean = 1353.26 ms) were not significantly different than those for the NNS group (mean = 1367.05 ms), nor did reaction times vary significantly across the three continuum conditions for either participant group. Mean reaction times across participants for each group in each condition are shown below in Figure 7-7.

Figure 7-7: Mean reaction times in milliseconds on forced-choice identification task employing real-word stimuli across three conditions of durational continuum for NS and NNS participants.



7.2.5 Discussion: Experiment 2

The results of Experiment 2 suggest that the duration of the pivotal consonant in possible liaison environments can indeed modulate the lexical interpretation of ambiguous sequences. As predicted, responses for both participant groups displayed a significant effect of continuum condition, i.e. both groups chose the V-initial (liaison) target significantly more often when presented with a shortened stimulus and the C-initial target more often when presented with a lengthened stimulus. Crucially, the baseline stimuli, which represent durational values intermediate to those of LCs and ICs, elicited roughly the same amount of V-initial and C-initial responses, indicating a guessing strategy on the part of both groups due to a lack of sufficient acoustic information in the signal. As was also predicted, a large amount of variation was observed across both groups, suggesting that this particular acoustic cue may not be exploited to the same extent across listeners.

Regarding differences among the three segments /n/, /t/ and /z/, the NS group consistently showed the predicted pattern of responses for all three segments, namely a

significant effect of continuum condition characterized by a decrease in the proportion of ‘liaison’ responses across the durational continuum; the shortened stimuli elicited the most ‘liaison’ responses followed by the baseline stimuli, and finally, the lengthened stimuli. However, the predicted differences in perceptual sensitivity among the three segments were not observed. Contrary to the results of the AX discriminations tasks presented in the previous chapter, the NS group did not show differences in sensitivity among the individual segments. Differences among the three segments were observed in the predicted direction, i.e. more sensitivity to /n/ followed by /t/ followed by /z/, however, these differences did not reach significance for the NS group.

The NNS group did show significant differences among the consonants, which is likely attributable in large part to the response pattern for the segment /z/. The predicted effect of continuum condition was observed for the segments /n/ and /t/, i.e. the shortened stimuli elicited the most ‘liaison’ responses followed by the baseline stimuli, and finally, the lengthened stimuli. However, the segment /z/ showed a significantly higher proportion of C-initial responses for both the shortened stimuli and the lengthened stimuli, while baseline responses were at chance levels as shown above in Figure 7-6. At present, we are unable to account for the NNS response pattern for this particular segment. NS responses showed the predicted effect of continuum condition for this segment, which would lead us to believe that the NNS group’s anomalous response pattern is not based on information in the acoustic signal. If NNS responses for the segment /z/ were guided by acoustic information in the signal, the NS group would be expected to display the same response pattern.

We should also note that both the NS and the NNS groups showed a slight bias for C-initial responses for stimuli containing the segment /t/. Though the predicted overall linear effect of continuum condition was observed, the baseline stimuli for this segment showed a significantly higher proportion of C-initial responses for both groups. We explore possible explanations for these response biases below in a series of post-hoc analyses.

Globally, there was no significant performance difference between the NS group and the NNS group, though differences in performance were observed for individual segments. Our results indicate that the NNS participants in this task are performing in a nativelike manner, suggesting that adult learners of French can develop sensitivity to fine-grained acoustic detail and exploit this detail in lexical access. We will discuss the performance of individual NNS participants in relation to NNS biographical factors in the following chapter, however we do wish to point out briefly that two NNS participants (NNS 6 and NNS 9) performed at ceiling for the segment /n/, identifying 100% of shortened /n/ as a liaison environment and 100% of lengthened /n/ as word-initial (see Table 7-2 above), offering strong evidence at the level of the individual of nativelike sensitivity to non-contrastive allophonic variation in spoken French.

In sum, given that the sequences used in the current task were phonemically identical, differing only in the duration of the pivotal consonants, the current data suggest that both native and non-native participants used this single acoustic cue in the localization of word boundaries in these manipulated sequences. However, the amount of variation among both groups in the identification of manipulated stimuli coupled with the

fact that several participants in both groups did not exhibit the predicted response pattern for individual segments bring into question the robustness of this cue in natural speech.

In the next experiment, we examine sensitivity to durational differences in non-word stimuli in order to investigate the effects of the removal of lexical information on the perception and disambiguation of ambiguous sequences.

7.3 Experiment 2.1: Forced-choice Identification Task (non-word stimuli)

7.3.1 Predictions

Experiment 2.1 consists of a forced-choice identification task employing non-word stimuli. As discussed in Chapter Six, non-words are employed in addition to real words in order to investigate whether the removal of higher level information such as lexical frequency and/or semantic acceptability has an effect on the processing of durational variation in environments of liaison. The results of the AX discrimination tasks in Experiments 1 and 1.1 showed that durational differences in real-word stimuli were discriminated significantly better than non-word stimuli by both NS and NNS participants. In addition, reaction times were shorter in the processing of real words than non-words, though this difference only reached significance for the two NS groups.

In §6.4, we discussed possible reasons for heightened sensitivity to acoustic variation in real-word stimuli relative to non-word stimuli in the AX discrimination tasks. We proposed that lexical information associated with real lexical items facilitates the perception of durational variation relative to non-word items through feedback from higher processing levels. Given that the current identification task entails that a lexical decision be made (whereas the discrimination task required only a *same/different* response), we would expect word-superiority effects to be even stronger for real-word stimuli in a forced-choice identification task. If, however, differences in perception favoring non-word stimuli are observed in the identification task, this result would suggest that the same top-down information that facilitates processing at the auditory level (as in an AX discrimination task) can slow processing at the lexical level.

7.3.2 Participants

Thirty-six participants (18 native speakers and 18 non-native speakers) who had participated in Experiment 1.1 took part in Experiment 2.1. See §6.3.2 for a description of these two participant groups.

7.3.3 Stimuli

The stimuli used in Experiment 2.1 consisted of manipulated tokens taken from the three-step durational continuum of non-word stimuli described in §5.4.

7.3.4 Procedure

The procedure in Experiment 2.1 was identical to that of Experiment 2, however this experiment employed non-word stimuli instead of real lexical items. Participants were informed prior to the experiment that stimuli would include words that do not exist in French. See Appendices XI and XII for experimental instructions and sample presentation screen.

7.3.5 Experiment 2.1: Results

Two NS participants (NS 29 and NS 31) and one NNS participant (NNS 21) were removed from analysis due to the fact that they chose either the V-initial or the C-initial response for all 216 trials of the experiment. The following analyses therefore include data from 16 NS participants and 17 NNS participants.

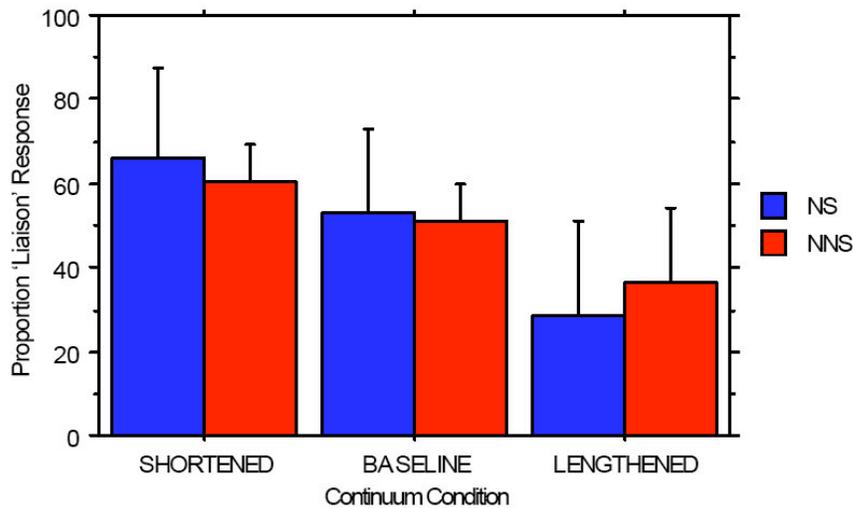
The proportion of ‘liaison’ (i.e. V-initial) responses was calculated for each participant for the three steps of the durational continuum: the shortened (liaison) version, the baseline version and lengthened (C-initial) version. Response proportions for each participant, along with mean proportions across participants, standard deviations, range and the results of single sample t-tests comparing response proportions to chance performance (50%) are presented below in Table 7-5. Note that five NS participants (NS 19, 20, 21, 22, 27) and three NNS participants (28, 32, 33) did not exhibit the predicted response pattern, i.e. a linear decline in the proportion of ‘liaison’ responses was not observed across continuum conditions.

Table 7-5: *Proportion of ‘liaison’ responses for NS and NNS participants in forced-choice identification task employing non-word stimuli across three conditions of durational continuum.*

	Shortened	Baseline	Lengthened		Shortened	Baseline	Lengthened
NS 19	16.67	8.33	8.33	NNS 19	70.83	55.56	29.17
NS 20	97.22	94.37	76.39	NNS 20	65.28	50.00	37.50
NS 21	44.44	41.67	45.83	NNS 21	removed from analysis		
NS 22	51.39	52.78	12.50	NNS 22	58.33	55.56	15.28
NS 23	69.44	43.06	19.44	NNS 23	47.22	36.11	15.28
NS 24	72.22	68.06	31.94	NNS 24	59.72	40.28	37.50
NS 25	80.56	54.17	11.11	NNS 25	56.94	50.00	41.67
NS 26	70.42	55.56	27.78	NNS 26	70.42	56.94	43.06
NS 27	44.44	44.44	29.17	NNS 27	65.28	45.83	18.06
NS 28	51.43	43.48	39.44	NNS 28	41.67	61.11	68.06
NS 29	removed from analysis			NNS 29	63.89	54.17	51.39
NS 30	65.71	55.71	44.44	NNS 30	58.33	41.67	20.83
NS 31	removed from analysis			NNS 31	58.33	41.67	12.50
NS 32	50.00	49.30	13.89	NNS 32	73.61	61.11	61.11
NS 33	68.06	51.39	41.67	NNS 33	56.94	62.50	58.33
NS 34	72.22	51.39	2.78	NNS 34	69.44	55.56	33.33
NS 35	84.72	45.83	4.17	NNS 35	65.28	59.72	52.78
NS 36	86.27	48.33	5.88	NNS 36	47.22	38.89	23.61
RANGE	16.67 – 97.22	8.33 – 94.37	2.78 – 76.39	RANGE	41.67 – 73.61	36.11 – 62.50	12.50 – 68.06
MEAN	63.98	50.33	25.82	MEAN	60.51	50.98	36.44
(SD)	(19.97)	(17.02)	(20.24)	(SD)	(8.96)	(8.68)	(17.55)
DIFF FROM CHANCE	p = .0134	NS	p = .0002	DIFF FROM CHANCE	p = .0002	NS	p = .0057

A factorial ANOVA was conducted in order to compare these proportions across the three continuum conditions for each participant group. A main effect of Continuum Condition was observed ($F(2, 96) = 27.263, p < .0001$), however there was no significant difference between the two Participant Groups ($F(1, 96) = .001, NS$) and no significant interaction ($F(2, 96) = 1.369, NS$). The proportions of ‘liaison’ responses for the three continuum conditions for non-word stimuli for both participant groups are shown graphically below in Figure 7-8.

Figure 7-8: Mean proportion of ‘liaison’ responses in forced-choice identification task employing non-word stimuli across three conditions of durational continuum for NS and NNS participants.

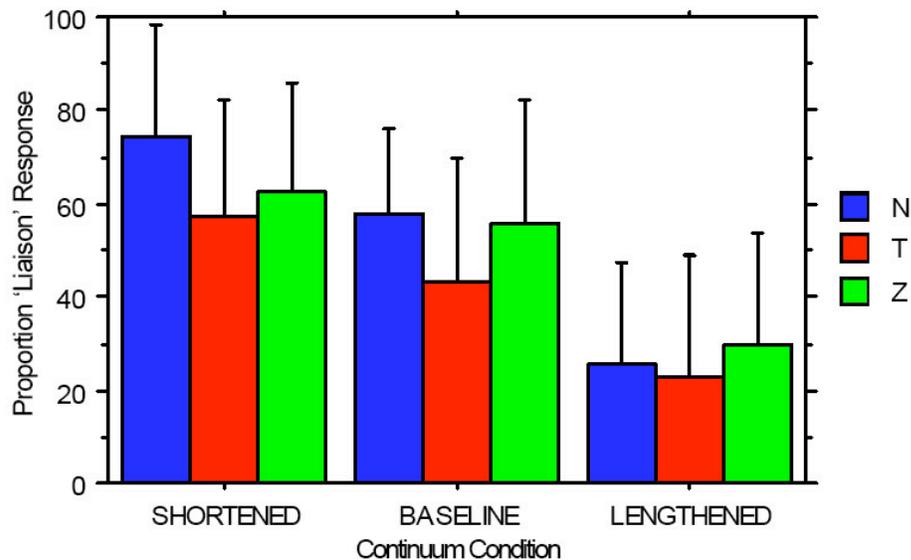


In line with the results of Experiment 2, the results of the forced-choice identification task employing non-word stimuli suggest that the duration of the pivotal consonant in ambiguous liaison environments can modulate the lexical interpretation of these sequences even when higher level lexical information is removed from W_2 in two-word ambiguous liaison sequences.

7.3.5.1 Analyses by Consonant

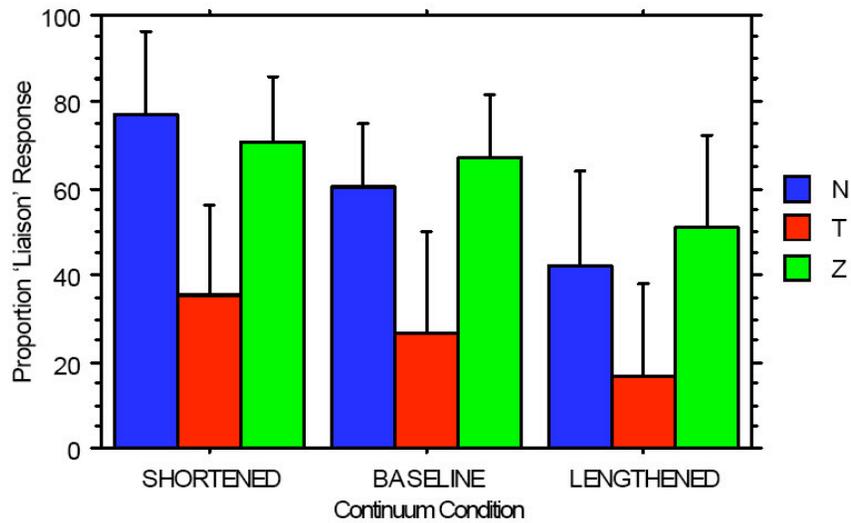
Possible differences among responses for the three segments /n, t, z/ for the two participant groups were also considered for non-word stimuli. First, proportions of ‘liaison’ responses were compared across the three consonants for each continuum condition for the NS group, which showed significant differences across the three Continuum Conditions ($F(2, 140) = 33.510, p < .0001$) and a significant difference among the Consonants ($F(2, 140) = 3.170, p = .0450$). No significant interaction was found ($F(4, 140) = .679, NS$). Specific comparisons in Scheffe post-hoc tests showed a significant difference only between the segments /n/ and /t/ ($p = .0447$). Proportion ‘liaison’ responses across conditions and consonants for the NS group are shown below in Figure 7-9.

Figure 7-9: Mean proportion of ‘liaison’ responses in forced-choice identification task employing non-word stimuli across three continuum conditions and consonants for NS participants.



We then analyzed proportions of ‘liaison’ responses across the three consonants for each continuum condition for the NNS group, which also showed a significant effect of Continuum Condition ($F(2, 141) = 20.400, p < .0001$) as well as a significant difference among the Consonants ($F(2, 141) = 55.517, p < .0001$). No significant interaction was found ($F(4, 141) = 1.082, NS$). Post-hoc specific comparisons showed that responses for the segment /n/ were significantly different from those for /t/ ($p < .0001$) and /t/ was significantly different from /z/ ($p < .0001$). There was no difference, however, between /n/ and /z/. Note that the segment /t/ elicited a significantly lower proportion of V-initial responses across all three continuum conditions than /n/ or /z/, again pointing to a C-initial bias for this segment. Proportion ‘liaison’ responses across conditions and consonants for the NNS group are shown below in Figure 7-10.

Figure 7-10: Mean proportion of ‘liaison’ responses in forced-choice identification task employing non-word stimuli across three continuum conditions and consonants for NNS participants.



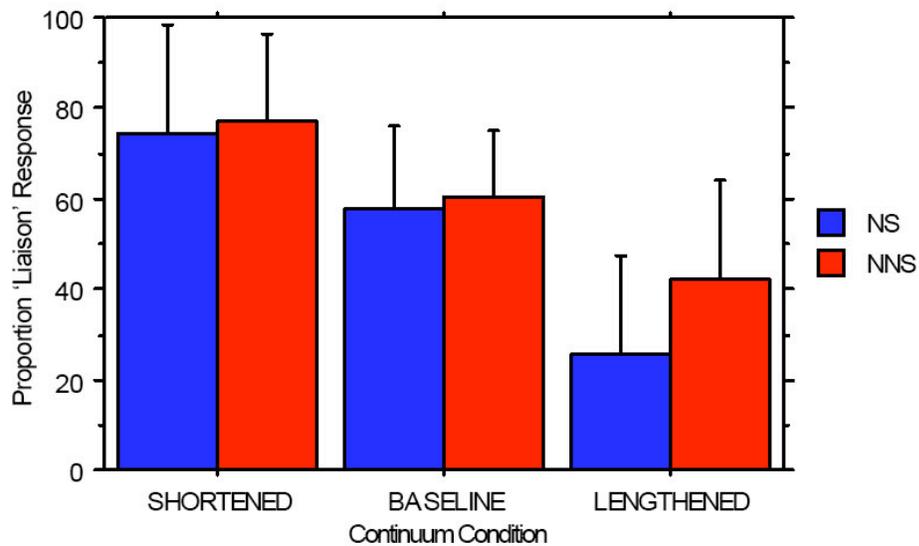
NS and NNS responses were then analyzed separately for each individual consonant. Responses to non-word stimuli containing only the segment /n/ were examined first. The proportion of ‘liaison’ (V-initial) responses was calculated for each of the three steps of the durational continuum: the shortened (liaison) version, the baseline version and lengthened (C-initial) version. Proportions of ‘liaison’ responses for the segment /n/ show that seven NS participants (NS 19, 21, 22, 26, 28, 30, 32) and five NNS participants (NNS 22, 25, 28, 28, 33) did not exhibit the predicted linear decline across continuum conditions. Note as well that baseline responses for the NNS group are significantly above chance, suggesting a V-initial response bias for this group. Response proportions for each participant, along with mean proportions across participants, standard deviations, range and the results of single sample t-tests comparing response proportions to chance performance (50%) are presented below in Table 7-6.

Table 7-6: Proportion of ‘liaison’ responses for NS and NNS participants in forced-choice identification task employing non-word stimuli across three conditions of durational continuum for stimuli containing /n/.

	Shortened /n/	Baseline /n/	Lengthened /n/		Shortened /n/	Baseline /n/	Lengthened /n/
NS 19	37.50	20.83	20.83	NNS 19	83.33	58.33	33.33
NS 20	100.00	95.65	66.67	NNS 20	95.83	66.67	45.83
NS 21	37.50	37.50	41.67	NNS 21	removed from analysis		
NS 22	58.33	66.67	8.33	NNS 22	83.33	83.33	25.00
NS 23	79.17	50.00	20.83	NNS 23	70.83	54.17	16.67
NS 24	83.33	70.83	37.50	NNS 24	95.83	54.17	45.83
NS 25	100.00	54.17	8.33	NNS 25	91.67	54.17	58.33
NS 26	62.50	62.50	16.67	NNS 26	91.67	75.00	58.33
NS 27	41.67	33.33	8.33	NNS 27	83.33	50.00	12.50
NS 28	65.22	68.18	47.83	NNS 28	20.83	45.83	79.17
NS 29	removed from analysis			NNS 29	58.33	33.33	41.67
NS 30	73.91	50.00	62.50	NNS 30	75.00	70.83	33.33
NS 31	removed from analysis			NNS 31	79.17	50.00	12.50
NS 32	50.00	54.17	8.33	NNS 32	75.00	66.67	62.50
NS 33	79.17	58.33	54.17	NNS 33	54.17	54.17	54.17
NS 34	100.00	54.17	0.00	NNS 34	87.50	50.00	25.00
NS 35	100.00	66.67	12.50	NNS 35	95.83	91.67	83.33
NS 36	100.00	54.17	0.00	NNS 36	70.83	66.67	33.33
RANGE	37.50 – 100.00	20.83 – 95.65	0.00 – 66.67	RANGE	20.83 – 95.83	33.33 – 91.67	12.50 – 83.33
MEAN	73.02	56.07	25.91	MEAN	77.20	60.29	42.40
(SD)	(23.45)	(17.12)	(22.44)	(SD)	(19.10)	(14.52)	(21.51)
DIFF				DIFF			
FROM	p = .0006	NS	p = .0003	FROM	p < .0001	p = .009	NS
CHANCE				CHANCE			

A factorial ANOVA was employed to compare response proportions for the segment /n/ across the three continuum conditions for each participant group. This analysis showed a main effect of Continuum Condition ($F(2, 96) = 37.925, p < .0001$). However, the difference between the two Participant Groups fell short of significance ($F(1, 96) = 3.422, p = .0674$) and there was no interaction between the two factors ($F(2, 96) = 1.496, NS$). Proportions of ‘liaison’ responses for non-word stimuli for all three conditions for both participant groups are shown below in Figure 7-11.

Figure 7-11: Mean proportion of 'liaison' responses in forced-choice identification task employing non-word stimuli across three conditions of durational continuum for stimuli containing the segment /n/ for NS and NNS participants.



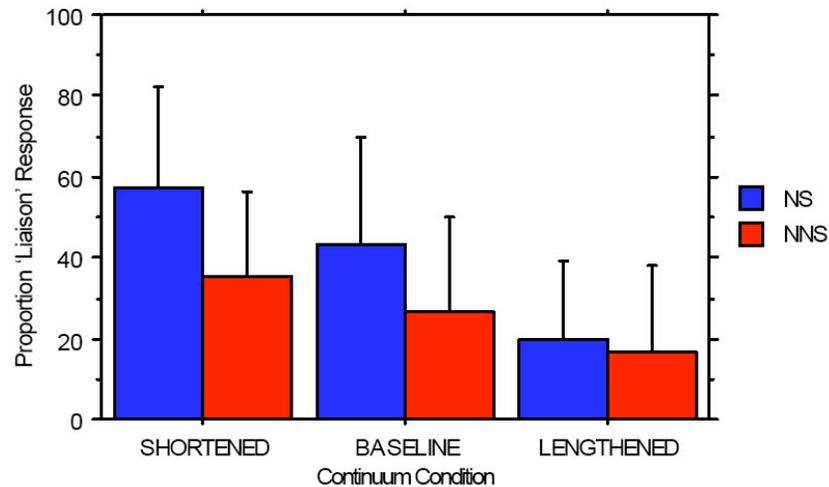
Responses to non-word stimuli containing only the segment /t/ were then examined for both participant groups. The proportions of 'liaison' (V-initial) responses for each of the three steps of the durational continuum are presented below in Table 7-7, along with mean proportions, standard deviations, range and the results of single sample t-tests comparing response proportions to chance performance (50%). Five NS participants (NS 19, 21, 24, 26, 27) and ten NNS participants (NNS 20, 23, 24, 25, 28, 29, 32, 33, 35, 36) did not show response patterns in the predicted direction.

Table 7-7: Proportion of ‘liaison’ responses for NS and NNS participants in forced-choice identification task employing non-word stimuli across three conditions of durational continuum for stimuli containing /t/.

	Shortened /t/	Baseline /t/	Lengthened /t/		Shortened /t/	Baseline /t/	Lengthened /t/
NS 19	12.50	0.00	4.17	NNS 19	58.33	41.67	12.50
NS 20	95.83	87.50	66.67	NNS 20	8.33	4.17	8.33
NS 21	4.17	0.00	0.00	NNS 21	removed from analysis		
NS 22	45.83	33.33	16.67	NNS 22	20.83	4.17	0.00
NS 23	79.17	33.33	16.67	NNS 23	8.33	0.00	0.00
NS 24	58.33	58.33	29.17	NNS 24	0.00	0.00	0.00
NS 25	70.83	54.17	8.33	NNS 25	25.00	33.33	20.83
NS 26	70.83	25.00	25.00	NNS 26	30.43	20.83	0.00
NS 27	54.17	66.67	37.50	NNS 27	45.83	33.33	8.33
NS 28	45.83	41.67	37.50	NNS 28	58.33	79.17	66.67
NS 29	removed from analysis			NNS 29	58.33	33.33	54.17
NS 30	58.33	56.52	45.83	NNS 30	41.67	8.33	0.00
NS 31	removed from analysis			NNS 31	79.17	50.00	33.33
NS 32	50.00	45.83	8.33	NNS 32	54.17	29.17	29.17
NS 33	54.17	33.33	16.67	NNS 33	58.33	62.50	58.33
NS 34	45.83	25.00	0.00	NNS 34	58.33	33.33	16.67
NS 35	79.17	50.00	0.00	NNS 35	4.17	0.00	0.00
NS 36	45.83	25.00	0.00	NNS 36	4.17	8.33	4.17
RANGE	4.17 – 95.83	0.00 – 87.50	0.00 – 66.67	RANGE	4.17 – 58.33	0.00 – 79.17	0.00 – 66.67
MEAN	54.43	39.73	19.53	MEAN	35.23	26.82	16.67
(SD)	(23.25)	(22.98)	(19.53)	(SD)	(21.17)	(23.27)	(21.14)
DIFF				DIFF			
FROM	NS	NS	p < .0001	FROM	p = .0137	p = .0012	p < .0001
CHANCE		(p = .0941)		CHANCE			

A factorial ANOVA was conducted to compare response proportions for the segment /t/ across the three continuum conditions for each participant group. There was a main effect of Continuum Condition ($F(2, 92) = 12.246, p < .0001$) and a main effect of Participant Group ($F(1, 92) = 8.724, p = .0040$), but no interaction ($F(2, 92) = 1.467, NS$). The proportions of ‘liaison’ responses for all three conditions for non-word stimuli for both participant groups are shown graphically in Figure 7-12 below.

Figure 7-12: Mean proportion of ‘liaison’ responses in forced-choice identification task employing non-word stimuli across three conditions of durational continuum for stimuli containing the segment /t/ for NS and NNS participants.



Furthermore, response proportions for both participant groups point to a strong C-initial bias for this segment. Note that, even though the predicted linear decline of response proportions across continuum conditions was obtained, responses across all three continuum conditions for both participant groups are either at or significantly below chance performance, indicating a strong response bias toward C-initial targets. Recall as well that a significant C-initial bias was observed for the segment /t/ in Experiment 2, which employed real-word stimuli. We return to possible explanations of response bias below in §7.5.1.

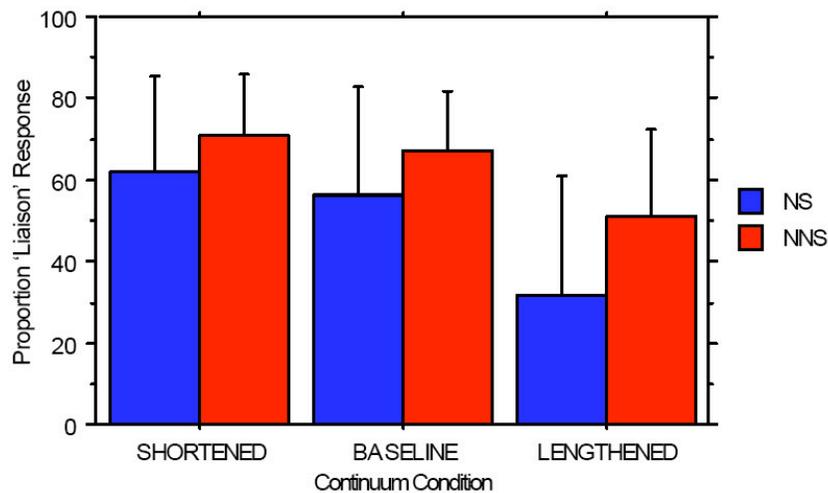
Finally, response proportions for non-word stimuli containing only the segment /z/ were examined for both participant groups. The proportion of ‘liaison’ (i.e. V-initial) responses was calculated for each of the three steps of the durational continuum: the shortened (liaison) version, the baseline version and lengthened (C-initial) version. These proportions are shown below in Table 7-8 along with mean proportions, standard deviations, range and the results of single sample t-tests comparing response proportions to chance performance (50%). Ten NS participants (NS 19, 20, 21, 22, 24, 26, 27, 28, 34, 36) and nine NNS participants (NNS 22, 24, 25, 28, 29, 31, 32, 33, 34) did not exhibit the expected response pattern. Note as well that response proportions for the NNS group for this segment showed a response bias in the opposite direction than what was observed for the segment /t/. NNS baseline response proportions were above chance, while response proportions for lengthened stimuli were not significantly different from chance, indicating a bias for V-initial responses.

Table 7-8: Proportion of 'liaison' responses for NS and NNS participants in forced-choice identification task employing non-word stimuli across three conditions of durational continuum for stimuli containing /z/.

	Shortened	Baseline	Lengthened		Shortened	Baseline	Lengthened
	/z/	/z/	/z/		/z/	/z/	/z/
NS 19	0.00	4.17	0.00	NNS 19	70.83	66.67	41.67
NS 20	95.83	100.00	95.83	NNS 20	91.67	79.17	58.33
NS 21	91.67	87.50	95.83	NNS 21	removed from analysis		
NS 22	50.00	58.33	12.50	NNS 22	70.83	79.17	20.83
NS 23	50.00	45.83	20.83	NNS 23	62.50	54.17	29.17
NS 24	75.00	75.00	29.17	NNS 24	83.33	66.67	66.67
NS 25	70.83	54.17	16.67	NNS 25	54.17	62.50	45.83
NS 26	78.26	79.17	41.67	NNS 26	87.50	75.00	70.83
NS 27	37.50	33.33	41.67	NNS 27	66.67	54.17	33.33
NS 28	43.48	21.74	33.33	NNS 28	45.83	58.33	58.33
NS 29	removed from analysis			NNS 29	58.33	33.33	79.17
NS 30	65.22	60.87	25.00	NNS 30	58.33	45.83	29.17
NS 31	removed from analysis			NNS 31	79.17	50.00	62.50
NS 32	50.00	47.83	25.00	NNS 32	91.67	87.50	91.67
NS 33	70.83	62.50	54.17	NNS 33	58.33	70.83	62.50
NS 34	70.83	75.00	8.33	NNS 34	62.50	83.33	58.33
NS 35	75.00	20.83	0.00	NNS 35	95.83	87.50	75.00
NS 36	70.83	75.00	8.33	NNS 36	66.67	41.67	33.33
RANGE	0.00 – 95.83	4.17 – 100.00	0.00 – 95.83	RANGE	45.83 – 95.83	41.67 – 87.50	20.83 – 91.67
MEAN	62.21	56.33	31.77	MEAN	71.08	67.40	51.22
(SD)	(23.34)	(26.33)	(29.26)	(SD)	(14.69)	(14.38)	(21.29)
DIFF				DIFF			
FROM	p = .0538	NS	p = .0249	FROM	p < .0001	p = .0001	NS
CHANCE				CHANCE			

A factorial ANOVA was subsequently performed for stimuli containing /z/ comparing response proportions across the three continuum conditions for each participant group. There was a main effect of Continuum Condition ($F(2, 93) = 12.037, p < .0001$), as well as a main effect of Participant Group ($F(1, 93) = 8.733, p = .0040$) and no significant interaction ($F(2, 93) = .526, NS$). See Figure 7-13 below.

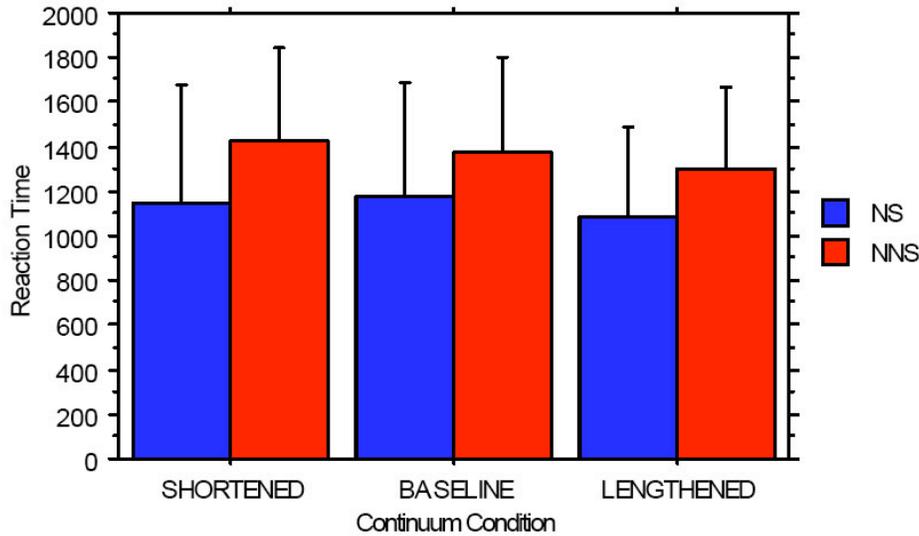
Figure 7-13: Mean proportion of 'liaison' responses in forced-choice identification task employing non-word stimuli across three conditions of durational continuum for stimuli containing the segment /z/ for NS and NNS participants.



7.3.5.2 Reaction Time

Reaction times were also compared between the two participant groups for responses to non-word stimuli in the identification task. Cut-off points were set and reaction times below 200 ms and above 4000 ms were discarded from analysis. Reaction times for each of the three continuum conditions were analyzed in a two-way factorial ANOVA, which showed a significant effect for Participant Group ($F(1, 96) = 7.023, p = .0094$), no significant effect of Condition ($F(2, 96) = .440, NS$) and no interaction between the two factors ($F(2, 96) = .081, NS$). This analysis showed that NS reaction times (mean = 1131.71 ms) were significantly shorter than NNS reaction times (mean = 1368.55 ms); however neither group had significantly different reaction times across the three continuum conditions as shown below in Figure 7-14.

Figure 7-14: Mean reaction times in milliseconds on forced-choice identification task employing non-word stimuli across three conditions of durational continuum for NS and NNS participants.



7.3.6 Discussion: Experiment 2.1

The results of Experiment 2.1, which employed non-word stimuli, are in line with those of Experiment 2, which employed real-word stimuli. The duration of the pivotal consonant in possible liaison environments can modulate the interpretation of ambiguous sequences even when lexical information is removed. As predicted, a significant effect of continuum condition was observed, i.e. both the NS and NNS participant groups chose the V-initial (liaison) target significantly more often when presented with a shortened stimulus and the C-initial target significantly more often when presented with a lengthened stimulus. Furthermore, the baseline stimuli globally elicited roughly the same amount of V-initial and C-initial responses, again pointing to a guessing strategy adopted by participants due to a lack of sufficient acoustic information in the signal.

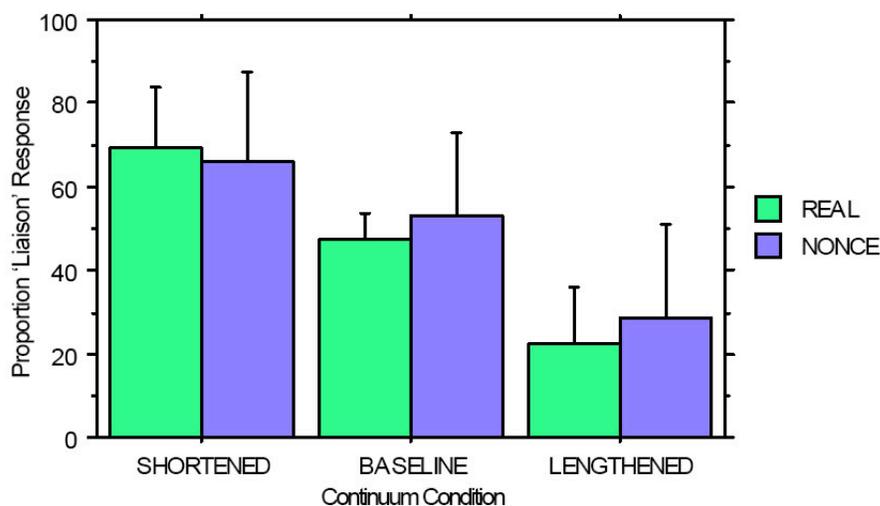
Regarding the individual segments /n/, /t/ and /z/, both participant groups showed the predicted pattern of responses for each of the three consonants, though to varying degrees. Shortened stimuli elicited the most ‘liaison’ responses followed by baseline stimuli, and, finally, lengthened stimuli. It should be noted however that though a consistent linear decline in ‘liaison’ responses was observed across the three continuum conditions, the segment /t/ elicited a significantly higher proportion of C-initial responses for both groups (see Table 7-7 above). The NNS group in particular showed a considerably higher proportion of /t/-initial responses across all three continuum conditions, while the segment /z/ for the NNS group elicited a higher proportion of V-initial responses. We will revisit possible sources of these biases in a more in-depth discussion of response biases for individual segments below in §7.5.1.1.

As in Experiment 2, no significant global difference emerged between the NS group and the NNS group, though differences were again observed for individual consonants. Though a substantial degree of variation in individual performance was observed across all participants, the results of Experiment 2.1 taken together with those of Experiment 2 suggest that adult learners of L2 French can develop nativelike sensitivity to non-contrastive allophonic variation.

7.4 Discussion: Experiments 2 and 2.1 (real versus non-word stimuli)

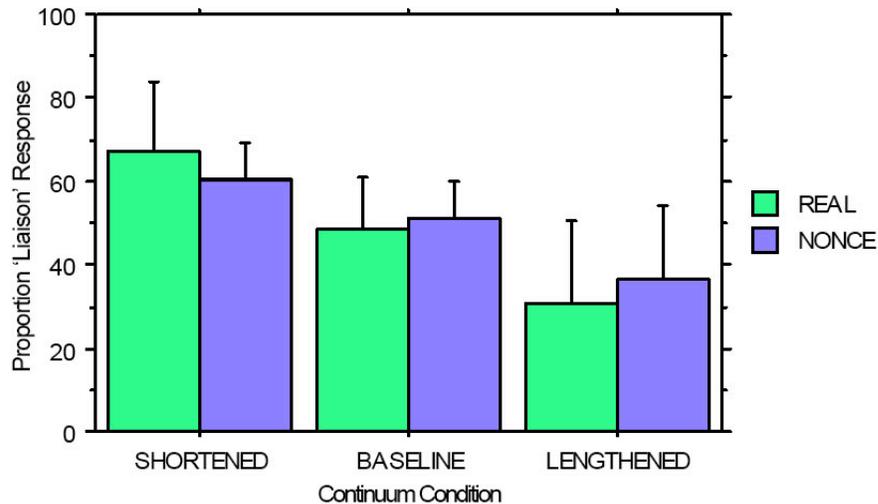
Given that significant differences were observed in responses to real-word stimuli and non-word stimuli in the AX discrimination tasks presented in Chapter Six, the possibility of differences in performance between the two word types was also explored in the current identification tasks. We first compared the proportion of ‘liaison’ responses for the NS group tested on real-word stimuli in Experiment 2 and the NS group tested on non-word stimuli in Experiment 2.1 across all three continuum conditions. Mean response proportions are shown below in Figure 7-15. A factorial ANOVA did not reveal a significant difference between the two NS groups tested on the two stimulus types ($F(1, 99) = .722$, NS).

Figure 7-15: Mean proportion of ‘liaison’ responses in forced-choice identification tasks for NS group tested in Experiment 2 (real words) and NS group tested in Experiment 2.1 (non-words) across three continuum conditions.



Responses were then examined for the NNS group tested on real-word stimuli in Experiment 2 and the NNS group tested on non-word stimuli in Experiment 2.1, which also failed to reveal a significant difference between the two groups ($F(1, 99) = .037$, NS). Mean proportions of ‘liaison’ responses for the two NNS groups are given below in Figure 7-16.

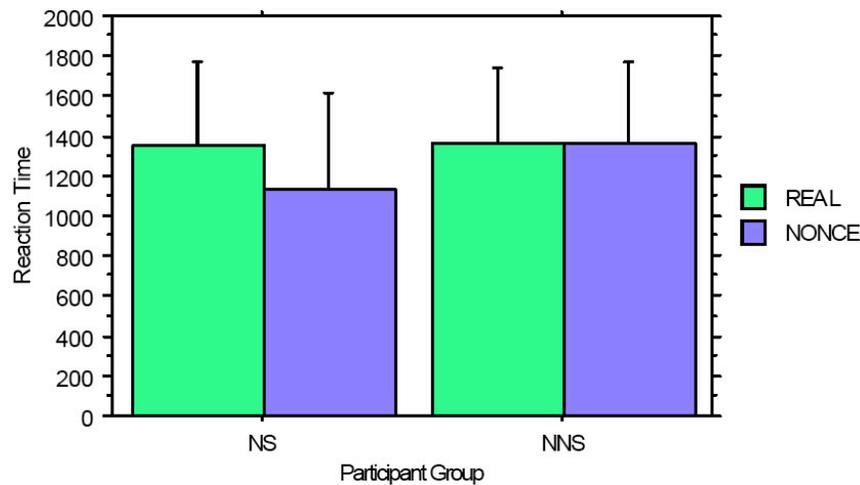
Figure 7-16: Mean proportion of 'liaison' responses in forced-choice identification tasks for NNS group tested in Experiment 2 (real words) and NNS group tested in Experiment 2.1 (non-words) across three continuum conditions.



Contrary to what was observed in the AX discrimination tasks in Experiments 1 and 1.1, in which both NS and NNS groups showed diminished sensitivity to segmental duration in non-word stimuli relative to real-word stimuli, neither the NS nor the NNS participants performed differently in the perception of real words and non-words in the identification task.

We also analyzed reaction times for each word type across participant groups. The effect of Word Type just missed significance ($F(1,206) = 3.523, p = .0619$), while a significant effect of Participant Group was observed ($F(1,206) = 4.632, p = .0325$). The interaction between the two factors also just fell short of significant ($F(1,206) = 3.616, p = .0586$), suggesting that the effect of Word Type was not equivalent for the two participant populations. Reaction times were significantly higher for the NS group tested on real-word stimuli in Experiment 2 than for the NS group tested on non-word stimuli in Experiment 2.1. The two NNS groups, on the other hand, showed no difference in response latencies between the two stimulus types employed in Experiments 2 and 2.1, as shown below in Figure 7-17.

Figure 7-17: Mean reaction times in milliseconds on forced-choice identification tasks in Experiments 2 (real words) and 2.1 (non-words) for NS and NNS groups.



This analysis revealed an interesting result, namely that the NS group tested on real-word stimuli exhibited *longer* reaction times than the NS group tested on non-word stimuli. Recall that these same two NS groups showed a reverse reaction time pattern in the AX discrimination task in Experiments 1 and 1.1; response latencies in the AX task were longer for the NS group tested on non-word stimuli than for the NS group tested on real-word stimuli. This reversal would suggest that, in acoustic-level processing such as that required by an AX discrimination task, the pre-existence of lexical representations as in the case of real words can facilitate processing, while in an identification task, which requires acoustic processing in addition to higher-order decision making, access to lexical information associated with existing mental representations may slow processing. Arguably, the NS group tested on non-word stimuli in the identification task in Experiment 2.1 may have continued to parse non-word sequences at the same acoustic level of processing required in the AX discrimination task in Experiment 1.1, focusing primarily on surface features of the signal. Conversely, the NS group tested on real-word stimuli may have found that the processing of real lexical items in the identification task in Experiment 2 was slowed by the interference of top-down information associated with each pair of items (e.g. lexical frequency, semantic and syntactic plausibility) relative to the (acoustic) processing of real-word items in the AX discrimination task in Experiment 1.

Interestingly, contrary to what was observed for the two NS groups, the NNS group tested on real-word stimuli in Experiment 2 and the NNS group tested on non-word stimuli in Experiment 2.1 showed no significant difference in reaction times. Response

latencies were not significantly different for participants tested on real words relative to participants tested on non-word stimuli in the forced-choice identification tasks. Of particular interest is the fact that these same two NNS participant groups also failed to exhibit a significant difference in response latencies in the processing of non-word stimuli relative to real-word stimuli in the AX discriminations tasks.

The differences in reaction time patterns observed in the two behavioral tasks between the two NS groups and the two NNS groups would suggest that both the NNS group tested on real words and the NNS group tested on non-words may have been processing both real- and non-word stimuli in both behavioral tasks (AX discrimination and identification) at the same lower acoustic level. The NS group tested on real-word stimuli, however, may have invoked higher-level processing and decision-making in the identification task. The NS group tested on non-word stimuli, on the other hand, may have continued to process non-lexical items at a lower level, resulting in lower reaction times relative to real lexical items. On this logic, NNS participants may have a relative (speed) advantage in processing the signal at the auditory level in that they are not required to suppress possible top-down influence of plausibility and/or familiarity to the same degree as native speakers. Supporting this hypothesis is the fact that reaction times in the AX discrimination task in Experiment 1.2 were lower for the BEG group relative to both the NS and the NNS groups tested on non-word stimuli (see §6.5.4.4 above).

7.5 Post-hoc Analyses

The purpose of Experiments 2 and 2.1 was to examine the effect of segmental duration on the lexical interpretation of globally ambiguous phrases in spoken French. A robust effect for duration was found, demonstrating that segmental duration alone can bias interpretation of ambiguous sequences in spoken French. Nonetheless, the above results indicate not only a large amount of variation, but also considerable response biases on the part of both NS and NNS participants. In the current section we examine these biases more closely and explore possible sources of bias in an attempt to shed additional light on the processing strategies of both native and non-native speakers in the disambiguation of liaison environments.

7.5.1 Response Bias

Previous work on the perception of liaison by NS and NNS of French has found response preferences for either V-initial words (liaison environments) or C-initial words in behavioral tasks in ambiguous minimal pairs such as those employed here. Shoemaker and Birdsong (2008; see also Shoemaker 2005), who employed a forced-choice identification task using unaltered minimal-pair phrases found that NS participants selected the C-initial target in 56.53% of experimental trials, which was significantly more than the V-initial target ($p = .02$). The NNS group in that study, however, showed no bias for either target type. As discussed above in §4.3, these authors conducted several analyses post-hoc in order to determine what factor was guiding participants' responses, but were unable to pinpoint the source of the NS bias; neither lexical frequency, nor semantic plausibility predicted responses.

Stridfeldt's (2003) data, which are also discussed in §4.3 above and was based on unaltered ambiguous non-word minimal pairs (e.g. *un auve* vs. *un nauve* [œ.nov]), exhibited an alternative trend; this study found a strong response bias on the part of the NNS participants (L1 Swedish), who showed a significant preference for the V-initial member (e.g. *un auve*) within each minimal pair of ambiguous non-word stimuli. The native speakers in Stridfeldt's study showed a response pattern in the same direction in that response proportions were higher for V-initial words, but this difference failed to reach statistical significance. Stridfeldt attributed the V-initial bias on the part of NNS participants to an overgeneralization of L2 strategies acquired for the processing of liaison consonants. The NNS had in effect assumed that all possible environments of liaison were instantiations of liaison.

The analyses of response proportions in the current study (see §7.2.4 and §7.3.5 above) point to biases on the part of both participant groups. While the proportions of V-initial responses in Experiments 2 and 2.1 were calculated separately for each of the three continuum conditions (shortened, baseline, and lengthened consonants), in the current section we investigate whether participants displayed a global bias across all three continuum conditions for either V-initial (liaison) or the C-initial responses in the identification task. Table 7-9 below summarizes the proportions of V-initial and C-initial responses for NS and NNS participant groups in Experiments 2 and 2.1. The table also includes the results of paired t-tests showing whether response proportions differ significantly within each experiment for each participant group. As can be seen below, the only significant difference that emerged was for the NS group tested on real-word stimuli in Experiment 2, who showed a significant bias toward C-initial words. This finding is in line with Shoemaker and Birdsong (2008) above who also found a preference for C-initial lexical items by native speakers, but no such preference for non-native speakers. The NNS group tested on non-word stimuli in Experiment 2.1 displayed a similar bias toward C-initial responses, however the difference failed to reach significance.

Table 7-9: *Proportion of V-initial versus C-initial responses Experiments 2 and 2.1 and results of paired t-tests comparing proportions.*

	Native Speakers			Non-Native Speakers		
	V-initial (liaison) response	C-initial response	Value of p	V-initial (liaison) response	C-initial response	Value of p
Experiment 2 (real-word stimuli)	46.37 %	53.63 %	p = .0046	49.43 %	50.57 %	NS
Experiment 2.1 (non-word stimuli)	50.14 %	49.86 %	NS	46.99 %	53.01 %	NS

Two explanations for the observed bias toward C-initial items present themselves. The first possibility is based on language usage and frequency. Bybee (2001b) notes that C-initial words are twice as common as V-initial words in French. Similarly, Adda-Decker et al. (2002) report that, of 14 different syllable structures observed in a corpus of 30 hours of spoken French, only three are V-initial. Given this distribution, native speakers may develop a higher expectation for encountering C-initial words in continuous speech, therefore biasing responses. However, specifically in the case of liaison, Nguyen et al. (2003) note that words following a potential liaison environment (e.g. *un* [œ̃] or *grand* [gʁɑ̃] or any word that can trigger liaison) are more likely to begin with one of many vowel sounds than with one particular consonant. In other words, in the case of the determiner *un* [œ̃] for example, the following word is more likely to begin with one of several possible vowels, than with the particular segment /n/.

An alternative account for a C-initial response bias is based purely on the acoustic signal. Recall that manipulated tokens employed in the behavioral tasks were produced as consonant initial (see §5.4 above), e.g. all manipulated tokens of [œ̃.nɛʁ] were originally produced as *un nerf* and not *un air*. Therefore, it is plausible that there are acoustic cues in the stimuli other than the duration of pivotal consonants, which we have not yet addressed, that are biasing NS responses toward a C-initial interpretation. For example, as discussed in Chapter Five, in the current production sample vowels preceding LCs were found to be significantly shorter than vowels preceding ICs. Given that the C-initial bias was observed for NS participants only, this would lead us to believe that native listeners are sensitive to some additional acoustic cue to which the non-native listeners are not.

We investigate these and other possibilities below in a more detailed analysis of response biases by individual consonant and by individual item.

7.5.1.1 Response Bias by Consonant

Given the degree of variation observed in participant performance among the three segments examined in this study, response biases for each individual consonant were also calculated, the results of which are summarized below in Table 7-10. This analysis revealed several biases in both directions (V-initial and C-initial) on the part of both NS and NNS participants. Cells shaded in grey indicate where NS groups and NNS groups were consistent in the directionality of response proportions for individual segments and where this difference was significant for both groups. Note that /t/ is the only of the three segments to elicit a consistent bias across all four groups. This segment showed a strong bias toward the C-initial member of each ambiguous pair.

Table 7-10: *Proportion of vowel-initial versus consonant-initial responses in forced-choice identification tasks in Experiments 2 and 2.1 for each of three consonants and results of paired t-tests comparing proportions. Cells shaded in grey indicate where NS groups and NNS groups were consistent in the directionality of response proportions for individual segments and where this difference was significant for both groups.*

	Native Speakers			Non-Native Speakers				
	Consonant	Proportion V-initial (liaison) response	Proportion C-initial response	Value of p	Consonant	Proportion V-initial (liaison) response	Proportion C-initial response	Value of p
Experiment 2 (real-word stimuli)	/n/	50.58 %	49.42 %	NS	/n/	56.56 %	43.44 %	P=.0550
	/t/	43.98	56.02	P=.0047	/t/	40.99	59.01	P=.0204
	/z/	45.37	54.63	NS	/z/	50.73	49.27	NS
Experiment 2.1 (non-word stimuli)	/n/	54.62	45.38	NS	/n/	56.94	43.06	P=.0035
	/t/	42.34	57.66	P=.0243	/t/	24.32	75.68	P=.0001
	/z/	53.45	46.55	NS	/z/	59.72	40.28	P=.0019

7.5.1.1.1 Response Bias: /n/

As can be seen above, performance on the segment /n/ showed a bias only for the two NNS groups; the NNS group tested on real-word stimuli showed a bias for V-initial responses that just missed significance, while the NNS group tested on non-word stimuli exhibited a highly significant preference for the V-initial member of each stimulus pair. The NS group tested on non-word stimuli showed a similar bias, but this bias did not reach statistical significance due to the large amount of variation among participants. The NS group tested on real-word stimuli showed an equal proportion of V-initial and C-initial responses.

V-initial response biases for this particular segment on the part of NNS speakers could be the result of an overgeneralization of L2 liaison processing strategies as proposed by Stridfeldt (2003). This is in also line with the processing strategy proposed by Nguyen et al. (2003) discussed above. Add to this the fact that /n/ is the second most frequent liaison consonant (after /z/; see Table 5-1 above) and a V-initial preference could presumably prove to be an efficient processing tool for learners of French.

The fact that NS speakers showed a similar, though statistically insignificant, bias in the same direction uniquely for non-word stimuli supports this hypothesis. When confronted with unfamiliar lexical items, as in the case of non-words employed in Experiment 2.1, the NS speakers may have adopted a similar processing strategy assuming that words following *un* are more likely to begin with a vowel than with /n/. In the case of real lexical items, as employed in Experiment 2, no such V-initial bias was observed for NS participants, suggesting that higher level information associated with each lexical item mitigated this strategy.

7.5.1.1.2 Response Bias: /t/

The plosive /t/ was the only segment to elicit a consistent bias in all four participant groups; this segment elicited significantly more C-initial responses across all participants. Obviously, processing strategies discussed above that resulted in a V-initial bias for the liaison consonant /n/ were not employed in the processing of /t/. Two alternative possibilities underlying a C-initial response bias for /t/ are considered. The first is signal-based, while the second is based on underlying representations of liaison /t/.

Recall that segmental durations of /t/ were taken from the closure portion of this segment only (see §5.4 above). We therefore explore the possibility that acoustic characteristics of the plosive /t/ other than closure duration might be guiding participant responses. As noted above in §5.3.2, VOT has shown variation between LCs and ICs. Dejean de la Bâtie (2003) and Wauquier-Gravelines (1996) both found that VOTs were longer in ICs than in LCs. In the current production sample, however, no significant difference was found, though a difference that just fell short of significance was observed in the opposite direction; VOTs were longer in LCs than in ICs in real-word stimuli. The difference between VOTs in LCs and ICs in non-word stimuli did not approach significance. Given conflicting results between the current production sample and previous production samples, as well as the lack of a consistent difference between real- and non-word tokens in the current production sample, it is difficult to make predictions as to how VOT could affect the perception of possible liaison environments.

VOT values for the four real-word items and the four non-word items included in the perceptual tasks are given below in Table 7-11. Recall that mean VOT values in the current production sample (see Table 5-11 above) represented an average value across 141 tokens of /t/, while only *four* real-word /t/ tokens and *four* non-word /t/ tokens were selected for instrumental manipulation and employed in the behavioral tasks. In other words, the VOT values below are measured from one particular token from one out of six speakers and do not represent mean values. Furthermore, the manipulation of /t/ stimuli in the current study entailed only the manipulation of the closure; VOT remained unaltered across the three continuum conditions. Therefore, the VOT values of the eight stimuli employed in the behavioral tasks do not represent a naturalistic distribution.

Note as well that the particular speaker chosen as the source of stimuli to be used in the perceptual tasks (Speaker Three) had the lowest mean VOT values of all the six speakers (see Figure 5-30 above) and therefore the VOTs of the individual tokens used in the perceptual tasks are lower than attested VOTs in spoken French. The literature reports VOT in French to be an average of 25-30 ms (see for example Caramazza & Yeni-Komshian 1974). In addition, VOT values for the four real-word items are considerably lower than VOTs for non-word items, possibly pointing to hyperarticulation on the part of this particular speaker in the case of unfamiliar lexical items, however this observation is purely conjectural.

Table 7-11: *Voice onset times in milliseconds of /t/ in real- and non-word stimuli used in behavioral tasks.*

Real-word stimuli (Experiment 2)		Non-word stimuli (Experiment 2.1)	
	VOT		VOT
<i>test</i>	18.51	<i>tupe</i>	67.63
<i>tact</i>	17.46	<i>tade</i>	19.75
<i>tamis</i>	17.97	<i>tauvis</i>	30.35
<i>tasseau</i>	18.85	<i>tépeu</i>	31.34

Correlations were conducted to ascertain whether VOT predicted the proportion of C-initial responses for each /t/-initial stimulus for each participant group, however any generalizations drawn from these analyses must be approached with caution given the extremely small number of data points. No significant correlation for the NS group in either a simple regression ($r = .193$, NS) or a Spearman rho analysis ($\rho = .500$, NS) was observed. An analysis of NNS responses also failed to exhibit a significant correlation between VOT and the proportion of C-initial responses in either a simple regression ($r = .505$, NS) or a Spearman rho analysis ($\rho = .429$, NS).

The small number of data points and the fact that the VOT values for the real lexical items do not represent values usually attested in spoken French make conclusions drawn from these analyses difficult, however these data combined with mixed evidence from production studies concerning VOT in environments of liaison would lead us to conclude that VOT does not represent a reliable cue to the presence or absence of liaison in spoken French. Further supporting this conclusion, Fougeron (2001) and Keating, Cho, Fougeron and Hsu (2003) found no consistent differences in VOT at varying prosodic levels (word-initial, syllable-initial, phrase-initial) in spoken French, suggesting that VOT does not represent a consistent cue to the localization of boundaries in French, however further research would be needed to substantiate this claim.

We explore a second possibility underlying the C-initial bias observed for the segment /t/, which is based on underlying representations of this particular segment in the word *grand*. In the stimuli employed in the current behavioral tasks, liaison realized with /t/ differs from liaison realized with /n/ and /z/ in that /t/ in *grand* ‘big’ is not the underlying form. Liaison with *grand* is ostensibly realized with /t/ (though see below), while *grand* in its feminine form, *grande* [gʁãd], is produced with /d/. Furthermore, /d/ surfaces in lexical items derived from *grand* as in *grandeur* [gʁã.dœʁ] ‘greatness, largeness’ and *grandir* [gʁã.diʁ] ‘to grow, to grow up’, pointing to /d/ as the underlying representation of this segment. One could argue that /t/ realized in liaison with *grand* is a learned epenthetic liaison consonant and does not actually represent the underlying form, therefore biasing perception toward a fixed /t/-initial interpretation.

The liaison consonant in the determiner *un* on the other hand has the same realization in liaison (/n/) as in its underlying form; *un* becomes *une* [yn] in the feminine form and also has derived forms such as *unité* [y.ni.te] ‘unit, unity’ and *unicité* [y.ni.si.te] ‘uniqueness’. Therefore, the existence of alternate underlying forms for liaison /t/ could

plausibly render mental representations of liaison environments with *grand* unstable relative to the other two segments under investigation here, consequently biasing /t/ responses toward a C-initial interpretation in that representations of word-initial /t/ are stable and fixed.

Supporting this hypothesis are data from recent corpus work on spoken French (Durand & Lyche 2008) which show that, in a reading task, the phrases *grand émoi* ‘great emotion’ and *grand honneur* ‘great honor’ were produced without liaison by 6 out of 100 speakers, even though this is considered to be a case of ‘obligatory’ liaison (pre-posed adjective + noun). Moreover, two speakers produced a liaison with /d/ instead of /t/ in productions of *grand émoi*.²⁴

Confirmation of a perceptual strategy based on the underlying forms of liaison consonants would require further research. One possible means to validate this hypothesis would be to compare the perception of /t/ in liaison environments with *grand* with liaison environments with *petit* /pə.ti/ ‘small, little’, as in the phrase *petit âne* [pə.ti.tan] ‘little donkey’. The underlying form of /t/ in *petit* is unquestionably /t/ as seen in the feminine form *petite* [pə.tit] and the derived form *petitesse* [pə.ti.tɛs] ‘smallness’. If mental representations of /t/ are more stable in *petit* than in *grand*, we would not expect a C-initial bias in ambiguous liaison environments including *petit*.

On this logic, we would also expect underlying mental representations of liaison /z/ in *les*, which has no alternate form, to differ from, for example, underlying mental representations of /z/ in the adjective *gros* [gʁo] ‘big, fat’. The segment /z/ surfaces in liaison, for example, in the phrase *gros ours* [gʁo.zuʁs] ‘big bear’, however *gros* has a second underlying form, /s/, found in the feminine form *grosse* [gʁɔs] and the derived form *grossesse* [gʁɔ.sɛs] ‘pregnancy’.

7.5.1.1.3 Response Bias: /z/

Regarding the segment /z/, which appeared exclusively in the determiner *les* in the current study, responses were varied. The NS group tested on real-word stimuli in Experiment 2 exhibited a small, but statistically insignificant, bias for C-initial items (54.63% of responses); the NNS group tested on real-word stimuli also showed no significant preference. Regarding the perception of non-word stimuli, the NS group tested in Experiment 2.1 showed a small, but insignificant, bias for V-initial items (53.45% of responses). The NNS group tested on non-word stimuli showed a significant bias in the same direction as the NS (V-initial bias = 59.72% of responses; $p = .0019$).

For non-word stimuli, it appears that both the NS and NNS groups have adopted the overgeneralization strategy discussed above leading listeners to assume that, when confronted with an unknown word in a possible liaison environment, the unknown item is more likely V-initial than C-initial.

²⁴ Recall as well that in the production sample of the current study the four tokens that were removed from acoustic analysis were all tokens of /t/ in a liaison environment. These tokens were removed because speakers either inserted a pause or a glottal stop instead of realizing /t/.

7.5.1.1.4 Discussion: Response bias by consonant

Taken as a whole, the analysis of individual segments is mixed — no clear global preference for either V-initial or C-initial responses emerged across all three consonants, however individual segments showed biases in both directions. This result is somewhat surprising given the fact that global response patterns in combined analyses of all three segments showed the predicted linear effect of durational variation, namely shortened consonants elicited significantly more V-initial responses and lengthened consonants elicited more C-initial responses. Nonetheless, the factors discussed above may have mitigated the robustness of the effect of segmental variation and most likely contributed to the degree of variation observed across participants. As discussed above and throughout this dissertation, multiple, and potentially competing, processing strategies are likely employed in the parsing of ambiguous sequences.

7.5.1.2 Response Bias by Item: Real-word stimuli

We now turn to an examination of response proportions within each individual response pair in an attempt to shed additional light on observed response biases. In order to examine whether participants showed a bias for one particular member of each individual minimal pair, response proportions in the forced-choice identification tasks for both Experiment 2 and Experiment 2.1 were calculated for each individual member of each ambiguous pair. Response proportions for real-word stimuli are summarized below in Table 7-12. The table also presents the results of paired t-tests analyzing whether response proportions within each stimulus pair differ significantly. Cells shaded in grey indicate where the NS group and NNS group were consistent in the directionality of their responses and where differences in response proportions were significant for both groups.

Response proportions for real-word stimuli for the NS group showed significant biases for 8 out of 12 stimulus pairs, while the NNS group showed biases for 6 out of 12 pairs. Following the results of Shoemaker and Birdsong (2008), in which NS and NNS participants' response were correlated on per-item biases, we also investigated whether biases per stimulus pair were correlated between the NS and the NNS groups in the current study. Response proportions for each stimulus pair were compared for the NS and NNS groups tested on real words, which revealed a significant correlation in both a simple regression ($r=.709$, $p=.0001$) and a Spearman rank order analysis ($\rho=.653$, $p=.0017$). This correlation further attests to the fact that NNS participants are exploiting the same information in the acoustic signal as NS participants.

Table 7-12: *Relative proportions of responses in identification task for real-word stimulus pairs used in Experiment 2. P values in parentheses indicate values that fall just short of significance. Cells shaded in grey indicate where the NS and NNS participants were consistent in the directionality of their responses.*

Native Speakers					Non-Native Speakers				
Proportion V-initial (liaison) Responses	Proportion C-initial Responses	Value of P in Paired t-test	Proportion V-initial (liaison) Responses	Proportion C-initial Responses	Value of P in Paired t-test	Proportion V-initial (liaison) Responses	Proportion C-initial Responses	Value of P in Paired t-test	
un air	39.51	un nerf	60.49	p = .0380	un air	46.30	un nerf	53.70	NS
un oeuf	61.42	un neuf	38.58	p = .0015	un oeuf	67.28	un neuf	31.48	p = .0040
un hectare	42.59	un nectar	57.41	p = .0070	un hectare	53.70	un nectar	46.30	NS
un aval	58.02	un naval	41.67	(p = .0735)	un aval	58.95	un naval	41.05	(p = .0699)
le grand Est	40.74	le grand test	59.26	p = .0064	le grand Est	33.95	le grand test	66.05	p = .0171
un grand acte	52.78	un grand tact	47.22	NS	un grand acte	45.68	un grand tact	54.32	NS
un grand ami	50.93	un grand tamis	49.07	NS	un grand ami	52.78	un grand tamis	45.37	NS
un grand assaut	28.40	un grand tasseau	71.60	p < .0001	un grand assaut	22.84	un grand tasseau	77.16	p < .0001
les ailes	61.42	les zèles	38.58	(p = .0704)	les ailes	61.73	les zèles	38.27	p = .0267
les aines	48.46	les Zens	51.54	NS	les aines	51.54	les Zens	48.46	NS
les ailés	30.86	les zélés	69.14	p = .0090	les ailés	54.94	les zélés	44.75	NS
les aunages	41.36	les zonages	58.64	NS	les aunages	34.57	les zonages	65.43	p = .0280
Mean Proportion V-initial responses	46.37%	Mean Proportion C-initial responses	53.63%	p = .0046	Mean Proportion V-initial responses	49.43%	Mean Proportion C-initial responses	50.57%	NS

7.5.1.2.1 Lexical Frequency

In the current section, we explore whether the relative lexical frequency of real-word items within each stimulus pair in Experiment 2 influenced participants' responses. As discussed in Chapter One, the frequency of a word can directly affect word recognition processes. Lexical items with higher frequency are recognized more readily than less frequent items (see for example Cluff & Luce 1990; Luce & Pisoni 1998). Specifically in the case of liaison in spoken French, Bybee's (2005) exemplar model proposes that mental representations of liaison are based largely on frequency of occurrence. Bybee holds that words that appear together frequently in liaison environments are encoded as lexical chunks, while words that appear less frequently

together are processed according to a sort of liaison *template* (see §2.3 above). The possibility that relative lexical frequency between the two items in each real-word stimulus pair affected responses was therefore also considered.

Lexical frequencies for each individual real-word target according the LEXIQUE database are reproduced from Chapter Five in Table 7-13 below. Recall that, in the stimulus set employed in the current study, the mean lexical frequency of V-initial items (119.34 occurrences per million) is roughly eight times the mean frequency of C-initial items (14.25 occurrences per million).

Table 7-13: *Frequency of occurrence (per million words) of real lexical items used in Experiment 2 according to the LEXIQUE database.*

Vowel-initial (liaison) item		Consonant-initial item	
air	588.00	nerf	30.30
oeuf	44.77	neuf	86.32
hectare	4.65	nectar	1.55
aval	3.04	naval	6.58
Est	65.74	test	30.50
acte	57.06	tact	3.74
ami	551.05	tamis	1.18
assaut	22.14	tasseau	0.01
ailles	46.96	zèles	7.80
aines	1.71	Zens	0.87
ailés	2.12	zélés	2.08
aunages	0.00	zonages	0.05
Mean Frequency	119.34	Mean Frequency	14.25

Correlations were conducted between the above frequency values and the proportion of responses for each stimulus pair. NS responses failed to show a correlation with frequency in both a simple regression ($r = -.147$, NS) or in a Spearman analysis ($\rho = -.203$, NS). Note as well that both of these correlation coefficients are in fact weakly negative, while an effect of frequency would be expected to produce a positive correlation coefficient with proportion of responses. NNS responses also failed to show a significant correlation with frequency in a simple regression ($r = -.046$, NS) and in a Spearman analysis ($\rho = -.149$, NS), both of which also showed weak negative correlations.

We consider one further possibility that is linked to lexical frequency and that could guide participant responses, which is the frequency of *co-occurrence* of the two words linked in liaison environments. Bybee (2001b) differentiates among two measures of frequency: *token frequency*, which refers to the frequency of a single word and which we have given above in Table 7-13, and *string frequency*, which refers to the frequency with which a sequence of words appears together. The string frequency for each W_1W_2

combination was also calculated from the *Lexique* database and is given below in Table 7-14. The values in the table represent the number of occurrences of the sequence W_1W_2 per million words. Due to limitations of search options in the LEXIQUE database, the string frequency values below are based only on film subtitles (50.4 million words), whereas token frequencies given above were based on both film subtitles as well as printed works (14.7 million words). As can be seen, the string frequency of 11 out of 24 sequences is zero according to this particular corpus. Note as well that all sequences including a /z/-initial W_2 had string frequencies of zero.

Table 7-14: *String frequency per million words of lexical items used in Experiment 2 according to LEXIQUE database.*

Vowel-initial W_2	String Frequency of W_1W_2	Consonant-initial W_2	String Frequency of W_1W_2
un air	12.58	un nerf	0.60
un oeuf	6.69	un neuf	1.63
un hectare	0.23	un nectar	0.08
un aval	0.04	un naval	0
grand Est	0	grand test	0.05
grand acte	0.03	grand tact	0.01
grand ami	2.01	grand tamis	0
grand assaut	0.02	grand tasseau	0
les ailes	5.38	les zèles	0
les aines	0	les Zens	0
les ailés	0	les zélés	0
les aunages	0	les zonages	0

Given the proportions of data points at zero, correlations are extremely unlikely to be significant, however the possibility of a correlation between string frequency and response proportions was nonetheless considered. NS responses failed to show a significant correlation with string frequency in both a simple regression ($r = -.006$, NS) and in a Spearman analysis ($\rho = .194$, NS). NNS responses also failed to show a significant correlation with frequency in a simple regression ($r = .132$, NS) and in a Spearman analysis ($\rho = .277$, NS).

In sum, frequency has been shown to play a significant role in lexical access, however, the current analyses suggest that neither the frequency of W_2 nor the frequency with which W_1 and W_2 occur together in spoken French had a significant effect on participant responses in Experiment 2. The more frequent lexical item of each stimulus pair did not systematically affect the distribution of responses for either participant group.

7.5.1.3 Response Bias by Item: Non-word stimuli

Response proportions for non-word stimuli pairs are given below in Table 7-15, along with the results of paired t-tests analyzing whether response proportions within each stimulus pair differ significantly. Response proportions for non-word stimuli for the

NS group showed significant biases for only 3 out of 12 stimulus pairs, while the NNS group showed significant biases for 8 out of 12 pairs. Proportion of responses for each non-word stimulus pair were compared for the two groups, which revealed no significant correlation in a simple regression ($r=.310$, NS), and a Spearman rank correlation that fell just short of significance ($\rho=.389$, $p=.0620$).

Table 7-15: *Relative proportion of responses in identification task for non-word stimulus pairs used in Experiment 2.1. P values in parentheses indicate values that fall just short of significance.*

Native Speakers					Non-Native Speakers				
Proportion V-initial (liaison) Responses	Proportion C-initial Responses	Value of P in Paired t-test	Proportion V-initial (liaison) Responses	Proportion C-initial Responses	Value of P in Paired t-test	Proportion V-initial (liaison) Responses	Proportion C-initial Responses	Value of P in Paired t-test	
un auvis	53.15	un nauvis	46.85	NS	un auvis	69.93	un nauvis	30.07	p = .0002
un épeu	51.40	un népeu	48.60	NS	un épeu	64.71	un népeu	35.29	p = .0054
un upe	64.24	un nupe	35.76	p = .0264	un upe	50.33	un nupe	49.67	NS
un ade	39.51	un nade	60.49	p = .0489	un ade	54.90	un nade	45.10	NS
un grand auvis	44.10	un grand tauvis	55.90	NS	un grand auvis	35.40	un grand tauvis	64.60	p = .0220
un grand épeu	33.10	un grand tépeu	66.90	p = .0086	un grand épeu	47.38	un grand tépeu	52.62	NS
un grand upe	40.97	un grand tupe	59.03	NS	un grand upe	15.08	un grand tupe	84.92	p > .0001
un grand ade	50.69	un grand tade	49.31	NS	un grand ade	22.88	un grand tade	77.12	p > .0001
les auvis	52.10	les zauvis	47.90	NS	les auvis	71.57	les zauvis	28.43	p = .0004
les épeus	52.45	les zépeus	47.55	NS	les épeus	71.89	les zépeus	28.11	p = .0007
les upes	43.01	les zupes	56.99	NS	les upes	50.66	les zupes	49.34	NS
les ades	47.90	les zades	52.10	NS	les ades	58.82	les zades	41.18	(p = .0770)
Mean Proportion V-initial responses	50.14%	Mean Proportion C-initial responses	49.86%	NS	Mean Proportion V-initial responses	46.99%	Mean Proportion C-initial responses	53.01%	NS

The lack of a strong correlation between NS and NNS responses to individual non-word stimuli pairs suggests that when higher level lexical information is lacking in the signal, differing strategies on the part of each participant group are adopted for the parsing of ambiguous sequences. The correlation found between NS and NNS responses

for real-word stimuli would suggest that both groups are making use of the same information when higher level information is available.

7.5.2 Additional Acoustic Factors

Previous studies have suggested that acoustic factors other than the duration of pivotal consonants can vary at and around word boundaries in environments of liaison (e.g. Dejean de la Bâtie 1993; Spinelli et al. 2003). The possibility that participants' responses were influenced by acoustic factors other than the segmental duration of pivotal consonants was therefore also considered in the present study. However, it is important to note that any generalizations drawn from an analysis of behavioral data based on additional acoustic factors should be taken with caution as the acoustic information in the manipulated stimuli used in the behavioral tasks does not reflect the same distribution of values as the production sample described in Chapter Five.

Recall that the intention of the behavioral tasks was to test the exploitation of the segmental duration of pivotal consonants by enhancing this particular acoustic parameter while keeping all other acoustic factors unchanged. For example, in the three manipulated versions of *un grand tamis* presented to participants in the behavioral tasks, only the closure duration of /t/ was manipulated while, for example, vocalic durations, voice onset time of pivotal /t/, formant transitions, etc. remained unchanged across all three manipulated tokens of the durational continuum. Nonetheless, the possibility that participant responses were influenced by acoustic information other than the duration of pivotal consonants in the forced-choice identification tasks employed in Experiments 2 and 2.1 was explored.

7.5.2.1 Vowel Duration

As discussed in §5.3.1, the production sample in Spinelli et al. (2003) showed that pre-boundary vowels preceding LCs were shorter by 3% than pre-boundary vowels preceding ICs, though no differences were found for post-boundary vowels in that study. Nguyen et al. (2007), however, found no significant durational differences between vowels preceding pivotal consonants in liaison environments versus non-liaison environments. Nguyen et al. did not report measurements for post-boundary vowels.

As discussed in §5.3.1.3 the current production sample is in line with the results of Spinelli et al. (2003). Pre-boundary vowels (V_1) produced before resyllabified LCs (mean: 88.18 ms) in our production sample were 3.82% shorter than pre-boundary vowels appearing before ICs (mean: 91.68 ms), a difference which was statistically significant ($p = .0012$). Also consistent with the results of Spinelli et al. (2003), our production sample revealed no significant differences in post-boundary vowels. For this reason, only V_1 durations are compared against behavioral data here.

If V_1 durations are guiding participants' responses in the identification task, longer vowels should elicit more C-initial responses. In other words, a positive correlation between V_1 duration and proportion of C-initial responses should be observed. To test this possibility, we performed correlations between V_1 duration and the proportion of C-initial responses for each individual stimulus. We first looked at responses from Experiment 2 (real-word stimuli) for both participant groups. For the NS

group, no significant correlation was found between V_1 duration and the proportion of C-initial responses for real-word items in a simple regression ($r = .168$, NS) or a Spearman analysis ($\rho = .046$, NS). An analysis of NNS responses also failed to show a significant correlation between V_1 duration and the proportion of C-initial responses for real-word items (simple regression, $r = -.056$, NS; Spearman analysis ($\rho = -.056$, NS).

Responses from Experiment 2.1 (non-word stimuli) were then examined. An analysis of NS responses failed to show a significant correlation between V_1 duration and the proportion of C-initial responses for non-word stimuli (simple regression, $r = .253$, NS; Spearman analysis, $\rho = 0$, NS). Responses for the NNS group also showed no significant correlation between V_1 duration and the proportion of C-initial responses for non-word stimuli (simple regression, $r = -.214$, NS; Spearman analysis, $\rho = -.238$, NS).

These results suggest that, though the durations of vowels preceding pivotal consonants may systematically vary as a function of the presence or absence of liaison, this variation did not affect participants' responses in the forced-choice identification task. A lack of correlation with participant responses coupled with the relatively small amount of variation ($\approx 3\%$) found between V_1 s in liaison environments and V_1 s in non-liaison environments would lead us to conclude that the duration of vowels preceding pivotal consonants does not represent a robust cue to liaison environments. However, further research in which the duration of pre-boundary vowels is manipulated while all other acoustic information is held constant would be needed to validate this conclusion.

7.5.2.2 Vowel/Consonant Ratio

We also wanted to consider the possibility that durational cues to liaison could be relational, i.e. that perceptual cues are governed by a proportional relationship between the durations of more than one segment. Given that the durations of both V_1 and pivotal consonants have been shown to vary according to the presence or absence of liaison, the proportional relationship of these two segments was also examined. Vowel/consonant (VC) ratios were calculated by dividing the duration of V_1 by the duration of ICs and LCs. VC ratios for the production current sample showed that $V_1 + LC$ ratios were 15% higher than $V_1 + IC$ ratios. This finding is in line with production measurements reported by Spinelli et al. (2003), which also showed an increase of roughly 15%. VC ratios for the current production sample as well as for the production sample from Spinelli et al. (2003) are given below in Table 7-16.

In order to investigate whether VC ratios influenced participant responses in the forced-choice identification tasks in Experiments 2 and Experiments 2.1, ratios were calculated for each stimulus item in each of the three continuum conditions and are given below in Table 7-17. Mean durations and mean VC ratios, as well as standard deviations, are given at the bottom of the table. It is worth noting, however, that vocalic durations were held constant in the instrumental manipulation process and therefore VC ratios in manipulated stimuli may not represent values that would be encountered in natural speech.

Table 7-16: Mean durations in milliseconds of vowels preceding pivotal consonants (V_1), pivotal consonants (C), and vowel/consonant (VC) ratio in liaison and non-liaison environments in the current production sample and in the production sample of Spinelli et al. (2003).

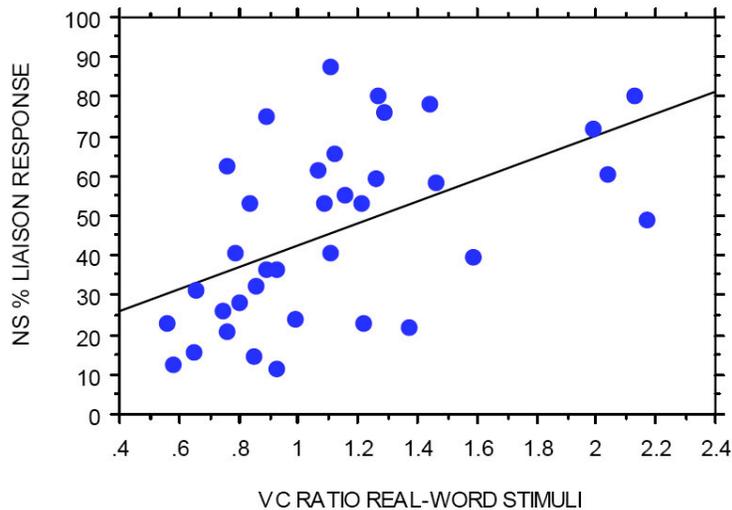
	Current Production Sample			Spinelli et al. (2003) Production Sample		
	V_1	C	VC ratio	V_1	C	VC ratio
Liaison Environment (LC)	88.18	84.40	1.04	97	59	1.64
Non-liaison Environment (IC)	91.68	100.63	0.91	100	71	1.41

Table 7-17: Mean durations in milliseconds of vowels preceding pivotal consonants (V_1), pivotal consonants (C), and vowel/consonant (VC) ratio for each stimulus item in each of three continuum conditions.

	Shortened			Baseline			Lengthened		
	V_1	C	VC ratio	V_1	C	VC ratio	V_1	C	VC ratio
un nerf	84.80	75.53	1.12	84.80	107.78	0.79	84.80	145.23	0.58
un neuf	95.60	75.53	1.27	95.60	107.78	0.89	95.60	145.23	0.66
un nectar	83.35	64.83	1.29	83.35	94.04	0.89	83.35	129.20	0.65
un naval	71.88	64.83	1.11	71.88	94.04	0.76	71.88	129.20	0.56
le grand test	94.01	46.09	2.04	94.01	84.71	1.11	94.01	123.32	0.76
un grand tact	98.26	46.09	2.13	98.26	84.71	1.16	98.26	123.32	0.80
un grand tamis	89.40	44.88	1.99	89.40	71.08	1.26	89.40	105.14	0.85
un grand tasseau	97.29	44.88	2.17	97.29	71.08	1.37	97.29	105.14	0.93
les zèles	107.88	75.11	1.44	107.88	100.86	1.07	107.88	128.11	0.84
les Zens	110.00	75.11	1.46	110.00	100.86	1.09	110.00	128.11	0.86
les zélés	115.63	72.82	1.59	115.63	94.97	1.22	115.63	117.18	0.99
les zonages	87.97	72.82	1.21	87.97	94.97	0.93	87.97	117.18	0.75
un nupe	95.26	75.53	1.26	95.26	107.78	0.88	95.26	145.23	0.66
un nade	91.85	75.53	1.22	91.85	107.78	0.85	91.85	145.23	0.63
un nauvis	77.49	64.83	1.20	77.49	94.04	0.82	77.49	129.20	0.60
un népeu	75.71	64.83	1.17	75.71	94.04	0.81	75.71	129.20	0.59
un grand tupe	107.83	46.09	2.34	107.83	84.71	1.27	107.83	123.32	0.87
un grand tade	104.71	46.09	2.27	104.71	84.71	1.24	104.71	123.32	0.85
un grand tauvis	90.00	44.88	2.01	90.00	71.08	1.27	90.00	105.14	0.86
un grand tépeu	83.35	44.88	1.86	83.35	71.08	1.17	83.35	105.14	0.79
les zupes	111.03	75.11	1.48	111.03	100.86	1.10	111.03	128.11	0.87
les zades	107.19	75.11	1.43	107.19	100.86	1.06	107.19	128.11	0.84
les zauvis	123.29	72.82	1.69	123.29	94.97	1.30	123.29	117.18	1.05
les zépeus	83.30	72.82	1.14	83.30	94.97	0.88	83.30	117.18	0.71
MEAN	95.29	63.21	1.58	95.29	92.24	1.05	95.29	124.70	0.77
(SD)	(13.41)	(13.30)	(0.41)	(13.41)	(12.02)	(0.19)	(13.41)	(12.47)	(0.13)

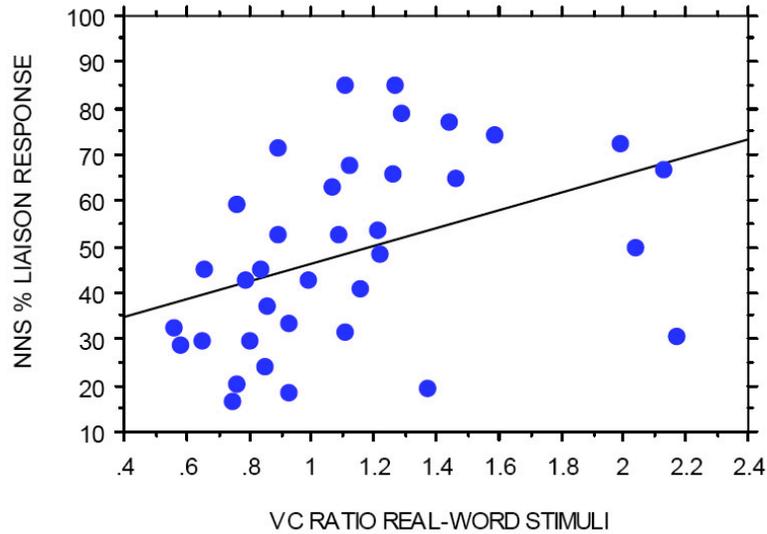
If VC ratios represent a reliable cue to environments of liaison, higher VC ratios should elicit a higher proportion of ‘liaison’ (V-initial) responses. In order to investigate this possibility, VC ratios for real-word stimuli were compared with proportion ‘liaison’ responses for the NS group, which correlated significantly in a Spearman analysis ($\rho = .555$, $p = .0011$), as well as in a simple regression ($r = .523$, $p = .0011$). A regression plot for this analysis is shown below in Figure 7-18.

Figure 7-18: Regression plot showing NS participants’ proportions of ‘liaison’ responses in forced-choice identification task employing real-word stimuli as a function of vowel/consonant ratio.



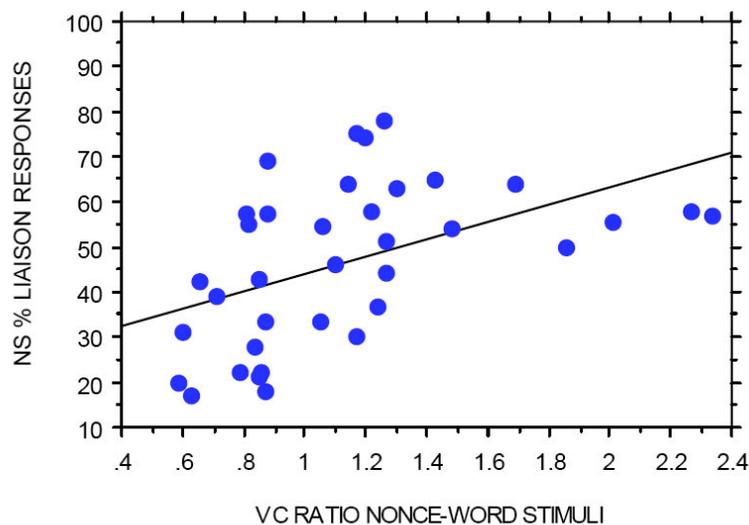
VC ratios for real-word stimuli were subsequently compared with proportion ‘liaison’ responses for the NNS group, which also revealed a significant correlation in a Spearman analysis ($\rho = .526$, $p = .0019$), as well as in a simple regression ($r = .404$, $p = .0145$). A regression plot showing NNS responses in relation to VC ratio is shown below in Figure 7-19.

Figure 7-19: Regression plot showing NNS participants' proportions of 'liaison' responses in forced-choice identification task employing real-word stimuli as a function of vowel/consonant ratio.



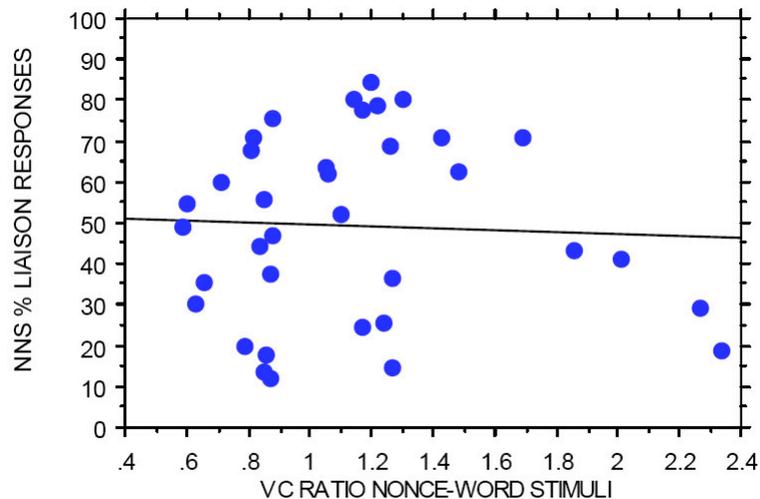
We then examined VC ratios for non-word stimuli. The proportion of 'liaison' responses for the NS group also showed a significant correlation with VC ratios in a Spearman analysis ($\rho = .585$, $p = .0005$), and in a simple regression ($r = .483$, $p = .0029$). A regression plot for this analysis is shown below in Figure 7-20.

Figure 7-20: Regression plot showing NS participants' proportions of 'liaison' responses in forced-choice identification task employing non-word stimuli as a function of vowel/consonant ratio.



Finally, VC ratios for non-word stimuli were compared to the proportion of ‘liaison’ responses for the NNS group, which failed to demonstrate a significant correlation in a Spearman analysis ($\rho = .106$, NS), or in simple regression ($r = -.043$, NS), where the correlation coefficient was in fact weakly negative. Figure 7-21 below shows a regression plot of NNS responses as a function of VC ratio.

Figure 7-21: *Regression plot showing NNS participants’ proportions of ‘liaison’ responses in forced-choice identification task employing non-word stimuli as a function of vowel/consonant ratio.*



NS listeners appear to be sensitive to the relationship between the vowel and pivotal consonant in liaison environments in spoken French. Both NS groups showed an increase in the proportion of ‘liaison’ responses as VC ratio increased. Results for the two NNS groups are mixed. The NNS group tested on real-word stimuli showed that same linear increase, while the NNS group tested on non-word stimuli did not, suggesting that the NS of French may be more sensitive to this relational cue than L2 learners.

We see two possible explanations for the observed correlations between VC ratio and proportion of ‘liaison’ responses. The first possibility is that the observed correlations are merely an artifact of the instrumental manipulation process rather than evidence of a relational cue that exists in natural speech. Given that the vocalic durations in each target item were kept constant across the three continuum conditions while only the duration of pivotal consonants was altered, these ratios may not represent values that would be found in natural speech as noted above. As the durations of pivotal consonants increased in the manipulation process, the VC ratio necessarily decreased, which could account for the correlation with ‘liaison’ response proportions.

The second possibility is that VC ratio does indeed signal the presence or absence of liaison in natural speech. To our knowledge, no work to date has explored the possibility of liaison being signaled by relational rather than absolute duration. Further

research in which vowel durations are manipulated along with consonant durations would be needed in order to corroborate whether VC ratio is in fact a reliable cue to liaison.

7.5.3 Discussion: Post-hoc Analyses

The results of the above post-hoc analyses of participant responses as a function of acoustic factors other than the segmental duration of pivotal consonants suggest that, although additional information may be available in the speech signal in instances of liaison, the segmental duration of pivotal consonants was guiding participants' responses in the identification of ambiguous phonemic sequences in these identification tasks. As noted above, further research would be required to substantiate the presence of additional acoustic cues to the presence or absence of liaison.

7.6 General Discussion: Forced-Choice Identification Task

The purpose of Experiments 2 and 2.1 was to evaluate the degree to which the duration of pivotal consonants in globally ambiguous phrases in spoken French can influence the lexical interpretation of these phrases when all other acoustic factors in the signal are held constant. The forced-choice identification tasks in Experiments 2 and 2.1 employed stimuli from the three-step durational continuum described in Chapter Five.

As predicted, stimuli in which the pivotal consonants were instrumentally shortened consistently elicited a significantly larger proportion of V-initial (liaison) responses, while stimuli in which the pivotal consonants were instrumentally lengthened consistently elicited a significantly larger proportion of C-initial responses. Crucially, the baseline stimuli elicited an equal number of V-initial and C-initial responses globally, suggesting that when insufficient acoustic information is available in the signal and context is lacking, a guessing strategy is adopted.

Taken as a whole, the pattern of data from both experiments supports the hypothesis that segmental duration can modulate the lexical interpretation of ambiguous liaison sequences in spoken French. Unlike previous studies, which have hypothesized that duration serves as a cue to disambiguation in environments of liaison, but have not tested this cue directly (Gaskell et al. 2002; Spinelli et al. 2003), the current study has demonstrated an effect of variation in segmental duration by manipulating this factor while all other acoustic factors remain unchanged.

These results are consistent with previous findings which have demonstrated that listeners use segmental duration to divide the speech stream into lexical units in English (Davis et al. 2002; Salverda et al. 2003), Dutch (Quené 1992; Shatzman & McQueen 2006), Italian (Tabossi et al. 2000), and French (Banel & Bacri 1994). The current results are also compatible with a large body of work which has shown that mismatches in sub-phonemic detail can perturb word recognition processes (Andruski, Blumstein & Buton 1994; Connine, Blasko & Wang 1994; Connine, Titone, Deelman & Blasko 1997; Frauenfelder, Scholten & Content 2001; Marslen-Wilson, Moss & van Halen 1996; Soto-Faraco, Marslen-Wilson & Warren 1994; Sebastián-Gallés & Cutler 2001).

Also as predicted, we observed a large degree of variation among participants, both NS and NNS, suggesting that, though these durational differences are systematically present in the speech stream and are encoded in a listener's phonological grammar, not all listeners exploit this cue to the same extent and that there may be other as yet

unidentified cues (acoustic or higher level) that are exploited more reliably. Indeed, a close inspection of response patterns revealed significant response biases within stimulus pairs on the part of both NS and NNS participants indicating that, even when durational differences are exaggerated as in the current stimulus sample, other cues may be influencing responses. Post-hoc analyses examining additional acoustic and higher-level information were conducted in an attempt to isolate additional strategies used by participants in the processing of ambiguous phonemic content. Hypotheses were made as to possible (competing) processing strategies underlying response biases, but no single factor emerged as a determinant in response patterns.

Regarding the performance of the NNS participants, the current results are in line with recent research offering evidence that late language learners can acquire sensitivity to non-contrastive allophonic variation in the L2 (e.g. Darcy et al. 2007). Specifically, our results suggest that highly advanced learners of L2 French can develop sensitivity to allophonic durational variation in environments of liaison in spoken French. It is also important to point out that, though we were not able to pinpoint factors other than segmental duration that may underlie response biases within stimulus pairs, NS and NNS response biases within each stimulus pair were strongly correlated for real-word stimuli, indicating that these two groups are making use of the same information in the signal, be it acoustic or otherwise, in determining responses.

We again point out that these durational differences as they occur in natural speech may not represent a consistently robust processing cue. However, crucially, the fact that listeners by and large responded in the predicted direction demonstrates that segmental duration does have cue value in the processing of spoken French. These results offer strong evidence that durational differences between LCs and ICs are indeed encoded in the phonological L1 and L2 grammar. In the next chapter, we discuss the current results and the use of segmental duration as boundary cue within a framework of currently accepted models of both L1 and L2 spoken word recognition.

CHAPTER EIGHT: NNS Biographical Data

8.0 Introduction

In this section, we analyze NNS data from the two behavioral tasks presented in Chapters Six and Seven as a function of biographical and experiential factors. Behavioral data from the NNS group tested on real-word stimuli in Experiments 1 and 2 and the NNS group tested on non-word stimuli in Experiments 1.1 and 2.1 are examined in light of self-reported biographical data in an attempt to identify factors contributing to the upper limits of attainment in the perceptual processing of L2 phonology.

8.1 L2 Experiential Factors in L2 Attainment

As discussed extensively in Chapter Three, much research on the acquisition of L2 phonological structure is centered upon the mediating role of age. The variable of age is operationalized in the majority of the literature as the age of arrival (AOA) in the L2 environment. AOA is taken to mark the beginning of significant immersion in the target language. This variable has proved to be the strongest predictor of performance in a L2 for late-learners (see Birdsong 1999, 2005 for a review).

However, as also discussed in Chapter Three, a large body of work has demonstrated that experiential factors not related to age play a significant role in L2 attainment levels. For example, Piske, Flege and MacKay (2000) point out several non-age related factors that can affect L2 phonological acquisition such as length of residence in the L2 environment and the amount of daily use of both the L2 and the L1. In exploring the effects of daily L2 usage, Flege and Liu (2001) showed that learners of L2 English (L1=Chinese) who had increased contact with native English speakers performed better on tasks including listening comprehension, grammatical judgments, and the perception of L2 segments than learners with lower levels of native speaker contact. Furthermore, Flege, Frieda and Nozawa (1997) offered evidence suggesting that the overall degree of foreign accent in L2 production positively correlated with self-reported amount of L1 use. Education level has also been shown to affect L2 performance in inflectional morphology (Flege, Yeni-Komshian & Liu 1999). Recall as well that both Birdsong (2003, 2007) and Bongaerts (1999) found that instances of nativelike pronunciation in L2 French (L1=English and Dutch, respectively) were linked to phonetic training. Motivational factors have also been proposed as having a significant effect on L2 attainment. Klein (1995) proposes that nativelike behavior on the part of non-native speakers may not be achievable without a desire to assimilate into the culture associated with the L2.

However, though the literature has shown that these experiential factors can affect attainment outcomes, they tend to account for much less of the variance observed in L2 ultimate attainment than does AOA. For example, Birdsong and Paik (2008) maintain

that only 10% of NNS performance variance is attributable to amount of L2 input and use, while 50% can be attributed to age-related effects.

8.2 NNS Biographical and Experiential Data

In an attempt to pinpoint factors, age-related or otherwise, that may underlie NNS participant performance in the current investigation, self-reported biographical and language proficiency information was collected prior to testing. As noted above in §6.2.3, both NNS groups completed an extensive questionnaire adapted from the Language Experience and Proficiency Questionnaire²⁵ (LEAP-Q; Marian et al. 2007; See Appendix II for complete questionnaire).

Recall that several NNS participants were removed from analyses of behavioral results in Chapters Six and Seven due to extreme response biases; these same participants were also excluded from the current analyses of biographical and experiential information. This resulted in the exclusion of one participant from the NNS group tested on real-word stimuli (NNS 5) and two participants from the NNS group tested on non-word stimuli (NNS 21 and NNS 29).

The two NNS groups were closely matched for AOA and proficiency. Average AOA for the NNS participants tested on real-word stimuli (Experiments 1 and 2) was 29.49 years (SD=9.5; range: 18 - 59). Average AOA for the NNS participants tested on non-word stimuli (Experiments 1.1 and 1.2) was 31.09 years (SD=8.5; range: 20 - 47.7). This factor showed no significant difference between the two participant groups ($F(1, 31) = 6.74$, NS).

The two NNS groups were also matched for L2 proficiency. Proficiency was operationalized as the average of self-reported proficiency levels in four domains (speaking, reading, listening and pronunciation) each reported on a scale of 1 – 10. Mean proficiency for the NNS participants tested on real-word stimuli (Experiments 1 and 2) was 7.48 on a 10-point scale (SD=1.65; range: 3 - 9.75). Mean proficiency for the NNS participants tested on non-word stimuli (Experiments 1.1 and 1.2) was 7.26 on a 10-point scale (SD=1.49; range: 4.50 - 9.25). The difference in self-reported proficiency levels between the groups was non-significant ($F(1, 31) = .162$, NS). AOA and proficiency for each NNS participant are given below in Table 8-1.

²⁵ Language proficiency can, of course, also be determined by metrics that are more objective than self-assessment such as standardized language tests. However, the questionnaire used in the current study was chosen because the authors specifically created it in order to provide a questionnaire that offers predictable correlations between self-reported and behavioral data. These authors conducted extensive quantitative studies which showed that the self-reported data on L2 proficiency gathered in this particular questionnaire reliably correlates with standardized measures.

Table 8-1: *Age of arrival (AOA) and global proficiency for NNS participants.*

NNS tested in Experiments 1 & 2 (real-word stimuli)			NNS tested in Experiments 1.1 & 2.1 (non-word stimuli)		
	AOA	Proficiency		AOA	Proficiency
NNS 1	23.3	8.38	NNS 19	23.8	6.75
NNS 2	20.0	8.50	NNS 20	37.3	6.00
NNS 3	29.0	8.25	NNS 21	removed from analysis	
NNS 4	59.0	3.00	NNS 22	39.8	6.50
NNS 5	removed from analysis		NNS 23	26.7	5.25
NNS 6	22.0	9.00	NNS 24	24.0	7.75
NNS 7	35.0	7.88	NNS 25	27.0	9.25
NNS 8	21.0	8.25	NNS 26	26.0	6.75
NNS 9	18.0	7.50	NNS 27	22.4	6.50
NNS 10	36.0	7.50	NNS 28	29.2	8.25
NNS 11	27.6	8.50	NNS 29	removed from analysis	
NNS 12	28.7	9.75	NNS 30	42.5	9.50
NNS 13	27.8	8.50	NNS 31	21.0	6.75
NNS 14	23.0	5.00	NNS 32	47.7	8.38
NNS 15	22.8	6.88	NNS 33	20.0	9.25
NNS 16	29.8	5.75	NNS 34	39.0	4.50
NNS 17	26.0	6.50	NNS 35	36.0	6.25
NNS 18	35.4	8.00	NNS 36	35.0	8.50
Mean	29.49	7.48	Mean	31.09	7.26
SD	9.5	1.65	SD	8.5	1.49
Range	18-59	3.00- 9.75	Range	20.0-47.7	4.50 - 9.25

We now turn to analysis of NNS participants' behavioral data from the discrimination and identification tasks as a function of biographical and experiential factors. The following factors were examined in these analyses:

1. age of arrival (AOA) in the L2 French environment, i.e. age of first immersion;
2. age of first exposure to L2 French (AOE, e.g. through classroom instruction or time spent in a French-speaking environment);
3. the proportion of daily use of French relative to the L1 and other languages spoken by the participant expressed as a percentage averaged over proportions of daily use of French at work, at home and with friends; proficiency as measured by an average of self-reported proficiency levels in speaking, reading, listening and pronunciation, each reported on a scale of 1 – 10;
4. frequency with which the participant is identified as a native speaker by native French speakers as reported on a scale of 1 – 10, with 1 representing 'never' and 10 representing 'always';
5. language preference as measured by an average percentage (relative to L1 English) of preference for reading, speaking, listening to radio and watching television or movies in French;

6. cumulative time spent in an English-speaking country;
7. cumulative time spent in an French-speaking country;
8. amount of formal (classroom) instruction in L2 French.

8.2.1 Experiments 1 and 1.1: AX Discrimination Task

Performance on the AX discrimination tasks in Experiments 1 (real-word stimuli) and 1.1 (non-word stimuli) was operationalized as a participant's average d-prime score across the six different pairs used in this task (e.g. 1_2, 1_3, 2_3, etc.) See Table 6-5 above for a complete reporting of real-word d-prime scores in Experiment 1 and Table 6-10 above for a complete reporting of non-word d-prime scores in Experiment 1.1. D-prime was chosen over mean accuracy as an indicator NNS performance on this task given that it is generally considered to be a more sensitive measure of perceptual discrimination. Mean d-prime scores were then analyzed as a function of biographical factors using both simple regressions and Spearman rank order correlations. The results of analyses for participants tested on real-word stimuli are given below in Table 8-2; analyses for participants tested on non-word stimuli are given below in Table 8-3.

The only consistent predictor of perceptual sensitivity to durational differences to emerge in the AX discrimination tasks across both NNS participant groups was the AOA of the participant in a French-speaking environment. A negative correlation between AOA and mean d-prime scores was observed. In other words, as age of immersion in the L2 environment increased for NNS participants, sensitivity to durational differences decreased. This result is consistent with the literature showing AOA to be a reliable predictor of attainment. No further experiential variables consistently predicted NNS participant behavior.

Table 8-2: *Correlation matrix: biographical factors as a function of mean d-prime scores for NNS participants in Experiment 1 (AX discrimination task employing real-word stimuli).*

BIOGRAPHICAL FACTOR	Value of r and p (simple regression)	Value of Rho and p (Spearman rank order)
Age of arrival	r= -.503, p = .0396	rho= -.444, p =.0760
Age of first exposure	r= -.317, NS	rho= -.112, NS
Proportion daily L2 usage	r= .002, NS	rho= .032, NS
Global proficiency	r= .144, NS	rho= .018, NS
Identified as native speaker	r= .320, NS	rho= .444, p =.0760
L2 usage preference	r= -.141, NS	rho= -.283, NS
Time spent in English-speaking country	r= -.425, NS	rho= -.390, NS
Time spent in French-speaking country	r= -.243, NS	rho= -.290, NS
Amount of formal study in L2	r= .124, NS	rho= .073, NS

Table 8-3: *Correlation matrix: biographical factors as a function of and mean d-prime scores for NNS participants in Experiment 1.1 (AX discrimination task employing non-word stimuli).*

BIOGRAPHICAL FACTOR	Value of r and p (simple regression)	Value of Rho and p (Spearman rank order)
Age of arrival	r= -.518, p = .0333	rho= -.485, p = .0522
Age of first exposure	r= -.087, NS	rho= .260, NS
Proportion daily L2 usage	r= .055, NS	rho= .222, NS
Global proficiency	r= -.115, NS	rho= -.147, NS
Identified as native speaker	r= .084, NS	rho= .299, NS
L2 usage preference	r= .002, NS	rho= .064, NS
Time spent in English-speaking country	r= -.457, p = .0654	rho= -.503, p = .0442
Time spent in French-speaking country	r= -.377, NS	rho= -.327, NS
Amount of formal study in L2	r= -.235, NS	rho= -.411, NS

8.2.2 Experiments 2 and 2.1: Forced-choice Identification Task

Performance on the forced-choice identification tasks in Experiments 2 and 2.1 was calculated as the average of two measures: the proportion of V-initial ('liaison') responses for shortened stimuli and the proportion of C-initial responses for lengthened stimuli (see Table 7-1 above for real-word response proportions in Experiment 2 and Table 7-5 above for non-word response proportions in Experiment 2.1). Responses to baseline stimuli were excluded from this analysis due to the fact that these stimuli represented average durational values across our production sample and therefore did not represent an indicator of sensitivity to this acoustic cue. The results of analyses for participants tested on real-word stimuli are given below in Table 8-4; analyses for participants tested on non-word stimuli are given below in Table 8-5.

Table 8-4: *Correlation matrix: biographical factors as a function of response proportions for NNS participants in Experiment 2 (forced-choice identification task employing real-word stimuli).*

BIOGRAPHICAL FACTOR	Value of r and p (simple regression)	Value of Rho and p (Spearman rank order)
Age of arrival	r= -.202, NS	rho= -.074, NS
Age of first exposure	r= -.150, NS	rho= -.174, NS
Proportion daily L2 usage	r= -.101, NS	rho= -.053, NS
Global proficiency	r= .241, NS	rho= .193, NS
Identified as native speaker	r= .404, p = .0965	rho= .448, p = .0648
L2 usage preference	r= .181, NS	rho= .044, NS
Time spent in English-speaking country	r= -.145, NS	rho= -.010, NS
Time spent in French-speaking country	r= -.317, NS	rho= -.139, NS
Amount of formal study in L2	r= .314, NS	rho= .265, NS

Contrary to an analysis of behavioral data from the AX discrimination tasks, AOA did not show a significant correlation with accuracy rates on the identification task. Correlation coefficients for this factor were consistently negative as would be predicted, however these coefficients failed to reach significance. Other experiential factors did not consistently predict responses.

Table 8-5: *Correlation matrix: biographical factors as a function of response proportions for NNS participants in Experiment 2.1 (forced-choice identification task employing non-word stimuli).*

BIOGRAPHICAL FACTOR	Value of r and p (simple regression)	Value of Rho and p (Spearman rank order)
Age of arrival	r= -.062, NS	rho= -.184, NS
Age of first exposure	r= .318, NS	rho= -.085, NS
Proportion daily L2 usage	r= .184, NS	rho= .287, NS
Global proficiency	r= -.401, NS	rho= -.273, NS
Identified as native speaker	r= -.331, NS	rho= -.196, NS
L2 usage preference	r= -.288, NS	rho= -.295, NS
Time spent in English-speaking country	r= -.117, NS	rho= -.123, NS
Time spent in French-speaking country	r= -.051, NS	rho= -.139, NS
Amount of formal study in L2	r= -.356, NS	rho= -.447, p = 0834.

8.2.3 Discussion

A correlation was observed between AOA and performance on the discrimination tasks, however correlations between AOA and performance on the identification tasks failed to reach significance. The differing outcomes between the two tasks would lead us to deduce that the processing levels and/or strategies involved in the two tasks are affected differently by age. The AX discrimination task involved decisions based primarily on the surface features of the input (though an effect of lexicality was observed as discussed above in §6.4), while the identification task involved the mapping of acoustic input onto mental representations in the lexicon. This difference would suggest that age has a more significant effect on lower-level processing of acoustic input than on higher-level lexical decision making.

Other biographical and experiential factors failed to show any consistent correlations with NNS performance on either task.

8.3 Nativelike Performance

Much work in psycho- and applied linguistics seeks to identify and quantify *nativelike* behavior on linguistic tasks on the part of non-native participants, i.e. non-native behavior that is indistinguishable from that of native controls. As Birdsong (in press) notes, “referencing learner performance to that of natives provides an easily understood metric of the potential for learner attainment.”

It is important to note, however, that native performance itself is a measure that must also be empirically established; it is neither uniform nor predictable. Once native performance has been quantified, nativelike behavior on the part of non-native subjects is usually operationalized as performance that falls either within the actual range of measurements obtained for native controls, or within 1 or 2 standard deviations above and below mean native measurements.

The quantification of *nativelike performance* in the current study is difficult given the degree of variation observed on the part of NS participants. This variation resulted in an extremely broad range of NS responses as well as large standard deviations on both the discrimination and identification tasks. For this reason we have employed even more stringent measures by which to quantify nativelikeness than are usually found in the literature. We have chosen to operationalize *nativelike performance* as NNS performance that is at or above native means themselves, as opposed to within 1 standard deviation above or below this mean as is often seen. We now turn to a discussion of instances of nativelike performance in the current study.

8.3.1 Experiments 1 and 1.1: AX Discrimination Task

In Experiment 1, which employed real-word stimuli in an AX discrimination task, 7 out of 18 NNS participants had mean d-prime scores across the six different conditions (i.e. 1_2, 1_3, 2_1, 2_3, 3_1, 3_2) at or above the NS mean of 1.09, as can be seen in Table 8-6 below.

Table 8-6: NS and NNS mean d-prime scores across six different pairs (1_2, 1_3, 2_1, 2_3, 3_1, 3_2) on Experiment 1 (AX discrimination task employing real-word stimuli). Mean scores, standard deviations, and range across participants are given at the bottom of the table. Cells shaded in grey indicate NNS performance at or above native mean.

	NS		NNS
	mean d-prime		mean d-prime
	across		across
	different pairs		different pairs
NS 1	0.94	NNS 1	0.36
NS 2	0.62	NNS 2	0.40
NS 3	0.58	NNS 3	0.79
NS 4	2.33	NNS 4	0.21
NS 5	1.16	NNS 5	removed from analysis
NS 6	0.97	NNS 6	1.36
NS 7	0.21	NNS 7	1.28
NS 8	0.90	NNS 8	1.63
NS 9	1.23	NNS 9	1.42
NS 10	0.68	NNS 10	0.47
NS 11	1.39	NNS 11	1.45
NS 12	1.24	NNS 12	0.86
NS 13	0.76	NNS 13	0.96
NS 14	0.55	NNS 14	1.66
NS 15	0.77	NNS 15	0.95
NS 16	1.74	NNS 16	1.35
NS 17	2.39	NNS 17	1.03
NS 18	1.16	NNS 18	0.57
MEAN	1.09	MEAN	0.98
SD	0.58	SD	0.46
RANGE	0.21 - 2.39	RANGE	0.21 - 1.66

Table 8-7 below shows mean d-prime scores from NS and NNS participants in Experiment 1.1, which employed non-word stimuli in an AX discrimination task. Two out of 17 NNS participants scored at or above the native mean of 0.93.

Table 8-7: NS and NNS mean d-prime scores across six different pairs (1_2, 1_3, 2_1, 2_3, 3_1, 3_2) on Experiment 1.1 (AX discrimination task employing non-word stimuli). Mean scores, standard deviations, and range across participants are given at the bottom of the table. Cells shaded in grey indicate NNS performance at or above native mean.

	NS mean d-prime across different pairs		NNS mean d-prime across different pairs
NS 19	0.34	NNS 19	0.56
NS 20	1.44	NNS 20	0.29
NS 21	-0.19	NNS 21	0.36
NS 22	1.11	NNS 22	0.75
NS 23	1.27	NNS 23	0.90
NS 24	0.61	NNS 24	0.88
NS 25	0.92	NNS 25	0.60
NS 26	0.52	NNS 26	0.47
NS 27	0.59	NNS 27	1.04
NS 28	1.04	NNS 28	0.51
NS 29	1.73	NNS 29	removed from analysis
NS 30	removed from analysis	NNS 30	1.64
NS 31	removed from analysis	NNS 31	0.37
NS 32	0.43	NNS 32	0.27
NS 33	-0.82	NNS 33	0.32
NS 34	0.82	NNS 34	0.43
NS 35	-0.2	NNS 35	0.40
NS 36	1.5	NNS 36	0.48
MEAN	0.93	MEAN	0.60
SD	0.58	SD	0.35
RANGE	-.019 – 1.73	RANGE	0.26 – 1.635

8.3.2 Experiments 2 and 2.1: Forced-choice Identification Task

NNS behavioral data from the forced-choice identification tasks was then analyzed. Recall from above that performance on the identification tasks was calculated as an average of the proportion of V-initial responses for shortened stimuli and the proportion of C-initial responses for lengthened stimuli. Table 8-8 below shows mean NS and NNS response proportions on Experiment 2, which employed real-word stimuli. Six out of 18 participants scored above the native mean of 73.50 %.

Table 8-8: Mean NS and NNS response proportions on Experiment 2 (forced-choice identification task employing real-word stimuli). Mean scores, standard deviations, and range across participants are given at the bottom of the table. Cells shaded in grey indicate NNS performance at or above native mean.

	NS mean response proportion		NNS mean response proportion
NS 1	67.36	NNS 1	53.48
NS 2	84.03	NNS 2	60.42
NS 3	69.45	NNS 3	63.19
NS 4	90.97	NNS 4	55.56
NS 5	92.36	NNS 5	55.56
NS 6	83.33	NNS 6	85.42
NS 7	57.64	NNS 7	80.56
NS 8	59.72	NNS 8	67.36
NS 9	76.98	NNS 9	91.67
NS 10	68.06	NNS 10	69.45
NS 11	73.62	NNS 11	79.17
NS 12	76.39	NNS 12	78.48
NS 13	56.95	NNS 13	70.14
NS 14	55.56	NNS 14	77.78
NS 15	66.67	NNS 15	63.20
NS 16	59.03	NNS 16	71.53
NS 17	92.52	NNS 17	34.72
NS 18	92.31	NNS 18	68.75
MEAN	73.50	MEAN	68.13
SD	13.24	SD	13.51
RANGE	55.55 - 92.52	RANGE	34.72 - 91.67

Table 8-9 below shows mean NS and NNS response proportions on Experiment 2.1, which employed non-word stimuli. Four out of 17 NNS participant scored above the native mean of 69.08 %.

Table 8-9: Mean NS and NNS response proportions on Experiment 2.1 (forced-choice identification task employing non-word stimuli). Mean scores, standard deviations, and range across participants are given at the bottom of the table. Cells shaded in grey indicate NNS performance at or above native mean.

	NS mean response proportion		NNS mean response proportion
NS 19	54.17	NNS 19	70.83
NS 20	60.42	NNS 20	63.89
NS 21	49.31	NNS 21	removed from analysis
NS 22	69.45	NNS 22	71.53
NS 23	75.00	NNS 23	65.97
NS 24	70.14	NNS 24	61.11
NS 25	84.73	NNS 25	57.64
NS 26	71.32	NNS 26	63.68
NS 27	57.64	NNS 27	73.61
NS 28	56.00	NNS 28	36.81
NS 29	removed from analysis	NNS 29	56.25
NS 30	60.64	NNS 30	68.75
NS 31	removed from analysis	NNS 31	72.92
NS 32	68.06	NNS 32	56.25
NS 33	63.20	NNS 33	49.31
NS 34	84.72	NNS 34	68.06
NS 35	90.28	NNS 35	56.25
NS 36	90.20	NNS 36	61.81
MEAN	69.08	MEAN	62.04
SD	12.97	SD	9.49
RANGE	49.31 - 90.28	RANGE	36.81 – 73.61

Note that five NNS participants tested on real-word stimuli performed at native levels on both the discrimination and identification tasks (NNS 6, 7, 9, 11, and 14) and one NNS (NNS 27) participant tested on non-word stimuli performed at native levels on both the discrimination and identification tasks.

8.3.3 Discussion

The current results contribute to a growing body of research on the upper limits of L2 attainment. Instances of natively-like performance in an L2 have been attested in

numerous experimental tasks dealing with L2 domains ranging from morphosyntax (Birdsong 1992; Birdsong & Molis 2001, Marinova-Todd 2003) to pronunciation (Birdsong 1992, 2003; Bongearts 1997, Marinova-Todd 2003). To our knowledge, the present study is the first to demonstrate nativelike attainment with respect to perceptual sensitivity to fine-grained acoustic detail in the L2.

Perhaps a more provocative finding concerning the performance of L2 participants in the current study is that non-contrastive phonetic detail in a L2 is acquired *at all*, let alone to nativelike levels. This finding raises the question as to how sensitivity to non-contrastive detail is acquired. We have already noted that the study of the acquisition of non-native *contrasts* has received much attention in the literature, however the acquisition of *non-contrastive* detail, in either a L1 or a L2, has not been tackled by many researchers. Exceptions include work offered by Peperkamp and colleagues, which we discuss below in §9.3.

CHAPTER NINE: Conclusion

9.0 General Discussion

As stated in the introduction, the two principal goals of this dissertation are as follows: 1) the investigation of the exploitation of segmental duration as a cue to disambiguation and segmentation in spoken French by native speakers and 2) an examination of the perceptual capacities of second language learners in the use of non-contrastive duration in the lexical interpretation of L2 French.

In the current chapter we address both of these goals in light of data presented in Chapters Five, Six and Seven. We will first recapitulate the major findings of the present study. We then address the first research goal in a discussion of how the exploitation of fine-grained acoustic detail in word recognition can be reconciled within currently accepted phoneme-based models of SWR. Our second research goal is addressed within a discussion of the acquisition of non-contrastive detail in a L2. Finally, we discuss limitations of the current study and possible further directions of research.

9.1 Result Summary

A production study and two behavioral experiments were conducted. The acoustic data from the production study presented in Chapter Five offers further evidence of durational differences between consonants that surface in environments of liaison and the same consonant in word-initial position—segments produced in liaison environments are significantly shorter than word-initial segments. Of particular note is the fact that this durational difference was observed in each of the six native speakers tested. The current production data adds to a burgeoning body of research demonstrating not only the use of segmental variation to signal the presence or absence of liaison in spoken French (Gaskell et al. 2002; Nguyen et al. 2007; Shoemaker 2006; Spinelli et al. 2003; Wauquier-Gravelines 1996), but also demonstrating acoustic variation at boundaries at multiple levels of prosodic organization in spoken French (Douchez & Lancia 2008; Fougeron 2001, 2007; Fougeron et al. 2002, 2003; Spinelli, Welby & Scheagis 2007; Welby 2003).

Chapter Six presented Experiments 1, 1.1 and 1.2, which examined the perceptual saliency of durational variation for both native and non-native speakers of French through the use of an AX discrimination task. This task employed real and non-word stimuli in which the durations of pivotal consonants had been instrumentally manipulated to create a three-step continuum of segmental duration. Durational values in this continuum were established from the distribution of the production sample presented in Chapter Five. Though these manipulated stimuli represented differences in absolute duration well above what would normally be considered just-noticeable differences, sensitivity to differences (as measured by *d*-prime analysis) was only robust when segments were separated by

two-degrees on the continuum, i.e. by durational values superior to two standard deviations from mean production values. Results suggest that allophonic differentiation in segmental duration is only perceptually salient when these differences represent extreme values in a normal distribution of speech. In addition, we observed that perception of segmental variation was superior in real-word stimuli relative to non-word stimuli, suggesting that higher-level information associated with real lexical items may facilitate the perception of variation in this phonological context.

One further significant result that emerged from the AX discrimination tasks was that sensitivity to durational differences in environments of liaison is conditioned by language exposure. Experiments 1 and 1.1 failed to show a conclusive difference in performance between the NS and NNS groups. For this reason, it was decided to test beginning learners of French in order to investigate whether nativelike performance on the part of the NNS groups was a result of language experience or could be attributable to general auditory processing strategies. The beginning group showed diminished sensitivity to durational variation in liaison environments relative to both the NS and the NNS groups, suggesting that the saliency of allophonic variation in this particular speech environment is indeed attributable to prolonged exposure to L2 French.

Experiments 2 and 2.1, presented in Chapter Seven, explored the exploitation of segmental duration in the lexical interpretation of sequences rendered ambiguous by the lexical assignment of pivotal consonants. Manipulated stimuli from the three-step durational continuum were employed in a forced-choice identification task. The results of these experiments demonstrated that utterance interpretation in spoken French can be influenced by segment duration alone. Participant responses were guided by the duration of the pivotal consonant. Participants interpreted a shortened consonant as an instance of liaison and a lengthened consonant as word-initial. Furthermore, participants adopted a guessing strategy in the baseline stimuli where pivotal consonants represented durations intermediate to those of LCs and ICs.

Of particular note is the fact that several advanced NNS participants showed nativelike sensitivity to durational variation in the interpretation to ambiguous sequences at or above native levels in both behavioral tasks, demonstrating that highly advanced learners can acquire the use of non-contrastive phonetic detail in L2 lexical processing.

We now turn to a discussion of the current findings within the broader picture of SWR processes.

9.2 The Exploitation of Fine-phonetic Detail within Current SWR Models

One essential question that emerges from the current study, as well as from a burgeoning body of research on the utilization of fine-grained phonetic detail in speech processing, is the issue as to how non-contrastive phonetic detail fits with currently accepted models of spoken word recognition. The exploitation of non-contrastive phonetic detail brings to the surface two specific challenges for current SWR models.

First, traditional phoneme-based recognition models (e.g. TRACE and Shortlist) assume that the mental lexicon is composed of word forms made up of *discrete* and *abstract* phoneme-sized units, which are accessed at the pre-lexical level. Once categorized, it is these phonemic units which access word units at the lexical level. These

models therefore propose that the phoneme is the smallest unit that accesses the lexicon. However, the exploitation of fine-phonetic detail as demonstrated in the current study suggests that listeners are not relying solely on abstract segmental representations of language, but rather that acoustic information is passed continuously to upper levels of processing. Current findings therefore suggest that the role of acoustic detail is not confined to pre-lexical processing, but that phonetic detail can directly affect lexical representations.

Similarly, evidence from bilingual processing poses challenges for phoneme-based models. Data from gating experiments have shown that bilingual listeners can differentiate lexical items in their two languages that share putatively identical phonemic content and that fine-grained detail can influence lexical access between a bilingual's two languages (Ju & Luce 2004). This line of research suggests that lexical items are not coded phonemically in the bilingual lexicon, but rather that bilinguals retain sensitivity to phonetic detail that serves to differentiate lexical items in the two languages.

A second assumption made by phoneme-based models is that the recognition of phonemes occurs sequentially; however, relativistic assessment of acoustic cues requires that phonemes be qualified in relation to both previous and following acoustic information. Specifically in the case of variation in segmental duration, Klatt (1976) pointed out that it may be impossible to make use of durational cues until after a listener has heard an entire utterance; only then can the listener evaluate whether lengthening or shortening is due to the inherent properties of the segment, stress, emphasis, etc. Furthermore, recent work exploring the perceptual effects of phonetic variation in non-adjacent segments (e.g. Holt 2005; Local 2003; Nguyen et al. 2004) has demonstrated that the categorization of phonemes can be affected by acoustic information above the level of the syllable.

These two issues, among others, have led many researchers over the past decade to question the comprehensiveness of phoneme-based SWR models and even to question the empirical validity of the phoneme itself. Lotto and Holt (2000) discuss what they refer to as *the illusion of the phoneme*, pointing out that the parameters traditionally invoked to characterize the phoneme — discreteness, abstraction, and language-specificity — are at odds with the reality of the speech signal, which is continuous, physically real, and not 'linguistically marked', in that the physical properties of the acoustic waveform do not differ from language to language (p. 4). These authors also voice their apprehension that the prominent role assigned to the phoneme in the study of speech perception has actually given rise to the illusion of more regularity (in both production and perception) than actually exists in the signal. They go on to note that many of the principal 'problems' that researchers seek to resolve in speech perception — e.g. lack of invariance and explicitly marked word boundaries, compensation for co-articulation, normalization of the signal — may not be 'problems' at all, but rather the consequence of researchers' theoretical assumptions about the nature of speech. These authors do not go so far as to deny the existence of the phoneme; their principal goal is rather to demonstrate that assumptions about its existence may lead researchers to overlook the presence of multiple streams of information in the signal.

One related issue that remains unresolved in the domain of SWR is that of the existence of a (universal) basic unit of perception. In Chapter One we discussed controversy in the domain of SWR as to what constitutes the basic unit of speech perception. Though psycholinguists are (mostly) in agreement that the speech signal is classified into some sort of abstract intermediate *pre-lexical* linguistic unit, there is less much agreement as to what this unit may be.

This controversy has also changed direction more recently as many researchers have come to question the need for such a primary unit at all. Nguyen and Hawkins (2003) in fact argue that there is no basic unit of speech perception, but rather that units of varying sizes may be simultaneously activated, with a natural bias for larger units to prevail upon smaller ones (see also Goldinger & Azuma 2003; Grossberg & Myers 2000). Nguyen and Hawkins (2003) observe,

Rather than there being one basic unit of speech perception, there may instead be a variety of competing candidates whose temporal domain depends not only on phonological, lexical and grammatical factors, but also on the dynamics of conversational interaction and on the particular demands of the experimental or other listening situation in which the listeners are placed. (p. 281)

A comprehensive debate on either the existence of the phoneme or candidates for the primary perceptual unit in speech is well beyond the scope of the present project, however, the current data, along with an established body of research demonstrating the continuous uptake of acoustic information in speech processing, bring into question many assumptions upon which the domain of SWR rests.

We discuss now alternative models of speech processing and modifications to existing models which could better accommodate the use of fine-grained acoustic detail.

9.2.1 Exemplar Models of Speech Processing

One alternative to phoneme-based models is to posit that the SWR system stores and makes use of multiple episodic traces of lexical representations. Exemplar-based models of spoken word recognition (e.g. Goldinger 1992, 1996; Johnson 1997; Pisoni 1997) propose that units are in effect stored as detailed acoustic traces in long-term memory. The lexicon is therefore made up of stores of multiple exemplars of words with varying acoustic detail; allophonic variation and fine-phonetic detail are stored at the lexical level with the mental representation of each word. Acoustic input is then matched against these exemplars at the lexical level with no intervening pre-lexical processing, allowing phonetic detail to be directly mapped onto lexical units.

Evidence for exemplar-based speech perception comes from a body of work suggesting that speakers are sensitive to speaker-specific acoustic information (see Lachs et al. 2003 for a review). For example, Palmeri, Goldinger and Pisoni (1993) demonstrated that listeners recall spoken words from a list more readily if presented with a token of the word produced by the same speaker. Exemplar models are also supported

by research showing that not only cues to speaker identity, but also speech rate, are retained in long-term memory and have an effect on word-recognition tasks (Bradlow, Nygaard & Pisoni 1999; Pisoni 1997). These phenomena are difficult to reconcile within a SWR system based on abstract phonological representations from which all ‘extraneous’ phonetic detail has been stripped through normalization.

As many researchers point out, an exemplar-based model is problematic in that it would require significant duplication of stored knowledge about each segment in that phonemes would need to be coded in all possible realizations. However, Hawkins (2003) counters this argument by noting that,

There seems no obvious reason why memories for words should be qualitatively different from other sorts of memories. They are developed from sensory percepts, and to the extent that they are abstract, the abstractions are developed from finding common factors amongst the many different pronunciations we have heard. (p. 379)

Specifically in the case of liaison, an exemplar approach could account for the current data. An exemplar model of liaison would entail that longer consonants be stored as examples of consonant-initial words while shorter consonants would be stored as (resyllabified) coda consonants. Thus a relatively longer /n/, for example, would better match stocked representations of *nerf* than representations of *un* in the sequence [œ.nɛʁ].

9.2.2 Probabilistic/Distributional Model of SWR

One further possibility is that the probabilistic distribution of fine-grained acoustic detail is encoded at the pre-lexical level and serves to bias initial lexical activation. Segmental duration can vary as a function of myriad factors in natural speech, thus it would be impractical to assume that duration is evaluated online as an absolute measure by listeners as would be assumed in a strict interpretation of an exemplar-based model; the same absolute value could be long in one context, but short in another.

A refinement to an exemplar model would therefore be to posit that exemplars are not necessarily stored as acoustic ‘photographs’ of lexical items, but rather that exemplars are processed probabilistically and that distributions are stored pre-lexically. Some exemplar-based models include mechanisms allowing for patterns to emerge from the accumulation of exemplars over time. Johnson (1997), for example, proposes that exemplars are not stored explicitly, but rather are encoded as weight modifications that lead to the creation of a quantized perceptual space onto which further input is mapped.

In the case of segmental duration, Shatzman and McQueen (2006) propose that, for example, longer consonants could serve to increase activation levels of syllable-initial allophones, whereas shorter consonants would serve to activate syllable-final allophones. These allophones would then be incorporated accordingly into larger lexical units. Furthermore, these authors note that the time course of the processing of durational variation observed in their eye-tracking study (discussed below in §9.5.1) would suggest

that durational information is likely not evaluated as an absolute measure, but rather relative to other information as it accumulates.

On this view, durational differences in these segments are not themselves encoded in the lexical representation, but rather may serve to bias the system in the localization of the word boundary. In the case of, for example, a relatively short /n/ in [œ.nɛʁ], the bottom-up activation process may either favor the candidate with a liaison consonant (e.g. *un*) and/or disfavor the consonant-initial word (e.g. *nerf*). Either of these possibilities would give more support to the selection of the vowel-initial candidate (e.g. *air*).

9.2.3 Parallel Segmental and Suprasegmental Analysis

One further account of fine-grained lexical access has recently been proposed involving parallel processing at multiple levels of linguistic organization. Instead of a SWR system based primarily on low-level segmental information, several authors have proposed a view in which segmental and suprasegmental analyses are in effect undertaken in parallel in the processing of the speech signal (Norris et al. 1997; Salverda et al. 2003; Shatzman and McQueen 2006). This view suggests that prosodic analysis is implemented in parallel to the segmental analysis, which would then act to favor intended hypotheses. Specifically, acoustic cues to word and syllable boundaries serve to favor lexical candidates whose boundaries align with prosodic boundaries predicted by these cues.

The goodness of fit of a particular instance of a phoneme is thus based not only on the distribution of acoustic properties that traditionally qualify the segment, but also on the perceived goodness of the segment in a particular prosodic position. This model therefore also functions in a probabilistic fashion, in that acoustic information reflects the likelihood that the duration of a particular segment corresponds to a predicted prosodic boundary. In the case of liaison in spoken French, for example, a longer consonant would suggest a preceding word boundary, while a shorter consonant would suggest resyllabification of a (latent) coda consonant.

This proposal adds to the controversy mentioned above as to what constitutes not only the primary unit of speech perception, but also the level at which this unit is exploited. Following this line of thought, Lotto and Holt (2000) note that, much like the notion of the phoneme, the notion of *processing levels* may need to be re-visited.

As we continue to study speech through forced-choice identification and discrimination paradigms, we should be cognizant of the fact that the representational level we are studying in these tasks may be one of several parallel representations that, perhaps, are not essential for the next “level” of perceptual or linguistic tasks. (p. 200)

The results of the current series of experiments do not offer sufficient data to differentiate among the above possibilities, however, they do underscore the fact that a sequential SWR model based on discrete phonemic units can not offer a comprehensive account of observed behavioral data.

9.3 The Exploitation of Fine Phonetic Detail in a L2

The current study also addresses an aspect of L2 phonological acquisition that has received little attention to date, namely the acquisition and exploitation of non-contrastive detail in a L2. The bulk of research, as well as models dealing with the acquisition of L2 phonological systems that have emerged from this research (e.g. Best 1995; Flege 1995), have dealt with the perception and/or production of phonetic *categories*.

While the current results have demonstrated that late learners *can* acquire sensitivity to non-contrastive allophonic detail in the L2, they offer no suggestion as to *how* this sensitivity is acquired. Recent work by Peperkamp and colleagues (Peperkamp 2003; Peperkamp, Pettinato & Dupoux 2003; Peperkamp, Le Calvez, Nadal & Dupoux 2006; Peperkamp & Dupoux 2007) has tackled the perception and acquisition of allophonic detail in both infants and adults. This body of work has suggested that allophonic variation is acquired mainly through statistical learning. Peperkamp et al. (2003) propose that rules governing instances of non-contrastive allophony emerge from a distributional analysis that exploits the fact that phonemes and their allophones are in complementary distribution.

However, Peperkamp et al. (2006) maintain that allophonic learning can not take place solely through exposure to distributions of variation. In an examination of a corpus of child-directed speech, these authors point to the existence of many near-complementary distributions that do not correspond to allophonic variation and which therefore could lead to the spurious assumption that two segments map onto the same underlying form. For example, they point out that, in French, the vowel /œ/ appears uniquely in closed syllables and therefore always before consonants (e.g. *peur* [pœʁ] ‘fear’), while the semi-vowel /ɥ/ occurs as the last segment in syllable onsets and therefore always before full vowels as in *pluie* [plɥi] ‘rain’. This distribution could lead to the assumption that these two segments are allophones of the same phoneme. Similarly, in English /h/ is always syllable-initial, while /ŋ/ is always syllable-final. By consequence these two phonemes are in complementary distribution, though they do not share an underlying form. Given distributions such as these, these authors note that the acquisition of allophonic variation is likely not dependent on statistical algorithms alone.

Continuing this line of inquiry, Peperkamp and Dupoux (2007) showed that statistical learning of allophonic distributions in adults can be augmented by semantic information associated with lexical items. These authors tested the acquisition of allophonic distributions in an artificial language by adults and showed that the learning of an allophonic rule was more robust when accompanied by semantic information. In the first experiment in this study, participants learned an allophonic voicing distribution through the association of the same image to a word containing both a voiced and an unvoiced allophone in the context that licenses the allophonic change. These participants thus learned that both variants mapped onto the same lexical item. In the second experiment, participants were exposed to the same distribution over the same amount of time, but only one variant of the allophone was associated to an image. Participants in the

first experiment were able to transfer the newly learned allophonic rule to novel items, while participants in the second experiment, who had been exposed to the same distribution, but without accompanying semantic information, were not.

The work of Peperkamp and colleagues suggests that statistical learning alone may not be sufficient for the acquisition of non-contrastive variation, but rather that allophonic variation is acquired through the application of statistical algorithms to distributions of speech as well as through the association of meaningful content. These results are in line with a body of research which has demonstrated that infant phonological learning is augmented by meaningful human interaction (see Kuhl 2004 for a review). Similarly, Carroll (1999) proposes that meaningful conversation is the main context in which the phonological properties of a second language are learned.

To our knowledge, research on the acquisition of non-contrastive detail has only been investigated through the acquisition of artificial speech by adults and infants and through algorithms trained on corpus data from natural speech. We are not aware of any work to date that has specifically examined the acquisition of allophonic variation by adults in real second language learning environments.

9.4 Limitations

We address briefly in this section limitations of the current investigation which may have affected observed outcomes.

9.4.1 Performance Variation among Participants

As we have already discussed in detail, a significant amount of variation was observed among participants in both participant groups. A good deal of variation was expected among NNS participants, as the learning of a L2 is subject to the idiosyncratic and individual experiences of every learner, however, the degree of variation observed in the NS groups was somewhat unexpected.

In §6.1 above we discussed several factors which may render the perception of segmental duration in this particular phonological environment less robust, namely the length and complexity of the stimuli in which the variation occurs and the fact that the variation investigated here represents allophonic (within-category) variation. In an attempt to further examine both the extent of variation among participants as well as the relationship between the perceptual saliency of these durational differences and their exploitation in lexical processing, we investigated whether participant performance in Experiments 1 and 1.1 predicted performance in Experiments 2 and 2.1. In other words, we explored the hypothesis that participants who showed stronger sensitivity in the AX discrimination task may also exhibit increased exploitation of these cues in the identification task. As in the previous chapter, performance on the discrimination tasks is measured by mean d' -prime scores across the six different pairs, while performance on the identification tasks is measured by an average of the proportion of V-initial responses for shortened stimuli and C-initial responses for lengthened stimuli. Table 9-1 below presents the results of correlations between these two measures, which show that this hypothesis was borne out for participants tested on real-word stimuli only.

Table 9-1: *Correlation matrix: behavioral results from Experiments 1 and 1.1 (AX discrimination tasks) as a function of behavioral results from Experiments 2 and 2.1 (forced-choice identification tasks).*

		Value of r and p (simple regression)	Value of Rho and p (Spearman rank order)
Participants tested on Real-word stimuli (Experiments 1 & 2)	NS d-prime scores *	r = .560,	Rho = .564,
	ident accuracy	p = .0157	p = .0201
	NNS d-prime scores *	r = .531,	Rho = .627,
	ident accuracy	p = .0284	p = .0121
Participants tested on Non-word stimuli (Experiments 1.1 & 2.1)	NS d-prime scores *	r = .352,	Rho = .293,
	ident accuracy	p = NS	p = NS
	NNS d-prime scores *	r = -.441,	Rho = -.447,
	ident accuracy	(p = .0876)	(p = .0834)

As can be seen, for real-word stimuli, sensitivity to durational differences in the discrimination task strongly predicted the use of duration as a segmentation cue in the identification task. Both NS and NNS participants tested on real-word stimuli showed significant correlations between the two tasks, suggesting that sensitivity to acoustic variation at the acoustic level predicted the exploitation of this variation at the lexical level.

Results for participants tested on non-word stimuli, however, showed no significant correlation for NS between the two tasks. Note also that NNS participants showed a correlation coefficient that not only failed to reach significance, but was negative as well. These results combined with those from real-word stimuli further point to an effect of lexicality on the exploitation of this particular acoustic cue.

We have already noted that the high degree of variation in our behavioral data suggests that this particular cue may not be exploited to the same extent by all listeners. The results of these correlations underscore this fact, however reasons as to why this might be the case remain to be determined. We see two plausible explanations. First, the amount of variation could be attributable to attentional differences on the part of individual participants. Given the difficulty of the tasks, some participants may simply have adopted a guessing strategy instead of devoting sufficient attentional and cognitive resources required for the completion of the tasks.

One other possibility is that the observed variation reflects the fact that the phenomenon of liaison itself is highly variable. As discussed above in Chapters Two and Seven, recent work from corpus analysis has suggested that the *production* of liaison is much more variable than once assumed — even in so-called ‘obligatory’ cases of liaison (Durand & Lyche 2008). The degree of variation observed in both the production and perception of liaison is not surprising given that, as Durand and Lyche (2008) note “liaison cannot be seen as a single phonological process, [given that] it is partly morphosyntactic, partly phonological, partly phonetic and partly the result of the speaker’s knowledge of the orthographic system, particularly in the areas most sensitive to sociostylistic variation.” (p. 34). It is worth noting as well that the realization of liaison is not only a highly variable phenomenon, but also seems to show a trend toward

diminished usage in modern spoken French (see Bybee 2001 a/b for discussion). This variability and diminished usage in the realization of liaison could feasibly affect perception as well.

9.4.2 Regional Accent

One additional factor that could have rendered these tasks more difficult, thus increasing individual variation, brings to the fore a methodological issue. The speaker chosen to be the source of instrumentally manipulated stimuli was a speaker of Southwestern French (from Agen, northwest of Toulouse). The Southern French accent is noticeably different from what is considered to be Standard French (the French of the Paris metropolitan area and Ile-de-France). This accent is characterized by, among other traits, qualitative differences in nasal vowels relative to Standard French, and the realization of word-final schwa.

Some research has shown the word recognition processes can be disrupted by regional accents. Wright et al. (2008) showed that both recognition accuracy and reaction times were affected in the perception of English vowel categories when vowels were produced by speakers of various regional accents. Girard, Goslin and Floccia (2004, 2005) also showed that reaction times were slowed in word recognition in French when lexical items were produced in a regional accent relative to items produced in Standard French. This methodological issue could account for diminished sensitivity to this particular acoustic cue and, hence, some of the variation observed among both participant groups.

9.5 Further Research Directions

In this section we explore possible directions of research on the processing of liaison environments by both native and non-native speakers which could add to the current body of knowledge in this domain. Given the degree of variation observed in the current study, the highly variable character of liaison as a phonological phenomenon, and other limitations discussed above, further research on this topic is warranted.

9.5.1 Eye-tracking Employing Manipulated Stimuli

The current study has employed two separate methodological paradigms in the investigation of the exploitation of segmental duration in spoken French — an AX discrimination task and a forced-choice identification task. Due to the offline nature of these behavioral tasks, they unfortunately offer little insight into how these durational differences might be used in *online* speech processing. These tasks record response accuracy and reaction time only once a trial is completed; therefore, they offer no evidence as to the time course of the actual implementation of this particular acoustic cue.

Further research could make use of methodologies that more closely model the processing of speech in close to real time, such as eye-tracking. This technique is used to investigate the time course in the exploitation of acoustic cues in speech processing by taking advantage of the fact that the probability of fixating an image varies with the

goodness of fit between the image and the acoustic input. (See Tanenhaus & Spivey-Knowlton 1996 for a comprehensive discussion of this methodology.)

Following the results of Shatzman & McQueen (2006), we would expect stimuli in which durational differences due to the presence or absence of liaison have been exaggerated to offer a window into the time course of the processing this particular cue in spoken French. In this study, reviewed above in §1.4.2.3.1, these authors showed in two separate eye-tracking experiments that the time course of the integration of acoustic information can be altered by the degree of information in the signal.

In both experiments, participants' eye movements were tracked as they listened to sequences in Dutch rendered ambiguous by the lexical assignment of /s/ (e.g. *eens pot* 'once jar' and *een spot* 'a spotlight'); participants were then asked to click on images on a computer screen. In the first experiment, the visual target was, for example, /p/-initial *pot* 'jar'. Images also included /sp/-initial competitors (e.g. *spin* 'spider'). Listeners heard two versions of the target embedded in the sentence *ze heft wel eens pot gezegd* 'she did say once jar' — one version (identity-spliced) in which the /sp/ sequence had been spliced from a different token of the same sentence, and another version (cross-spliced) in which the /sp/ sequence had been spliced from a version produced as cluster-initial, *een spot* 'a spotlight'. As discussed in Chapter One, participants fixated more and longer on the visual target *pot* after hearing the identity-spliced versions, while participants initially fixated more and longer on the competitor *spin* after hearing the cross-spliced versions, suggesting sensitivity to fine-grained differences between the two versions.

In the second experiment, participants heard phrases which were produced as the same physical utterance, but in which the duration of /s/ had been altered. The results of Experiment 1 were replicated, and even stronger fixations to competitors were observed, indicating that this particular acoustic cue was modulating lexical access. Of particular interest is the fact that eye-movements showed that fixations started significantly later when listening to the spliced versions of the target in Experiment 1 than when listening to the exaggerated versions of the target in Experiment 2, suggesting that when there is less acoustic information available, listeners wait for more information to accumulate before fixating the lexical target. When listening to manipulated stimuli, where acoustic differences between the target and competitor are enhanced, however, fixations started earlier, suggesting that the exaggerated durations of the pivotal segment /s/ offered enough acoustic information to allow the listener to choose the target before letting further input accumulate.

As noted above, the difference in time course between the two conditions (cross-spliced and enhanced duration) suggests that segmental duration is used in a probabilistic fashion, i.e. that it is evaluated in relation to other acoustic information. This methodology could give similar insight into the processing of acoustic cues to liaison.

9.5.2 Perceptual Training in L2 Phonology

In Chapter Three we brought up the question as to which cues to speech processing and segmentation could be amenable to perceptual training for the L2 learner. Kuhl (2000) notes that features which characterize *motherese*, the speech to which infants are exposed in L1 acquisition, such as “exaggerated acoustic cues, multiple instances by

many talkers, and mass listening experience” have also proved beneficial to L2 acquisition (p. 11855). We explore in this section the first feature to which Kuhl refers — *exaggerated acoustic cues*. As we noted in §3.4, phonetic training has resulted in the improvement of the perception of some L2 segmental contrasts (e.g., Bradlow, Pisoni, Akahane-Yamada & Tohkura 1997; Lively, Pisoni, Yamada, Tohkura & Yamada 1994; McCandliss, Fiez, Protopapas, Conway & McClelland 2002; McClelland, Fiez & McCandliss 2002, see also Logan & Pruitt 1995). Much of this body of work has successfully employed the instrumental exaggeration of acoustic features, suggesting that perceptual cues with clear acoustic correlates could benefit from similar training techniques.

McClelland et al. (2002) and McCandliss et al. (2002), for example, employed an adaptive training regime within a framework of Hebbian learning in the training of the English /l-r/ contrast for speakers of Japanese. A Hebbian model of associative learning, in which the repetitive activation of neuronal synapses leads to an increase in the strength of synaptic connections, predicts that adult learners of a second language experience difficulties in the perception of L2 contrasts due to the strength of neural connections established by the acquisition of the L1 phonological system. When the learner is then confronted with L2 sounds that may be similar to established L1 categories, the existing connections are actually strengthened as the similar sound is assimilated to an existing category. Subsequently, new categories are difficult to create.

McClelland et al. (2002) and McCandliss et al. (2002) hypothesized that the same neuronal system that leads to the entrenchment of the L1 phonology could be exploited in the training of L2 contrasts by provoking the creation of new neural connections through adaptive training. In a demonstration of adult brain plasticity, these authors showed that Japanese speakers’ perception of the /l-r/ contrast improved with training on stimuli in which acoustic differences between the two segments had been exaggerated and then gradually brought within normal production range. In the McClelland et al. (2002) study, Japanese participants were able to better distinguish the sounds after three 20-minute sessions.

While perceptual training has seen much success in the acquisition of non-native phonemic *contrasts*, we are aware of no work to date concerning the training of L2 learners on cues to speech segmentation or the perception of allophonic variation in a L2. However, it seems feasible that techniques similar to those used in adaptive perceptual training could be applied to cues such as segmental duration or stress accent, which have clearly identifiable acoustic correlates. In the case of the perception of liaison consonants in spoken French, for example, non-native listeners could be subjected to adaptive training on these segments in which the durational differences between LCs and ICs have been exaggerated, as we have done in the current study. Further research would be needed to test the empirical accuracy of this hypothesis.

9.5.3 L2 Production

The notion that perception precedes production in the acquisition of L2 phonology is a well established one (e.g. Flege 1987). However, to our knowledge, no research to date has explored the production of allophonic variation in a second language.

Further research could therefore explore the relationship between production and perception by eliciting productions of ambiguous stimuli similar to those employed in the current study from advanced speakers of French as a second language. Following a body of research that has uncovered instances of nativelike L2 pronunciation on the part of adult learners at both segmental and global levels (e.g. Birdsong 2003, 2007; Bongaerts 1999, 1991), we predict that highly advanced learners would show the same segmental durational variation in environments of liaison as native speakers.

9.6 Conclusion

In this dissertation we have examined the role of segmental duration in the interpretation and processing of environments of ambiguity in spoken French. The current study has investigated the perceptual capacities of both native French speakers and adult learners of L2 French in discriminating durational differences that arise between segments produced in word-initial position and segments that surface in liaison. These results build on previous research on the perception and processing of liaison environments in spoken French. Priming data from natural speech in work by Spinelli et al. (2003) showed that ambiguity in liaison environments does not impede lexical access to vowel-initial candidates even though liaison and resyllabification would be predicted to complicate this process. These authors suggested that speakers differentiate liaison environments acoustically from non-liaison environments by varying the duration of pivotal consonants, though the authors did not test this hypothesis directly.

Shoemaker and Birdsong (2008; see also Shoemaker 2005) continued this line of research with a more direct test of listeners' abilities to disambiguate globally ambiguous phrases through the use of a forced-choice identification task employing the same auditory stimuli used in the Spinelli et al. (2003) study. Shoemaker and Birdsong also compared the perceptual capacities of native French speakers and highly advanced learners of L2 French. Though response distributions correlated significantly for these two participant groups, this study failed to demonstrate a direct relationship between the duration of pivotal consonants and response patterns for either group.

This dissertation has further expanded this body of research by directly demonstrating that segmental duration can indeed affect the processing and lexical interpretation of liaison environments by manipulating this single acoustic cue while holding all other information in the signal constant. In exaggerating these differences the methodology employed here has offered participants "the best of conditions" (Warner et al. 2004: 266) in which to differentiate ambiguities due to possible liaison in spoken French. In isolating and exaggerating the durational differences in the pivotal consonants in ambiguous pairs, the series of experiments reported here has shown that French listeners do indeed use this cue in speech segmentation and word recognition.

Furthermore, we have demonstrated that advanced learners of L2 French are also sensitive to this non-contrastive phonetic detail. We have also provided evidence that sensitivity to this cue is conditioned by language exposure by showing that beginning learners of French show diminished sensitivity to segmental duration relative to both NS and advanced NNS. Of particular interest for the study of the upper limits of L2 acquisition is the fact that several advanced L2 learners exhibited sensitivity to this cue at

nativelike levels, offering evidence that the adult perceptual system retains plasticity into adulthood.

As discussed at length above, the speech signal is characterized by substantial amounts of variation and uneven distributions of acoustic factors. Given the fact that listeners are likely exposed to a distribution of spoken French in which individual tokens of consonants in initial and liaison position may or may not exhibit the durational variation discussed here, the fact that listeners interpret speech in the predicted direction when exposed to manipulated tokens of these segments shows that this parameter does indeed have cue value. If this cue were not encoded as a phonological rule in listeners' grammars, no effect of duration would be observed in the behavioral tasks we have employed here.

If researchers are to understand how listeners segment continuous speech, it is necessary to integrate evidence culled from a variety of laboratory techniques and speech environments. Isolating individual cues allows us to form a sort of schematic of a listener's phonological grammar and the inventory of tools used in the comprehension of a highly variable speech signal. Though the particular cue investigated here may not be extremely robust in natural speech, it is nonetheless a part of native (and non-native) speaker's phonological inventory and therefore must be included in any comprehensive model of spoken word recognition.

Appendix I: Consent Form (Production of Materials)

CONSENT FORM

Title: *Durational cues to speech segmentation in French*

IRB PROTOCOL # 2007-04-0039

Conducted By: Ellenor Shoemaker (eshoemaker@mail.utexas.edu)

Faculty Sponsor: Dr. David Birdsong, Department of French and Italian,
University of Texas at Austin, Austin, TX 78712 USA

You are being asked to participate in a research study. This form provides you with information about the study. The person in charge of this research will also describe this study to you and answer all of your questions. Please read the information below and ask any questions you might have before deciding whether or not to take part. Your participation is entirely voluntary. You can refuse to participate without penalty or loss of benefits to which you are otherwise entitled. You can stop your participation at any time and your refusal will not impact current or future relationships with UT Austin or participating sites. To do so simply tell the researcher you wish to stop participation. The researcher will provide you with a copy of this consent for your records.

The purpose of this study is to investigate durational differences between consonants that surface in liaison environments and consonants that appear in other word positions.

If you agree to be in this study, we will ask you to do the following:

- Sit in a recording booth and read a prepared list of phrases which will be recorded onto a solid state recorder.

Total estimated time to participate in this study is 30 minutes.

Risks of being in the study:

- There are no known physical risks associated with these procedures beyond those of everyday life.
- Risks to privacy are minimal because you are only providing information on age, gender and native language and you will be reading a prepared list of phrases, thus no personal information will be included in the recordings.
- This study may involve risks that are currently unforeseeable. If you wish to discuss the information above or any other risks you may experience, you may ask questions now or call the Principal Investigator listed on the front page of this form.

Potential benefits of being in the study:

- There are no benefits for participation in this study.

Compensation:

- You will receive \$5 for your participation in this study.

Confidentiality and Privacy Protections:

- The only personal information gathered in this study will be age, gender and native language.
- The data resulting from your participation may be made available to other researchers in the future for research purposes not detailed within this consent form. In these cases, the data will contain no identifying information that could associate you with it, or with your participation in any study.

The **records** of this study will be stored securely and kept confidential. Authorized persons from The University of Texas at Austin, members of the Institutional Review Board, and (study sponsors,

if any) have the legal right to review your research records and will protect the **confidentiality** of those records to the extent permitted by law. All publications will exclude any information that will make it possible to identify you as a subject. Throughout the study, the researchers will notify you of new information that may become available and that might affect your decision to remain in the study.

Contacts and Questions:

If you have any questions about the study please ask now. If you have questions later, want additional information, or wish to withdraw your participation call the researchers conducting the study. Their names, phone numbers, and e-mail addresses are at the top of this page. If you have questions about your rights as a research participant, complaints, concerns, or questions about the research please contact Lisa Leiden, Ph.D., Chair of The University of Texas at Austin Institutional Review Board for the Protection of Human Subjects, (512) 471-8871 or email: orsc@uts.cc.utexas.edu.

You will be given a copy of this information to keep for your records.

Statement of Consent:

I have read the above information and have sufficient information to make a decision about participating in this study. I consent to participate in the study.

Signature: _____ Date: _____

Permission for use of the recording:

We may wish to present some of the tapes from this study at scientific conventions or as demonstrations in classrooms. Please sign below if you are willing to allow us to do so with your tape. I hereby give permission for the video (audio) tape made for this research study to also be used for educational purposes.

Signature: _____ Date: _____

Signature of Investigator: _____ Date: _____

Appendix II: Non-Native Participant Biographical Questionnaire

Biographical questionnaire: Non-Native

This questionnaire concerns your language experience over the course of your lifetime. Feel free to elaborate where you think it would be helpful to our study. If there are questions you prefer not to answer, please feel free to skip them.

All responses are confidential. Your name, telephone number, and email address will be useful in case we need follow-up information but will be shared with no one.

THANK YOU FOR YOUR PARTICIPATION!

Today's date: _____

Name: _____

Date of birth: _____

Place of birth: _____

Email: _____

Telephone number: _____

Gender: M F

1) Please list all languages that you speak in order of acquisition (native language first) and the age at which you started learning that language.

Language 1. _____ ENGLISH _____ Age of exposure ____ 0 ____

Language 2. _____ Age of exposure _____

Language 3. _____ Age of exposure _____

Language 4. _____ Age of exposure _____

2) Please give the percentage of time that you are *currently* and *on average* exposed to each of your languages. (Your percentages should add up to 100%.)

3) Please give the percentage of time that you *currently use* each of your languages *at work*. (Your percentages should add up to 100%.)

4) Please give the percentage of time that you *currently use* each of your languages *at home*. (Your percentages should add up to 100%.)

5) Please give the percentage of time that you *currently use* each of your languages *with friends*. (Your percentages should add up to 100%.)

6) When choosing to read a text in any of your languages, what percentage of the time would you choose to read it in each of your languages? Assume that the text was originally written in a language that you do not know. (Your percentages should add up to 100%.)

7) When choosing a language to speak with another person who is equally fluent in all of your languages, what percentage of time would you choose to speak each of your languages? (Your percentages should add up to 100%.)

8) When listening to the radio, what percentage of the time would you choose to listen in each of your languages? (Your percentages should add up to 100%.)

9) When watching television or movies, what percentage of the time would you choose to watch in each of your languages? (Your percentages should add up to 100%.)

10) On a scale of 1 (least nativelike) to 10 (most nativelike), please rate your speaking proficiency in each of your languages (including English).

Language 1. ENGLISH _____
Language 2. _____
Language 3. _____
Language 4. _____

11) On a scale of 1 (least nativelike) to 10 (most nativelike), please rate your reading proficiency in each of your languages (including English).

Language 1. ENGLISH _____
Language 2. _____
Language 3. _____
Language 4. _____

12) On a scale of 1 (least nativelike) to 10 (most nativelike), please rate your listening proficiency in each of your languages (including English).

Language 1. ENGLISH _____
Language 2. _____
Language 3. _____
Language 4. _____

13) On a scale of 1 (least nativelike) to 10 (most nativelike), please rate your pronunciation in each of your languages (including English).

Language 1. ENGLISH _____
Language 2. _____
Language 3. _____
Language 4. _____

14) On a scale of 1 to 10, how frequently do others identify you as a native speaker of each of your languages? (1 = never, 5 = half the time, 10 = always)

Language 1. ENGLISH _____
 Language 2. _____
 Language 3. _____
 Language 4. _____

15) Please list any time spent in an English-speaking country. Indicate the cities and periods of time below.

I lived in _____ from _____ to _____
 TOTAL = _____ years _____ months

16) Please list any time spent in a French-speaking country. Indicate the cities and periods of time below.

I lived in _____ from _____ to _____
 TOTAL = _____ years _____ months

17) In your learning of French, what percentage do you feel that you learned through **formal** language instruction and what percentage do you feel that you learned in more **informal** environments (e.g. interactions with other people, watching TV/movies, etc). (Your percentages should add up to 100%.)

18) Please indicate the approximate periods during which you formally studied French. Circle 'school' or 'university' as appropriate.

In school / university, I studied French from _____ to _____
 In school / university, I studied French from _____ to _____
 In school / university, I studied French from _____ to _____
 In school / university, I studied French from _____ to _____
 TOTAL = _____ years _____ months

19) How many years of formal education have you completed? _____

20) What is the highest level of formal education that you completed? _____

21) Please include any additional comments or information about your language use and history that you feel could be important to this survey or that you feel should have been addressed above:

Appendix III: Consent Form (Behavioral Tasks: French)

DECLARATION DE CONSENTEMENT

Titre: *Indices consonantiques sur la segmentation de la parole en français*

N° IRB # 2007-04-0039

Chercheur Principal/responsable: Ellenor Shoemaker (eshoemaker@mail.utexas.edu)

Directeur de recherches: Dr. David Birdsong, Department of French and Italian,
University of Texas at Austin, Austin, TX 78712 USA

Vous êtes invité(e) à participer à une étude qui compare des consonnes de liaison et des consonnes qui apparaissent en début de mots. Le présent document explique ce qui est attendu de vous, dans le cadre de cette étude. La responsable de l'étude vous précisera ces détails et est à votre disposition en vue de répondre à toutes vos questions. Veuillez lire attentivement les informations ci-dessous, pour décider de votre éventuelle participation. Votre décision n'affectera nullement vos relations avec l'Université du Texas à Austin. Si vous répondez favorablement, vous ferez partie des 36 locuteurs français qui prendront part à l'expérience. Si, pour une quelconque raison, vous changez d'avis, vous êtes autorisé(e), à n'importe quel moment, à refuser de participer à l'expérience. Le cas échéant, il vous suffira de nous le faire savoir.

L'objectif de cette étude est de déterminer si les locuteurs français sont capables de distinguer les différences acoustiques entre les consonnes de liaison et les consonnes initiales.

Si vous acceptez de participer à cette expérience:

- Dans la première partie de l'expérience, il vous sera demandé d'écouter des paires de groupes de mots, que vous entendrez au moyen d'un casque audio relié à un ordinateur portable. Au fur et à mesure que vous entendez ces paires, vous devrez indiquer si les groupes de mots sont les mêmes, ou bien s'ils sont différents.
- Dans la deuxième partie de l'expérience, il vous sera demandé d'écouter une liste de groupes de mots. Au fur et à mesure que vous entendez les groupes de mots, des mots s'afficheront sur l'écran. Vous devrez indiquer lequel de ces deux groupes de mots vous venez d'entendre.

Cette activité ne vous prendra pas plus de 75 minutes.

Risques liés à votre participation à cette étude :

- Ce projet ne comporte ni risques ni inconvénients matériels ou psychologiques.
- Les risques relatifs à votre confidentialité sont minimes puisque nous limitons les informations prises sur les participants à l'âge, le sexe, la langue maternelle et les langues étrangères parlées.
- Cette étude pourrait comprendre des risques qu'il est actuellement hors de nos capacités de prévoir. Si vous désirez un complément d'informations, vous pouvez formuler vos questions tout de suite ou directement contacter le chercheur responsable [indiqué ci-dessus].

Bénéfices Potentiels: Aucun avantage immédiat.

Compensation: Vous recevrez 8 € pour votre participation.

Confidentialité:

- Votre anonymat sera strictement respecté durant et après le projet.
- Les seules informations que nous vous demandons sont: votre âge, votre sexe, votre

langue maternelle et les autres langues que vous parlez.

- Les données pourront être utilisées dans d'autres études futures, dont ce formulaire ne fait pas état. Le cas échéant, votre anonymat sera respecté.

Les **données** de cette étude seront gardées dans un endroit sûr. Seuls les personnels autorisés de l'Université du Texas à Austin, ainsi que les membres du *Institutional Review Board* auront le droit d'accéder à ces données. Ils se portent garant de la confidentialité des informations que vous livrez. Toute publication issue de cette étude n'inclura aucune information identifiante.

Contacts et Questions:

Nous vous remercions de votre participation. Si vous avez des questions, n'hésitez pas à nous les poser. Si vous désirez plus de renseignements par rapport à votre participation à cette étude, veuillez vous adresser au chercheur responsable (voir coordonnées ci-dessus). Si vos questions portent sur vos droits en tant que participant, ou si vous voulez nous adresser des réclamations ou des plaintes, veuillez prendre contact avec Lisa Leiden, Ph.D., Chair of The University of Texas at Austin Institutional Review Board for the Protection of Human Subjects (tél : 001.512.471.8871 ; email : orsc@uts.cc.utexas.edu).

Une copie de ce formulaire vous sera fournie.

Déclaration de consentement :

En apposant votre signature ci-dessous, vous reconnaissez avoir compris les conditions de l'étude exposées ci-dessus, et exprimez votre consentement en vue d'y participer.

Signature : _____ Date: _____

Signature du chercheur: _____ Date: _____

Appendix IV: Consent Form (Behavioral Tasks: English)

CONSENT FORM

Title: *Durational cues to speech segmentation in French*

IRB PROTOCOL # 2007-04-0039

Conducted By: Ellenor Shoemaker (eshoemaker@mail.utexas.edu)

Faculty Sponsor: Dr. David Birdsong, Department of French and Italian,
University of Texas at Austin, Austin, TX 78712 USA

You are being asked to participate in a research study. This form provides you with information about the study. The person in charge of this research will also describe this study to you and answer all of your questions. Please read the information below and ask any questions you might have before deciding whether or not to take part. Your participation is entirely voluntary. You can stop your participation at any time and your refusal will not impact current or future relationships with the University of Texas at Austin or participating sites. To do so simply tell the researcher you wish to stop participation. The researcher will provide you with a copy of this consent for your records.

The purpose of this study is to investigate the perceptual capacities of highly advanced non-native French speakers in disambiguating pairs of phrases rendered ambiguous by liaison.

If you agree to be in this study, we will ask you to do the following:

- Fill out a biographical questionnaire about your language experience.
- Sit at a laptop computer and listen to a series of phrases in French presented aurally through headphones.
- In the first portion of the experiment, you will hear a series of phrases presented in pairs. After hearing each pair, you will be asked to indicate whether the two phrases you have just heard are the same or different.
- In the second portion of the experiment, you will hear a series of phrases presented individually, after each phrase you will see a pair of words presented visually on the laptop screen. You will then be asked to indicate which of the pair you have just heard.

Total estimated time to participate in this study is 90 minutes.

Risks of being in the study:

- There are no known physical risks associated with these procedures beyond those of everyday life.
- Risks to privacy are minimal because you are only providing information on age, gender and information on language history and usage.
- This study may involve risks that are currently unforeseeable. If you wish to discuss the information above or any other risks you may experience, you may ask questions now or call the Principal Investigator listed on the front page of this form.

Potential benefits of being in the study:

- There are no benefits for participation in this study.

Compensation:

- You will receive 8 € for your participation.

Confidentiality and Privacy Protections:

- The only personal information gathered in this study will be age, gender, native language

and information on your language use throughout your lifetime.

- The data resulting from your participation may be made available to other researchers in the future for research purposes not detailed within this consent form. In these cases, the data will contain no identifying information that could associate you with it, or with your participation in any study.

The **records** of this study will be stored securely and kept confidential. Only authorized persons from The University of Texas at Austin, members of the Institutional Review Board, and (study sponsors, if any) have the legal right to review your research records and will protect the **confidentiality** of those records to the extent permitted by law. All publications will exclude any information that will make it possible to identify you as a subject. Throughout the study, the researchers will notify you of new information that may become available and that might affect your decision to remain in the study.

Contacts and Questions:

If you have any questions about the study please ask now. If you have questions later, want additional information, or wish to withdraw your participation call the researchers conducting the study. Their names, phone numbers, and e-mail addresses are at the top of this page. If you have questions about your rights as a research participant, complaints, concerns, or questions about the research please contact Lisa Leiden, Ph.D., Chair of The University of Texas at Austin Institutional Review Board for the Protection of Human Subjects, (512) 471-8871 or email: orsc@uts.cc.utexas.edu.

You will be given a copy of this information to keep for your records.

Statement of Consent:

I have read the above information and have sufficient information to make a decision about participating in this study. I consent to participate in the study.

Signature : _____ Date: _____

Signature of Investigator: _____ Date: _____

Appendix V: Native Speaker Participant Biographical Questionnaire

QUESTIONNAIRE BIOGRAPHIQUE: LOCUTEUR NATIF

Ce questionnaire porte sur vos expériences linguistiques tout au long de votre vie. Toutes les réponses sont confidentielles et votre anonymat sera strictement respecté. Votre nom et votre adresse email nous pourront être utiles au cas où il nous faudra de plus amples informations. S'il y a des questions auxquelles vous préféreriez ne pas répondre, vous pouvez les sauter.

Nom et prénom: _____
Adresse email: _____
Date de naissance: _____
Lieu de naissance: _____
Sexe: M___ F___

1) Quelles langues est-ce vous parlez et depuis combien de temps?

2) Combien de temps avez-vous passé aux pays anglophones ? Merci d'indiquer le(s) pays et les dates du séjour.

3) Combien de temps avez-vous passé aux pays francophones ? Merci d'indiquer le(s) pays et les dates du séjour.

MERCI DE VOTRE PARTICIPATION !

Appendix VI: Experimental Instructions: Experiment 1

Ax Discrimination Task Employing Real-Word Stimuli

(An English translation follows text.)

Bienvenue.

Dans la première partie de l'expérience vous allez entendre des groupes de mots.

Votre tâche consiste à décider si ces deux groupes de mots sont identiques ou différents.
Appuyez sur la touche 1 si les deux groupes de mots que vous venez d'entendre sont identiques
ou sur la touche 2 si les deux groupes de mots sont différents.

Appuyez sur la barre d'espace dès que vous serez prêt à commencer l'entraînement.

Très bien!

C'est la fin de l'entraînement.

Rappel: appuyez sur la touche 1 si les groupes de mots sont identiques
ou sur la touche 2 si les groupes de mots sont différents.

Appuyez sur la barre d'espace dès que vous serez prêt à commencer l'expérience.

Très bien!

C'est la fin de la première partie de l'expérience.

Veillez prendre quelques minutes pour vous reposer avant de commencer la deuxième partie.

Appuyez sur la barre d'espace quand vous serez prêt à continuer.

ENGLISH TRANSLATION:

Welcome.

In the first part of the experiment, you are going to hear pairs of phrases in French.

Your task is to decide whether the two phrases are identical or different.

Press 1 on the keyboard if the phrases are identical or 2 if the phrases are different.

Press the space bar to when you are ready to begin the training portion of the experiment.

Very good!

That's the end of the training portion.

Reminder: Press 1 on the keyboard if the phrases are identical or 2 if the phrases are different.

Press the space bar when you are ready to begin the experiment.

Very good!

That's the end of the first part of the experiment.

Please take a few minutes to rest before beginning the second part of the experiment.

Press the space bar when you are ready to continue.

Appendix VII: Experimental Instructions: Experiment 1.1

Ax Discrimination Task Employing Non-Word Stimuli

(An English translation follows text.)

Bienvenue.

Dans la première partie de l'expérience vous allez entendre des groupes de mots. Le dernier mot de chacun de ces groupes n'existe pas en français.

Votre tâche consiste à décider si ces deux groupes de mots sont identiques ou différents. Appuyez sur la touche 1 si les deux groupes de mots que vous venez d'entendre sont identiques

ou sur la touche 2 si les deux groupes de mots sont différents.

Appuyez sur la barre d'espace dès que vous serez prêt à commencer l'entraînement.

Très bien!

C'est la fin de l'entraînement.

Rappel: appuyez sur la touche 1 si les groupes de mots sont identiques

ou sur la touche 2 si les groupes de mots sont différents.

Appuyez sur la barre d'espace dès que vous serez prêt à commencer l'expérience.

Très bien!

C'est la fin de la première partie de l'expérience.

Veillez prendre quelques minutes pour vous reposer avant de commencer la deuxième partie.

Appuyez sur la barre d'espace quand vous serez prêt à continuer.

ENGLISH TRANSLATION:

Welcome.

In the first part of the experiment, you are going to hear pairs of phrases in French. The last word in each phrase is not a real word.

Your task is to decide whether the two phrases are identical or different.

Press 1 on the keyboard if the phrases are identical or 2 if the phrases are different.

Press the space bar to when you are ready to begin the training portion of the experiment.

Very good!

That's the end of the training portion.

Reminder: Press 1 on the keyboard if the phrases are identical or 2 if the phrases are different.

Press the space bar when you are ready to begin the experiment.

Very good!

That's the end of the first part of the experiment.

Please take a few minutes to rest before beginning the second part of the experiment.

Press the space bar when you are ready to continue.

Appendix VIII: Experimental Instructions: Experiment 1.2

Ax Discrimination Task Employing Non-Word Stimuli Tested On Beginning Learners

Welcome.

In this experiment, you are going to hear pairs of phrases in French.

Your task is to decide whether the two phrases are identical or different.

Press 1 on the keyboard if the phrases are identical or 2 if the phrases are different.

Press the space bar to continue.

You will complete a short training portion before starting the actual experiment.

Reminder: Press 1 on the keyboard if the phrases are identical or 2 if the phrases are different.

Press the space bar to begin the training portion of the experiment.

Well done!

That's the end of the training portion.

Press the space bar when you are ready to begin the experiment.

Well done!

That's the end of the experiment.

Thank you for your participation.

Appendix IX: Experimental Instructions: Experiment 2

Forced-choice identification task employing real-word stimuli

(An English translation follows text.)

Dans la deuxième partie de l'expérience, vous allez entendre un groupe de mots. Tout de suite après vous verrez une paire de mots.

Votre tâche consiste à choisir lequel de ces deux mots vous venez d'entendre.

Appuyez sur la barre d'espace pour continuer.

Par exemple, vous allez entendre un groupe de mots et tout de suite après vous verrez une paire de mots comme la suivante:

(1) termite (2) hermite

Appuyez sur la touche 1 si le mot que vous venez d'entendre est le mot de gauche ou sur la touche 2 si le mot que vous venez d'entendre est le mot de droite.

Appuyez sur la barre d'espace dès que vous serez prêt à commencer l'entraînement.

Très bien!

C'est la fin de l'entraînement.

Rappel: Appuyez sur la touche 1 si le mot que vous venez d'entendre est le mot de gauche ou sur la touche 2 si le mot que vous venez d'entendre est le mot de droite.

Appuyez sur la barre d'espace dès que vous serez prêt à commencer l'expérience.

C'est la fin de l'expérience.

Merci d'avoir participé.

Au revoir.

ENGLISH TRANSLATION

Welcome.

In the second part of the experiment, you are going to hear a phrase in French. Immediately afterwards you will see a pair of words.

Your task is to choose which of the two words you have just heard.

Press the space bar to continue.

For example, you will hear a phrase and immediately afterwards you will see pairs of words such as the following:

(1) termite (2) hermite

Press the 1 key if you have just heard the word on the left or the 2 key if you have just heard the word on the right.

Press the space bar as soon as you are ready to begin the training portion of the experiment.

Very good!

That's the end of the training portion.

Reminder: Press the 1 key if you have just heard the word on the left or the 2 key if you have just heard the word on the right.

Press the space bar when you are ready to begin the experiment.

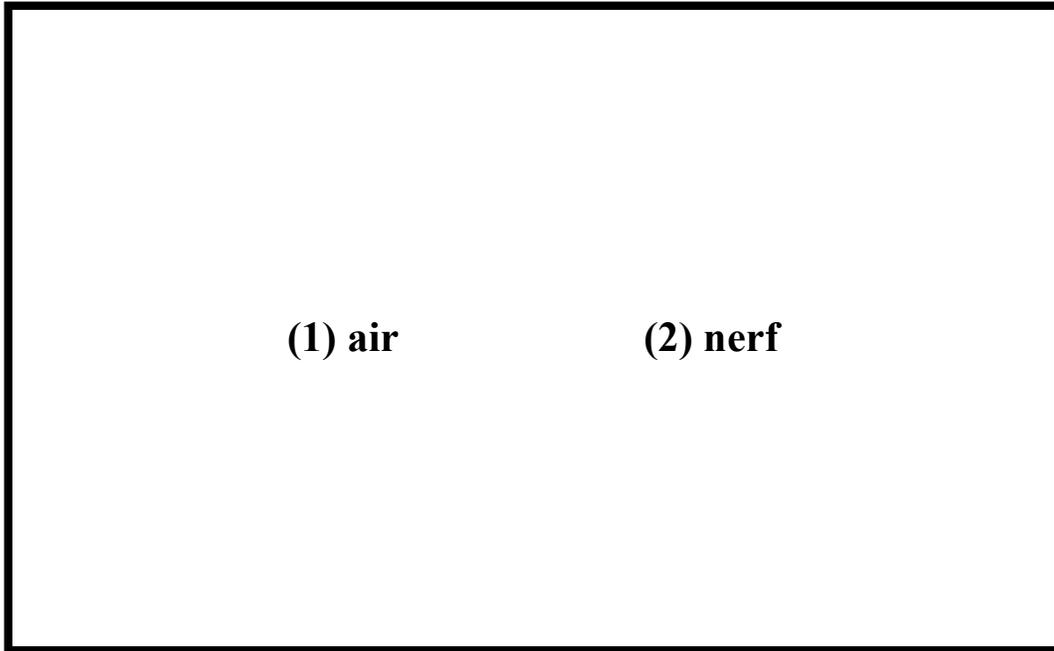
That is the end of the experiment.

Thank you for participating

Goodbye.

Appendix X: Sample Presentation Screen: Experiment 2

Forced-Choice Identification Task Employing Real-Word Stimuli



Appendix XI: Experimental Instructions: Experiment 2.1

Forced-choice identification task employing non-word stimuli

(An English translation follows text.)

Dans la deuxième partie de l'expérience, vous allez entendre un groupe de mots, dont le dernier mot n'existe pas en français. Tout de suite après vous verrez une paire de non-mots.

Votre tâche consiste à choisir lequel de ces deux non-mots vous venez d'entendre.

Appuyez sur la barre d'espace pour continuer.

Par exemple, vous allez entendre un groupe de mots et tout de suite après vous verrez une paire de non-mots comme la suivante:

(1) nuteau (2) uteau

Appuyez sur la touche 1 si le mot que vous venez d'entendre est le mot de gauche

ou sur la touche 2 si le mot que vous venez d'entendre est le mot de droite.

Appuyez sur la barre d'espace dès que vous serez prêt à commencer l'entraînement.

Très bien!

C'est la fin de l'entraînement.

Rappel: Appuyez sur la touche 1 si le mot que vous venez d'entendre est le mot de gauche

ou sur la touche 2 si le mot que vous venez d'entendre est le mot de droite.

Appuyez sur la barre d'espace dès que vous serez prêt à commencer l'expérience.

C'est la fin de l'expérience.

Merci d'avoir participé.

Au revoir.

ENGLISH TRANSLATION

Welcome.

In the second part of the experiment, you are going to hear a phrase in French. The last word in this phrase is not a real French word. Immediately afterwards you will see a pair of non-words.

Your task is to choose which of the two non-words you have just heard. Press the space bar to continue.

For example, you will hear a phrase and immediately afterwards you will see pairs of non-words such as the following:

(1) nuteau (2) uteau

Press the 1 key if you have just heard the word on the left or the 2 key if you have just heard the word on the right.

Press the space bar as soon as you are ready to begin the training portion of the experiment.

Very good!

That's the end of the training portion.

Reminder: Press the 1 key if you have just heard the word on the left or the 2 key if you have just heard the word on the right.

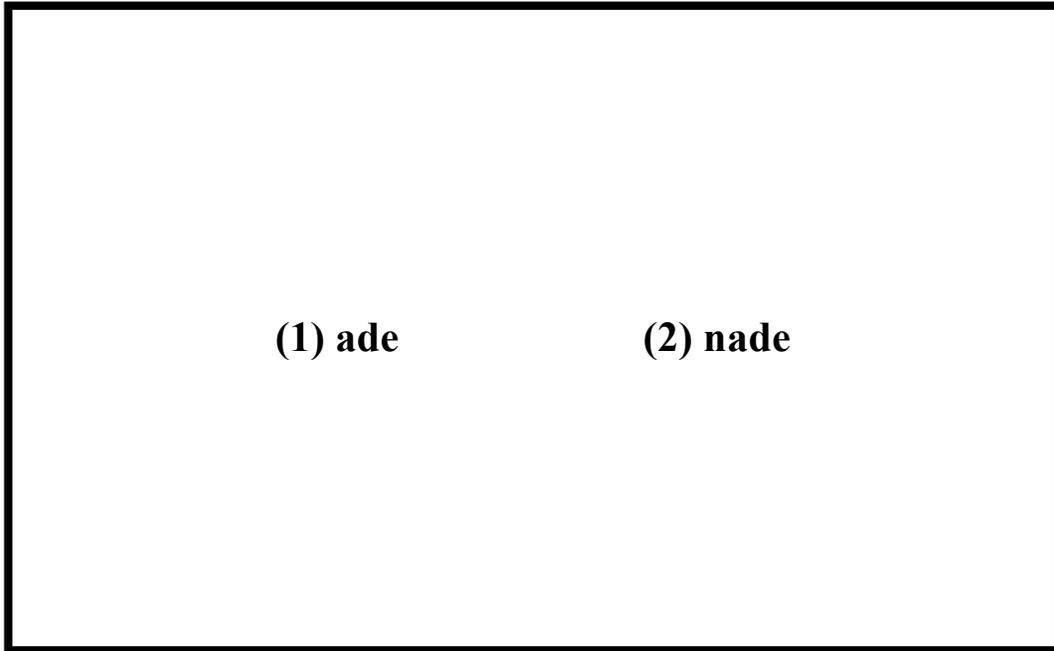
Press the space bar when you are ready to begin the experiment.

That is the end of the experiment.

Thank you for participating. Goodbye.

Appendix XII: Sample Presentation Screen: Experiment 2.1

Forced-Choice Identification Task Employing Non-Word Stimuli



References

- Adda-Decker, M., Boula de Mareüil, P., Adda, G. & Lamel, L. (2002). Investigating syllabic structure and its variation in speech from French radio interviews. In *PMLA-2002*, 89-94.
- Agren, J. (1973). *Etude sur quelques liaisons facultatives dans le français de conversation radiophonique*. Uppsala: Uppsala University Press.
- Allen, J. & Miller, J. (2001). Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception & Psychophysics*, 63, 798-810.
- Allopenna, D., Magnuson, S. & Tanenhaus, M. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419-439.
- Altenberg, E. (2005). The perception of word boundaries in a second language. *Second Language Research*, 21 (4), 325–358.
- Andruski, J., Blumstein, S. & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52(3), 163-187.
- Archibald, J. (2009). Second Language Phonology. In W. Ritchie & T. Bhatia (Eds.), *The new handbook of Second Language Acquisition*. Amsterdam: Elsevier.
- Aslin, R., Pisoni, D., Hennessy, B. & Perey, A. (1981). Discrimination of voice onset time by human infants: new findings and implications for the effects of early experience. *Child Development*, 52, 1135–1145.
- Aslin, R., Saffran, J. & Newport, E. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 321-324.
- Banel, M-H. & Bacri, N. (1994). On metrical patterns and lexical parsing in French. *Speech Communication*, 15, 115–126.
- Basset, E. (2000). *La liaison à 3, 7 et 11 ans: description et acquisition*. Unpublished *mémoire de maîtrise*, Université de Grenoble 3, Grenoble, France.
- Beckman, M. & Edwards, J. (1990). Lengthening and shortening and the nature of prosodic constituency. In J. Kingston and M. E. Beckman (Eds.), *Laboratory Phonology I*. (pp. 152-178). Cambridge University Press, Cambridge.
- Best, C. (1995). A direct realist view of cross-language speech perception: New Directions in Research and Theory. In Winifred Strange (ed.) *Speech perception and linguistic experience: Theoretical and methodological issues*. (pp. 171–204). Baltimore: York Press.

- Best, C., McRoberts, G. & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, 109, 775-794.
- Best, C., McRoberts, G. & Sithole, N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 345-360.
- Best, C. & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. Munro (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34). Amsterdam: John Benjamins.
- Bever, T. (1981). Normal acquisition processes explain the critical period for language learning. In K. Diller (Ed.), *Individual Differences in Language Learning Aptitude* (pp. 176-198). Rowley, MA: Newbury House.
- Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, 68, 706-755.
- Birdsong, D. (1999). Introduction: Whys and why nots of the critical period hypothesis. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis*. (pp. 1-22). Mahwah, NJ: Erlbaum.
- Birdsong, D. (2003). Authenticité de prononciation en français L2 chez des apprenants tardifs Anglophones: Analyses segmentales et globales. *Acquisition et Interaction en Langue Etrangère*, 18, 17-36.
- Birdsong, D. (2004). Second language acquisition and ultimate attainment. In A. Davies and C. Elder (Eds.), *Handbook of Applied Linguistics* (pp. 82-105) London: Blackwell.
- Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 109-127). New York: Oxford University Press.
- Birdsong, D. (2006). Age and second language acquisition and processing: A Selective Overview. *Language Learning*, 56(s1), 9-49.
- Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In O.-S. Bohn and M. Munro (Eds.) *Second language speech learning: The role of language experience in speech perception and production*. Amsterdam: John Benjamins.
- Birdsong, D. (2009). Age and the end state of second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *The new handbook of second language acquisition*. Amsterdam: Elsevier.
- Birdsong, D. & Molis, M. (2001). On the evidence for maturational effects in second language acquisition. *Journal of Memory and Language*, 44, 235-249.

- Birdsong, D. & Paik, J. (2008). Second language acquisition and ultimate attainment. In B. Spolsky & F. Hult (Eds.), *Handbook of educational linguistics*. London: Blackwell.
- Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. Gass & J. Schachter (Eds.), *Linguistic Perspectives on Second Language Acquisition* (pp. 41-68). Cambridge University Press, Cambridge UK.
- Bloom, Lois. (1993). *The Transition from Infancy to Language: Acquiring the power of expression*. Cambridge, England and New York: Cambridge University Press.
- Blumenfeld, H. & Marian, V. (2005). Covert bilingual language activation through cognate word processing: An eye-tracking study. *Proceedings of the XXVII Annual Meeting of the Cognitive Science Society* (pp. 286-291). Mahwah, NJ: Lawrence Erlbaum.
- Boersma, P. & Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.6.36) [Computer program]. Retrieved May 12, 2007 from <http://www.praat.org/>
- Bohn, O.-S. & Flege, J.E. (1993). Perceptual switching in Spanish/English bilinguals: Evidence for universal factors in stop voicing judgments. *Journal of Phonetics*, 21, 267-290.
- Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.) *Second language acquisition and the critical period hypothesis*. Mahwah, NJ: Lawrence Erlbaum.
- Bongaerts, T., Mennen, S. & van der Slik, F. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica*, 54 (2), 298-308.
- Bongaerts, T., Planken, B. & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Second Language Research*, 19, 447-465.
- Bongaerts, T., van Summeren, C., Planken, B. & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition* 19, 447-465.
- Boomer, D. & Laver, J. (1968). Slips of the tongue. *British Journal of Disorders of Communication*, 3(1), 2-12.
- Boucher, V. (1988). A parameter of syllabification for VstopV and relative-timing invariance. *Journal of Phonetics*, 16, 299-326.
- Bowden, H., Sanz, C. & Stafford, C. (2005). Individual differences: Age, sex, working memory, and prior knowledge. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: methods, theory, and practice* (pp. 105-140). Georgetown University Press.

- Bradley, D., Sanchez-Casas, R. & Garcia-Albea, J. (1993). The status of the syllable in the perception of English and Spanish. *Language and Cognitive Processes*, 8, 197-233.
- Bradlow, A., Nygaard, L. & Pisoni, D. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, 61, 206–219.
- Bradlow, A., Pisoni, D., Akahane-Yamada, R. & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101 (4), 2299-2310.
- Bybee, J. (2001a). *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, J. (2001b). Frequency effects on French liaison. In J. Bybee & P. Hopper, (Eds.) *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins.
- Bybee, J. (2005). La liaison: Effets de fréquence et constructions. *Langages*, 158, 24-37.
- Byrd, D. (1994). *Articulatory timing in English consonant sequences*. Doctoral dissertation, UCLA (distributed as *UCLA Working Papers in Phonetics*, Vol. 86, pp. 1-196).
- Caramazza, A. & Yeni-Komshian, G. (1974). Voice onset time in two French dialects, *Journal of Phonetics*, 2, 239-245.
- Cairns, P., Shillcock, R., Chater, N. & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus based approach to speech segmentation. *Cognitive Psychology*, 33, 111-153.
- Carroll, S. (1999). Putting ‘input’ in its proper place. *Second Language Research*, 15, 337–388.
- Chevrot, J.P & Fayol, M. (2001). Acquisition of French liaison and related child errors. In M. Almgren, A. Barrena, M. J. Ezeizabarrena, I. Idiazabal & B. MacWhinney (Eds.), *Proceedings of the 8th Conference of the International Association for the Study of Child Language*. (pp. -760-774). Cascadilla Press.
- Cho, T. (2007). Prosodic strengthening in speech production and perception: The current issues. *Speech Sciences*, 14(4), 7-24.
- Cho, T. & Keating, P. (2001). Articulatory and acoustic studies on domain- initial strengthening in Korean. *Journal of Phonetics*, 29, 155-190.
- Cho, T., McQueen, J. & Cox, E. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, 35(2), 210-243.
- Christie, W. (1977). Some multiple cues for juncture in English. *General Linguistics*, 17, 212-222.

- Christophe, A., Dupoux, E., Bertoncini, J. & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, 95, 1570-1580.
- Christophe, A., Peperkamp, S., Pallier, C., Block, E. & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language*, 51, 523-547.
- Clahsen, H. & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, 27, 3-42.
- Clarke, C. & Garrett, M. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of American*, 116 (6), 3647–3658.
- Cluff, M. & Luce, P. (1990). Similarity neighborhoods of spoken two-syllable words: retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance*, 16(3), 551-63.
- Coenen, E., Zwitserlood, P. & Bölte, J. (2001). Variation and assimilation in German: Consequences of assimilation for word recognition and lexical representation. *Language and Cognitive Processes*, 16, 535 - 564.
- Connine, C., Blasko, D. & Wang, J. (1994). Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences and the role of sentence context. *Perception & Psychophysics*, 56, 624–636.
- Connine, C., Titone, D., Deelman, T. & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, 37, 463–480.
- Content, A., Bagou, O., Frauenfelder, U. & Fougeron, C. (2004). Indices de frontières lexicales avec des pseudo-mots: une étude préliminaire. *Actes des 25èmes Journées d'Etudes sur la Parole*. Fès, Maroc.
- Content, A., Dumay, N. & Frauenfelder, U. (2000). The role of syllable structure in lexical segmentation in French: Helping listeners avoid mondegreens. In A. Cutler, J.M. McQueen and R. Zondervan (Eds.) *Spoken word access processes* (pp. 39-42). Nijmegen, Netherlands.
- Content, A., Kearns, R. & Frauenfelder, U. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language*, 45, 177-199.
- Content, A., Meunier, C. Kearns, R. & Frauenfelder, U. (2001). Sequence detection in pseudo-words in French: Where is the syllable effect? *Language and Cognitive Processes*, 16(5/6), 609–636.
- Côté, M. (2005). Le statut lexical des consonnes de liaison. *Langages*, 158, 66-78.

- Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguists and computational perspectives* (pp. 105-121). Cambridge, MA: MIT Press.
- Cutler, A. (2001). Listening to a second language through the ears of a first. *Interpreting*, 5, 1-23.
- Cutler, A. (2002). Native listeners. *European Review*, 10, 27-41.
- Cutler, A. (2005). Lexical statistics of competition in L2 versus L1 listening. *Journal of the Acoustical Society of America*, 118 (3), 1901.
- Cutler, A. & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, 31, 218-236.
- Cutler, A. & Carter, D. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133-142.
- Cutler, A., Dahan, D. & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40 (2), 141-202.
- Cutler, A., Demuth, K. & McQueen, J. (2002). Universality versus Language Specificity in Listening to Running Speech. *Psychological Science*, 13, 258-262.
- Cutler, A., McQueen, J. & Suomi, K. (1997). Finnish vowel harmony and speech segmentation. *Max Planck Institute Annual Report*, 9.
- Cutler, A., Mehler, J., Norris, D. & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25, 385-400.
- Cutler, A., Mehler, J., Norris, D. & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology*, 19, 141-177.
- Cutler, A., Mehler, J., Norris, D. & Segui, J. (1989). Limits of bilingualism. *Nature*, 340, 229-230.
- Cutler, A., Mehler, J., Norris, D. & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals, *Cognitive Psychology*, 24, 381-410.
- Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 113-121.
- Cutler, A. & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, 33, 824-844.
- Cynader, M., Timney, B. & Mitchell, D. (1980). Period of susceptibility of kitten visual cortex to the effects of monocular deprivation extends beyond six months of age. *Brain Research*, 191, 545-550.

- Dahan, D., Magnuson, J. & Tanenhaus, M. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, 42, 317-367.
- Dahan, D., Swingle, D., Tanenhaus, M. & Magnuson, J. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, 42, 465-480.
- Dahan, D. & Tanenhaus, M. (2004). Continuous mapping from sound to meaning in spoken language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 498-513.
- Darcy, I., Peperkamp, S. & Dupoux, E. (2007). Plasticity in compensation for phonological variation: the case of late second language learners. In J. Cole & J. Hualde (Eds.) *Laboratory Phonology*, 9, Mouton de Gruyter, Berlin.
- Davies, A. (2004). The native speaker in applied linguistics. In Davies, A., & Elder, C. (Eds.), *Handbook of Applied Linguistics*. New York: Blackwell.
- Davis, M., Marslen-Wilson, W. & Gaskell, M. (2002). Leading up the lexical garden-path: segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 218-244.
- Dejean de la Bâtie, B. (1993). Word boundary ambiguity in spoken French. Unpublished doctoral dissertation. Monash University, Victoria, Australia.
- Dejean de la Bâtie, B. & Bradley, D. (1995). Resolving word boundaries in spoken French: Native and non-native strategies. *Applied Psycholinguistics*, 16, 59-81.
- Delattre, P. (1940). Le mot est-il une unité phonétique en français ? *Le Français Moderne*, 8(1), 47-56.
- Delattre, P. (1947). La liaison en français, tendances et classification. *The French Review*, 21, 148-157.
- Delattre, P. (1951). Principes de phonétique française: à l'usage des étudiants anglo-américains. Middlebury, Vermont: Ecole française d'été, 2 ed.
- Delattre, P. (1965). *Comparing the phonetic features of English, German, Spanish and French*. Heidelberg: Julius Groos Verlag.
- Delattre, P. (1966). *Studies in French and comparative phonetics*. The Hague: Mouton & Co.
- Dell, F. & Vergnaud, J.-R. (1984). Les développements récents en phonologie: Quelques idées centrales. In F. Dell, D. Hirst & J.-R. Vergnaud (Eds.), *Forme sonore du langage* (pp. 1-42). Paris: Hermann.
- Diehl, R., Souther, A. & Convis, C. (1980). Conditions on rate normalization in speech perception. *Perception & Psychophysics*, 27, 435-443.

- Dijkstra, A., Van Jaarsveld, H. & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition*, 1 (1), 51-66.
- Douchez, C. & Lancia, L. (2008). Caractéristiques articulatoires des consonnes de liaison: Etude pilote. *Journées d'Etudes sur la Parole*, Avignon, France.
- Dugua, C. (2002). Liaison et segmentation du lexique français: vers un scénario développemental. Unpublished Diplôme d'Etudes Approfondies, Université Grenoble 3, Grenoble, France.
- Dumay, N., Banel, M., Frauenfelder, U. & Content, A. (1998). Le rôle de la syllabe: Segmentation lexicale ou classification? Actes des XXIIème Journées d'Etude sur la Parole, (pp. 33-36). Martigny, Switzerland.
- Dumay, N., Content, A. & Frauenfelder, U. (1999). Acoustic-phonetic cues to word boundary location: Evidence from word-spotting. *Proceedings of the fourteenth international congress of phonetic sciences* (pp. 281-284). Berkeley, CA: University of California.
- Dupoux E. & Green K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 914-927.
- Dupoux, E., Pallier, C., Kakehi, K. & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes*, 5(16), 491-505.
- Dupoux, E., Pallier, C., Sebastián-Gallés, N. & Mehler, J. (1997). A distressing 'deafness' in French? *Journal of Memory and Language*, 36, 406-421.
- Dupoux, E., Sebastian-Galles, N., Navarete, E. & Peperkamp, S. (2008). Persistent stress 'deafness': the case of French learners of Spanish. *Cognition*, 106(2), 682-706.
- Durand, J., Laks, B. & Lyche, C. (2002). La phonologie du français contemporain: usages, variétés et structure. In C. Pusch & W. Raible (Eds.), *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance corpus Linguistics - Corpora and spoken language* (pp. 93-106). Tübingen: Gunter Narr Verlag.
- Durand, J., Laks, B. & Lyche, C. (2005). Un corpus numérisé pour la phonologie du français. In G. Williams (Ed.), *La linguistique de corpus* (pp. 205-217). Rennes: Presses Universitaires de Rennes
- Durand, J & Lyche, C. (2008). French liaison in the light of corpus data. *French Language Studies*, 18, 33-66.
- Eckman, F. (1977). Markedness and the Contrastive Analysis Hypothesis. *Language Learning*, 27(2), 315-330.

- Eimas, P. D. (1975). Speech perception in early infancy. In L. B. Cohen & P. Salapatek (Eds.), *Infant Perception, 2: From Sensation to Cognition*. New York: Academic Press.
- Eimas, P., Hornstein, S. & Payton, P. (1990). Attention and the role of dual codes in phoneme monitoring. *Journal of Memory and Language, 29*, 160–180.
- Eimas P., Siqueland E., Jusczyk P. & Vogorito J. (1971). Speech perception in infants. *Science, 171*, 303-306.
- Eisner, F. & McQueen, J. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67*, 224-238.
- Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge MA: MIT Press.
- Elman, J., Diehl, R. & Buchwald, S. (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America, 62*, 971–974.
- Encrevé, P. (1988). *La liaison avec et sans enchaînement*. Paris: le Seuil.
- Encrevé, P. & Scheer, T. (2005). *Autosegmental association is not automatic*. 13th Phonology Meeting, Manchester.
- Flege, J. (1987). The production of 'new' and 'similar' phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47-65.
- Flege, J. (1991). Perception and Production: The Relevance of Phonetic Input to L2 Phonological Learning. In Hueber, T. & Ferguson, C. (Eds.) *Crosscurrents in Second Language Acquisition and Linguistic Theories*. (pp. 249-289). Amsterdam: John Benjamins.
- Flege, J. (1992). Speech learning in a second language. In C. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological Development, Models, Research, and Applications* (pp. 233-273). Timonium, MD: York Press.
- Flege, J. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 233-273). Timonium, MD: York Press.
- Flege, J. (1997). English vowel production by Dutch talkers: more evidence for the “similar” vs. “new” distinction. In J. Leather & A. James (Eds.), *Second- Language Speech* (pp. 11-52). Berlin: Mouton de Gruyter.
- Flege, J. (1999). Age of Learning and Second-language Speech. In D. Birdsong (Ed.) *Second Language Acquisition and the Critical Period Hypothesis*. (pp. 101-132). Hillsdale, NJ: Lawrence Erlbaum.

- Flege, J. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske & A. Rohde (Eds.), *An Integrated View of Language Development: Papers in Honor of Henning Wode*. (pp. 217-244). Trier: Wissenschaftlicher Verlag.
- Flege, J. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.) *Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities*. (pp. 319-355). Berlin: Mouton de Gruyter.
- Flege, J., Birdsong, D., Bialystok, E., Mack, M., Sung, H. & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics*, 34, 153-175.
- Flege, J., Frieda, E. & Nozawa, T. (1997). Amount of native-language (L1 use) affects the pronunciation of an L2. *Journal of Phonetics*, 25, 169-186.
- Flege, J. & Hillenbrand, J. (1984). Limits on pronunciation accuracy in adult foreign language speech production. *Journal of the Acoustical Society of America*, 76, 708-721.
- Flege, J. & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, 23, 527-552.
- Flege, J. & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26, 1-34.
- Flege, J., MacKay, I. & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, 106, 2973-2987.
- Flege, J., MacKay, I. & Piske T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, 23, 567-598.
- Flege, J. & Munro, M. (1994). The word unit in L2 speech production and perception. *Studies in Second Language Acquisition*, 16, 381-411.
- Flege, J. Yeni-Komshian, G & Liu, S. (1999). Age constraints on second language acquisition. *Journal of Memory and Language*, 41, 78-104.
- Floccia, C. & Bertoncini, J. (1993). From initial speech perception to language-specific processing: An overview of cross-linguistic research. In J., Altarriba (Eds.) *Cognition and culture: A cross cultural approach to cognitive psychology*. (pp. 265-279). Amsterdam, Netherlands: North-Holland / Elsevier Science Publishers.
- Floccia, C., Goslin, J., Girard, F. & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, 32, 1276-1293.

- Fodor, J. (1983). *The Modularity of mind: an essay on faculty psychology*. Cambridge, Mass.: MIT Press.
- Forster, K. (1976). Accessing the mental lexicon. In R. J. Wales & E. W. Walker (Eds.), *New approaches to language mechanisms*. Amsterdam: North-Holland.
- Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, 29, 109-135.
- Fougeron, C. (2007). Word boundaries and contrast neutralization in the case of *enchaînement* in French. In Cole, J. & Hualde J. (Eds.) *Papers in Laboratory Phonology IX: Change in Phonology*. (pp. 609-642). Berlin: Mouton de Gruyter.
- Fougeron, C. & Keating, P. (1997). Articulatory strengthening at edges of prosodic domain. *Journal of the Acoustical Society of America*, 106, 3728-3740.
- Fougeron, C., Bagou, O., Stefanuto, M. & Frauenfelder, U. (2002). A la recherche d'indices de frontière lexicale dans la resyllabation. *Actes des XXIVèmes Journées d'Etude sur la Parole*, Nancy (pp. 125-128).
- Fougeron, C., Bagou, O., Stefanuto, M., Frauenfelder, U. (2003). Looking for acoustic cues of resyllabification in French. *15th International Congress of Phonetic Sciences*, Barcelona.
- Fox, R. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 526-540.
- Frauenfelder, U. & Floccia, C. (1999). The Recognition of Spoken Words. In A. Friederici (Ed.) *Language Comprehension: A Biological Perspective*. (pp. 1-41.) Heidelberg: Springer-Verlag Berlin.
- Frauenfelder, U., Scholten, M. & Content, A. (2001). Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes*, 16, 583-607.
- Frauenfelder, U., Segui, J. & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 77-91.
- Frauenfelder, U. & Tyler, L. (1987). The process of spoken word recognition: An introduction. *Cognition*, 25, 1-20.
- Fujisaki, H., Nakamura, K. & Imoto, T. (1975). Auditory Perception of Duration of Speech and Non-speech Stimuli. In G. Fant & M. Tatham, (Eds.) *Auditory Analysis and Perception of Speech*. (pp. 197-220). New York: Academic.
- Ganong, W. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance*, 6(1), 110-125.

- Gaskell, M. & Marslen-Wilson, W. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 144-158.
- Gaskell, M. & Marslen-Wilson, W. (1997). Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes*, 12, 613-656.
- Gaskell, M. & Marslen-Wilson, W. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 380-396.
- Gaskell, M. & Marslen-Wilson, W. (2001). Lexical ambiguity and spoken word recognition: bridging the gap. *Journal of Memory and Language*, 44, 325-349.
- Gaskell, G., Spinelli, E. & Meunier, F. (2002). Perception of Resyllabification in French. *Memory and Cognition*, 30, 798-810.
- Gerard, L. & Scarborough, D. (1989). Language-specific lexical access of homographs by bilinguals. *Journal of Experimental Psychology: Learning Memory and Cognition*, 15, 305-313.
- Girard, F., Goslin, J. & Floccia, C. (2004). Un accent régional gêne-t-il la compréhension? *Proceedings of the workshop MIDL*, Paris.
- Girard, F., Goslin, J. & Floccia, C. (2005). Influence of regional accents in speech perception. *Proceedings of Context '05*, Paris.
- Goetry, V. & Kolinsky, R. (2000). The role of rhythmic cues for speech segmentation in monolingual and bilingual listeners. *Psychologica Belgica*, 40, 115-52.
- Golato, P. (2002). Word parsing by late-learning French-English bilinguals. *Applied Psycholinguistics*, 23(3), 417-446.
- Goldinger, S. (1992). Words and voices: Implicit and explicit memory for spoken words. Unpublished doctoral dissertation, Indiana University.
- Goldinger, S. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166-1183.
- Goldinger, S. & Azuma, T. (2003). Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics*, 31, 305-320.
- Goldinger, S., Luce, P. & Pisoni, D. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, 28, 501-518.

- Gough, P. & Cosky, M. (1977). One second of reading again. In N. Castellan, D. Pisoni & G. Potts (Eds.), *Cognitive theory* (Vol. 2). Hillsdale, NJ: Erlbaum.
- Gow, D. & Gordon, P. (1993). The distinctiveness of word onsets. Paper presented at the meeting of the Acoustical Society of America, Ottawa, Canada.
- Gow, D. & Gordon, P. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 344-359.
- Grammont, M. (1938/1960). *Traité pratique de prononciation française*. Paris: Delagrave.
- Greenspan, S., Nusbaum, H. & Pisoni, D. (1988). Perception of synthetic speech produced by rule: Intelligibility of eight text-to-speech systems. *Behavioral Research Methods, Instruments, and Computers*, 18, 100-107.
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*, 28, 267-283.
- Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Processes*, 3(3), 233-274.
- Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, 1, 131-149.
- Grosjean, F. & Gee, J. (1987). Prosodic structure and spoken word recognition. *Cognition*, 25, 135-155.
- Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, 31, 423-445.
- Grossberg, S. & Myers, C. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review*, 107, 735-767.
- Gussenhoven, C. (1986). English plosive allophones and ambisyllabicity. *Gramma*, 10(2), 119-41.
- Hannahs, S. (2007). French phonology and L2 acquisition. In Dalila Ayoun, ed. *French Applied Linguistics*. (pp. 50-74). Amsterdam: John Benjamins.
- Halle, P., Best, C. & Levitt, A. (1999). The perception of American English approximants by French native speakers. *Proceedings of the XIVth International Congress of Phonetic Sciences*, (pp. 905-908). San Francisco, CA.
- Harley, T. (2001). *The Psychology of Language: From Data to Theory*. Psychology Press, East Sussex.

- Harrington, J., Watson, G. & Cooper, M. (1989). Word boundary detection in broad class and phoneme strings. *Computer Speech and Language*, 3, 367-382.
- Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, 31, 373-405
- Hazan, V. & Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech*, 36, 17-38.
- Healy, A. F. (1976). Detection errors on the word THE: Evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception & Performance*, 2, 235-242.
- Hoard, J. (1966). Juncture and syllable structure in English. *Phonetica*, 15, 19-109.
- Holt, L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Review*, 16 (4), 305-312.
- Huggins, A. (1972). Just noticeable differences for segment durations in natural speech. *Journal of the Acoustical Society of America*, 51, 1270-1278.
- Hyltenstam, K. & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty and M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539-588). Malden, MA and Oxford, UK: Blackwell.
- Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A. & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, B47-B57.
- Jakobson, R. & Waugh, L. (1979). *The Sound Shape of Language*. Bloomington, Ind: Indiana University Press and London: Harvester.
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145-165). San Diego: Academic Press.
- Johnson, J. & Newport, L. (1989). Critical period effects in second language acquisition: the influence of maturational stage on the acquisition of ESL. *Cognitive Psychology*, 21, 60-99.
- Ju, M. & Luce, P. (2004). Falling on sensitive ears – Constraints on bilingual lexical activation. *Psychological Science*, 15, 314-18.
- Jusczyk, P. (1993). From general to language specific capacities: the WRAPSA model of how speech perception develops. *Journal of Phonetics*, 21, 3-28.
- Jusczyk, P. (1997). *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

- Jusczyk, P., Cutler, A. & Redanz, N. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64, 675-687.
- Jusczyk, P. & Luce, P. (2002). Speech Perception and Spoken Word Recognition: Past and Present. *Ear and Hearing*, 23, 2-40.
- Jusczyk, P., Rosner, B., Cutting, J., Foard, F., and Smith, L. (1977). Categorical perception of nonspeech sounds by 2-month-old infants. *Perception & Psychophysics*, 21(1), 50-54.
- Kahn, D. (1980). Syllable-based generalizations in English phonology. New York: Garland.
- Kawai, N. & Carrell, T. (2005). Discrimination of phoneme length differences in word and sentence contexts. *Journal of the Acoustical Society of America*, 118 (3), 2033.
- Keating, P., Cho, T., Fougeron, C. & Hsu, C.-S. (2003). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden & R. Temple (Eds.), *Papers in laboratory phonology VI* (pp. 144-162). Cambridge, UK: Cambridge University Press.
- Klatt, D. (1973). Interaction between two factors that influence vowel duration. *Journal of the Acoustical Society of America*, 54 (4), 1102-1104.
- Klatt, D. (1974). The duration of [S] in English words. *Journal of Speech Hearing Research*, 17, 51-63.
- Klatt, D. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 18, 686-706.
- Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59 (5), 1208-1221.
- Klatt, D. (1980). Speech Perception: A Model of Acoustic-Phonetic Analysis and Lexical Access. In R. Cole (Ed.) *Production and Perception of Fluent Speech*. Hillsdale, NJ: Lawrence Erlbaum.
- Klatt, D. (1981). Lexical Representations for Speech Production and Perception. In T. Myers, J. Laver & J. Anderson (Eds.), *The Cognitive Representation of Speech* (pp. 11-31). Amsterdam: North-Holland Publishing Company.
- Klatt, D. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.) *Lexical Representation and Process* (pp. 189-226). Cambridge, MA: MIT Press.
- Klatt, D. & Cooper, W. (1975). Perception of Segment Duration in Sentence Contexts. In A. Cohen & S. G. (Eds.) *Structure and Process in Speech Perception*. Berlin: Nootboom Springer-Verlag.
- Klein, W. (1995). Language acquisition at different ages. In D. Magnusson (Ed.), *The Lifespan Development of Individuals: Behavioral, Neurobiological, and Psychosocial Perspectives. A synthesis*. (pp. 244-264). Cambridge UK: Cambridge University Press.

- Kolinsky, R. (1998). Spoken word recognition: A stage processing approach to language differences. *European Journal of Cognitive Psychology*, *10*, 1–40.
- Kraljic, T. & Samuel, A. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, *51*, 141–178.
- Kuhl, P. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perceptual Psychophysics*, *50*, 93–107.
- Kuhl, P. (2000). A new view of language acquisition. *Proceedings of the National Academy of Science*, *97*, 11850-11857.
- Kuhl, P. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews: Neuroscience*, *5*, 831-843.
- Kuhl, P., Conboy, B., Padden, D., Nelson, T. & Pruitt, J. (2005). Early speech perception and later language development: Implications for the "critical period." *Language Learning and Development*, *1*, 237-264.
- Kuhl, P. & Miller, J. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, *190*, 69–72.
- Kuhl, P. & Miller, J. (1978). Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, *63*, 905–17.
- Kuhl, P., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S. & Iverson, P. (2006). Infants show facilitation for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*, 13-21.
- Lachman, R., Lachman, J. & Butterfield, E. (1979). The information-processing paradigm. In R. Lachman, J. Lachman, & E. Butterfield (Eds.), *Cognitive psychology and information processing: An introduction*. (pp. 88-129). Mahway, NJ: Lawrence Erlbaum Associates.
- Lachs, L., McMichael, K., Pisoni, D. (2003). Speech perception and implicit memory: Evidence for detailed episodic encoding of phonetic events. In J. Bowers & C. Marsolek (Eds.), *Rethinking implicit memory* (pp. 215-235). Oxford: Oxford University Press.
- Lado, R. (1957). *Languages Across Cultures*. Ann Arbor Michigan: University of Michigan Press.
- Lahiri, A. & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the mental lexicon. *Cognition*, *38*, 245-294,
- Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica, Supplementum*, *5*, 1-54.

- Lehiste, I. (1961). Acoustic studies of boundary signals. *Proceedings of 4th ICPHS* (pp. 178-187).
- Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 6(2), 2018-2024.
- Lehiste, I. (1976). Influence of fundamental frequency pattern on the perception of duration. *Journal of Phonetics*, 4, 113–117.
- Lehiste, I., Olive, J. & Streeter, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, 60 (5), 1199-1202.
- Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.
- Léon, P. (1992). *Phonétisme et prononciations du français avec des travaux d'application et leurs corrigés*. Paris: Nathan.
- Lhote, E. (1995). *Enseigner l'oral en interaction. Percevoir, écouter, comprendre*. Paris: Hachette.
- Li, P. (1996). Spoken word recognition of code-switched words by Chinese–English bilinguals. *Journal of Memory and Language*, 35, 757–774.
- Liberman, A. (1996). *Speech: A special code*. Cambridge, MA: MIT Press.
- Liberman, A., Delattre, P., Cooper, F. & Gerstman, L. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs: General & Applied*, 68, 1-13.
- Liberman, A., Harris, K., Hoffman, H. & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries, *Journal of Experimental Psychology*, 61, 379-388,
- Liberman, A. & Studdert-Kennedy, M.. (1978). Phonetic perception. In R. Held, H. Leibowitz, & H-L. Teuber (Eds.), *Handbook of Sensory Physiology, Vol. VIII, Perception* (pp.143-178). Heidelberg: Springer Verlag.
- Lively, S., Pisoni, D., Yamada, R., Tohkura, Y. & Yamada, T. (1994). Training Japanese listeners to identify English /ɾ/ and /l/ III. Long-term retention of new phonetic categories. *Journal of the Acoustic Society of America*, 96, 2076-2087.
- Local, J. (2003). Variable domains and variable relevance: interpreting phonetic exponents. *Journal of Phonetics*, 31, 321–339.
- Logan, J. & Pruitt, J. (1995). Methodological considerations in training listeners to perceive nonnative phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 351-377). Timonium, MD: York Press.

- Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, 12, 251-285.
- Lotto, A. & Holt, L. (2000). The illusion of the phoneme. In S. J. Billings, J. P. Boyle & A. M. Griffith (Eds.), *Chicago Linguistic Society, Volume 35: The Panels*. (pp. 191-204). Chicago: Chicago Linguistic Society.
- Luce, P. & Pisoni, D. (1998). Recognizing spoken words: the neighborhood activation model. *Ear & Hearing*, 19 (1), 1-36.
- MacKay, I., Meador, D. & Flege, J. (2001). The identification of English consonants by native speakers of Italian. *Phonetica*, 58, 103-125.
- Major, R. (1994). Current trends in interlanguage phonology. In M. Yavas (Ed.) *First and second language phonology*. San Diego: Singular Publishing Group.
- Mann, V. & Repp, B. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, 69, 548-558.
- Marchman, V. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience*, 5, 215-234.
- Marian, V., Blumenfeld, H. & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, 50, 940- 967.
- Marian, V. & Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, 24 (2), 173-193.
- Marinova-Todd, S. (2003). *Comprehensive analysis of ultimate attainment in adult second language acquisition*. Unpublished doctoral dissertation, Harvard University, Cambridge, MA.
- Marslen-Wilson, W. (1985). Speech shadowing and speech comprehension. *Speech Communication*, 4, 55-73.
- Marslen-Wilson, W. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing*. Cambridge, MA: MIT Press.
- Marslen-Wilson, W., Moss, H. & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 1376-1392.
- Marslen-Wilson, W. & Tyler, L. (1980). The temporal structure of spoken language understanding. *Cognition*, 8, 1-71

- Marslen-Wilson, W. & Tyler, L. (1987). Against modularity. In J. Garfield (Ed.), *Modularity in Knowledge Representation and Natural Language Understanding*. Cambridge, Mass: MIT Press.
- Marslen-Wilson, W. & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, *101*, 653–675.
- Marslen-Wilson, W. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*, 29-63.
- Marslen-Wilson, W. & Zwitserlood, P. (1989). Accessing spoken word: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 576-585.
- Massaro, D. (1975). Preperceptual Auditory Storage in Speech Recognition. In (Eds.) A Cohen and S. Nootboom, *Structure and Process in Speech Perception*, Springer- Verlag, Heidelberg.
- Massaro, D. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, *79*, 124-145.
- Mastromonaco, S. (1999). Liaison in French as a Second Language. Unpublished doctoral dissertation, University of Toronto.
- Matter, J. (1989). Some fundamental problems in understanding French as a foreign language. In H. Dechert & M. Raupach (Eds.), *Interlingual Processes* (p. 105-119). Tübingen: Narr.
- Mattys, S. (2003). Stress-based speech segmentation revisited. In *Proceedings of Eurospeech: The 8th Annual Conference on Speech Communication and Technology*. (pp. 121–124). Geneva.
- Mattys, S. (2004). Stress versus coarticulation: Towards an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 397-408.
- Mattys, S., Jusczyk, P., Luce, P. & Morgan, J. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*, 465–494.
- Mattys, S. & Melhorn, J. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *Journal of the Acoustical Society of America*, *122*, 554-567.
- Mattys, S., White, L. & Melhorn, J. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*, 477-500.
- Maye, J., Aslin, R. & Tanenhaus, M. (2003). In search of the weckud wetch: Online adaptation to speaker accent. In *Proceedings of the 16th Annual CUNY Conference on Human Sentence Processing*, March 27–29, Cambridge, MA.

- Maye, J., Aslin, R. & Tanenhaus, M. (2008). The Weckud Wetch of the Wast: Lexical Adaptation to a Novel Accent. *Cognitive Science*, 32, 543–562.
- Maye, J., Weiss, D. & Aslin, R. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, 11 (1), 122–134.
- Maye, J., Werker J. & Gerken L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101-B111.
- McAllister, R. (1997). Perceptual foreign accent: L2 users comprehension ability. In J. Leather & A. James (Eds.) *Second-Language Speech*. Berlin: Mouton de Gruyter.
- McAllister, R., Flege, J. & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, 30, 229-258.
- McCandliss, B., Fiez, J., Protopapas, A., Conway, M. & McClelland, J. (2002). Success and failure in teaching the r-l contrast to Japanese adults: predictions of a hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, and Behavioral Neuroscience*, 2(2), 89-108.
- McClelland, J. & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.
- McClelland, J., Fiez, J. & McCandliss, B. (2002). Teaching the /r/-/l/ discrimination to Japanese adults: behavioral and neural aspects. *Physiology and Behavior*, 77, 657-62.
- McClelland, J., Mirman, D & Holt, L. (2006). Are there interactive processes in speech perception? *TRENDS in Cognitive Science*, 10 (8), 363-369.
- McQueen, J. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, 39, 21-46.
- McQueen, J. (2001). Spoken word access processes: an introduction. *Language and Cognitive Processes*, 16, 469-490.
- McQueen, J. (2007). Eight questions about spoken-word recognition. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 37-53). Oxford: Oxford University Press.
- McQueen, J. & Cutler, A. (1992). Words within words: Lexical statistics and lexical access. *Proceedings of the Second International Conference on Spoken Language Processing* (Vol. 1, pp. 221-224). Banff, Canada;
- McQueen, J., Norris, D. & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 621-638.

- McQueen, J., Norris, D. & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, 25 (5), 1363-1389.
- McQueen, J., Otake, T. & Cutler, A. (2001). Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language*, 45, 103–132.
- Mehler, J., Dommergues, J.Y., Frauenfelder, U. & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20, 298-305.
- Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J. & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29, 143-178.
- Miller, J. (1987). Rate-dependent processing in speech perception. In A. Ellis (Ed.), *Progress in the Psychology of Language*, (pp. 119– 57). Hillsdale, NJ: Erlbaum
- Miller, J. & Jusczyk, P. W. (1989). Seeking the neurobiological basis of speech perception. *Cognition*, 33, 111-137.
- Mirman, D., McClelland, J. & Holt, L. (2006). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin & Review*, 13(6), 958–965.
- Mitterer, H. & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception & Psychophysics*, 65, 956-969.
- Morais, J., Cary, L., Alegria, J. & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7, 323-331.
- Morel, M.-A. & Danon-Boileau, L. (1998). *Grammaire de l'intonation – l'exemple du français*. Paris: Ophrys.
- Morgan, J. & Saffran, J. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, 66, 911–936.
- Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76, 165-178.
- Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation and instruction. *Studies in Second Language Acquisition*, 21, 81-108.
- Nakatani L. & Dukes, K. (1977). Locus of Segmental Cues to Word Juncture. *Journal of the Acoustical Society of America*, 62, 714-719.
- Nakatani, L. & Schaffer, J. (1978). Hearing “words” without words. *Journal of the Acoustical Society of America*, 63, 234-245.

- Nas, G. (1983). Visual word recognition in bilinguals: evidence for a cooperation between visual and sound based codes during access to a common lexical store. *Journal of Verbal Learning and Verbal Behavior*, 22, 526-534.
- Nazzi, T., Bertoncini, J. & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 756-766.
- Nazzi, T. & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, 41(1), 233-243.
- New B., Pallier C., Ferrand L. & Matos R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'Année Psychologique*, 101, 447-462.
- Newport, E. (1990). Maturation constraints on language learning. *Cognitive Science* 14, 11-28.
- Norris, D. (1990). *A dynamic-net model of human speech recognition.*, In G. Altmann (Ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (pp.87-105). Cambridge, MA: MIT Press.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52(3), 189-234.
- Norris, D. & McQueen, J. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115 (2), 357-395.
- Norris, D., McQueen, J. & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1209-1228.
- Norris, D., McQueen J. & Cutler A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23 (3), 299-370.
- Norris, D., McQueen, J. & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204-238.
- Norris, D., McQueen, J., Cutler, A. & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, 34 (3), 191-243.
- Nguyen, N. (2005). La perception de la parole, In, N. Nguyen, S. Wauquier-Gravelines & J. Durand (Eds.), *Phonologie et Phonétique, forme et substance* (pp. 425-447). Paris: Hermès.
- Nguyen N., Fagyal, Z. & Cole, J. (2004). Perceptual relevance of long-domain phonetic dependencies. *Journées d'Etudes Linguistiques*, Nantes : France.
- Nguyen N. & Hawkins, S. (2003). Temporal integration in the perception of speech: introduction. *Journal of Phonetics*, 31, 279-287.

- Nguyen, N., Wauquier, S., Lancia, L. & Tuller, B. (2007). Detection of liaison consonants in speech processing in French: Experimental data and theoretical implications. In P. Prieto, J. Mascaró & M.-J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 3–23). Amsterdam: John Benjamins.
- Nyrop, K. (1925). *Manuel de phonétique du français parlé*. (4th edition). Copenhagen.
- Oller, D. (1973). The Effect of Position in Utterance on Speech Segment Duration in English. *Journal of the Acoustic Society of America*, 54, 1235-1247.
- Otake, T., Hatano, G., Cutler, A. & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32, 258-278.
- Pallier, C., Bosch, L. & Sebastián-Galles, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64 (3), B9-B17.
- Pallier, C., Christophe, A. & Mehler, J. (1997). Language-specific listening. *Trends in Cognitive Science*, 1(4), 129-132.
- Pallier, C., Colomé, A. & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: exemplar-based versus abstract lexical entries. *Psychological Science*, 12(6), 445-449.
- Pallier, C., Sebastián-Galles, N., Felguera, T., Christophe, A. & Mehler, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language*, 32, 373-389.
- Palmeri, T., Goldinger, S. & Pisoni, D. (1993). Episodic encoding of speaker's voice and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 309-328.
- Passy, P. (1917). *Les sons du français*. (8th edition). Paris: Didier.
- Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics*, 11, 73–89.
- Pegg, J. & Werker, J. (1997). Adult and infant perception of two English phones. *Journal of the Acoustical Society of America*, 102, 3742-53.
- Peperkamp, S. (2003). Phonological acquisition: Recent attainment and new challenges. *Language and Speech*, 46, 87–113.
- Peperkamp, S. & Dupoux, E. (2002). A typological study of stress 'deafness'. In: C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 203-240). Berlin: Mouton de Gruyter,

- Peperkamp, S. & Dupoux, E. (2007). Learning the mapping from surface to underlying representations in an artificial language. In J. Cole & J. Hualde (Eds.), *Papers in laboratory phonology IX* (pp. 315–338). Berlin: Mouton de Gruyter.
- Peperkamp, S., Dupoux, E. & Sebastián-Gallés, N. (1999). Perception of stress by French, Spanish, and bilingual subjects. *Proceedings of EuroSpeech '99, Vol. 6*, 2683-2686.
- Peperkamp, S., Le Calvez, R. Nadal, J.-P. & Dupoux E. (2006). The acquisition of allophonic rules: statistical learning with linguistic constraints. *Cognition, 101*, B31-B41.
- Peperkamp, S., Pettinato, M. & Dupoux, E. (2003). Allophonic variation and the acquisition of phoneme categories. In B. Beachley, A. Brown & F. Conlin (Eds.) *Proceedings of the 27th Annual Boston University Conference on Language Development: Volume 2* (pp. 650-661). Somerville, MA : Cascadilla Press,
- Perani, D., Paulesu, E., Sebastián-Gallés, N., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S.F., Fazio, F. & Mehler, J. (1998). The bilingual brain: proficiency and age of acquisition of the second language. *Brain, 121*, 1841-1852.
- Pike, K. (1947). Grammatical prerequisites to phonemic analysis. *Word, 3*, 155-172.
- Pinker, S. (1994). *The language instinct: How the mind creates language*. New York: W. Morrow.
- Piske, T., Flege, J. & MacKay, I. (2000). Factors reflecting degree of foreign accent. *New Sounds*. Amsterdam.
- Piske, T., MacKay, I. & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*, 191-215.
- Pisoni, D. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *Journal of the Acoustical Society of America, 61*, 1352–61.
- Pisoni, D. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. Mullennix (Eds.), *Talker variability in speech processing*. (pp. 9-32). San Diego: Academic Press.
- Pisoni, D. & Luce, P. (1987). Acoustic-phonetic representations in word recognition. *Cognition, 25*, 21-52.
- Pisoni, D. & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception and psychophysics, 15 (2)*, 285-290.
- Pitt, M. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory & Cognition, 21*, 1037–1052.
- Pitt, M. & Samuel, A. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology, 29*, 149-188.

- Polivanov, E. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague* 4, 79–96.
- Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America*, 95, 1286-1296.
- Polka, L., Colantonio, C. & Sundara, M. (2001). A cross-language comparison of /d-/D/ perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America*, 109(5), 2190-2201.
- Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, 20, 331-350.
- Reicher, G. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 275-280.
- Rivera-Gaxiola, M., Silva-Pereyra, J. & Kuhl, P. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. *Developmental Science*, 8 (2), 162–172.
- Rose, Y. & Wauquier-Gravelines S. (2007). Acquisition of speech in French. In S. McLeod (Ed.), *International guide of Speech Acquisition*, (pp. 364-386). Thomson Delmar Learning.
- Rubin, P., Turvey, M. & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in non-words. *Perception & Psychophysics*, 19, 394-398.
- Saffran J., Newport E., & Aslin R. (1996). Word segmentation: The role of distributional cues. *Journal of Memory & Language*, 35, 606-621.
- Saffran, J. & Thiessen, E. (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39, 484-494.
- Salverda, A., Dahan, D. & McQueen, J. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90, 51-89.
- Sanders, L. & Neville, H. (2000). Lexical, syntactic, and stress-pattern cues for speech segmentation. *Journal of Speech, Language, and Hearing Research*, 43, 1301–1321.
- Sanders, L., Neville, H. & Woldorff, M. (2002). Speech segmentation by native and non-native speakers: the use of lexical, syntactic, and stress-pattern cues. *Journal of Speech, Language and Hearing Research*. 45, 519–29.
- Saunders, G. (1988). The structure of errors in the perception of French speech. *Revue de Phonétique Appliquée*, 86, 43-100.
- Scarborough, D., Gerard, L. & Cortese, C. (1984). Independence of lexical access in bilingual word recognition. *Journal of Verbal Learning and Verbal Behavior*, 23, 84-99.

- Schaegis, A., Spinelli, E. & Welby, P. (2005). Perception of phonemically ambiguous spoken sequences in French. *Proceedings of 27th Annual Conference of the Cognitive Science Society*. Stresa, Italy.
- Schneider, W., Eschman, A. & Zuccolotto, A. (2002). E-Prime User's Guide. Pittsburgh: Psychology Software Tools Inc.
- Schulpen, B., Dijkstra, A., Schriefers, H. & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 29 (6), 1155–1178.
- Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Rowley, MA: Newbury House.
- Sebastian-Galles, N. & Bosch, L. (2005). Phonology and bilingualism. In J. Kroll & A. de Groot (Eds.) *Handbook of Bilingualism* (pp. 68-87). Oxford University Press
- Sebastian-Galles, N., Dupoux, E., Segui, J. & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish: The role of Stress. *Journal of Memory and Language*, 31, 18-32.
- Sebastian-Galles, N. & Soto-Faraco, S. (1999). On-line processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, 72, 112-123.
- Segui, J., Dupoux, E. & Mehler, J. (1990). The role of the syllable in speech segmentation, phoneme identification and lexical access. In G. Altmann (Ed.), *Cognitive Models of Speech Processing* (pp. 263-280). Cambridge Mass: MIT Press.
- Selkirk, E. (1972). *The Phrase Phonology of English and French*. Doctoral dissertation, MIT.
- Selkirk, E. (1982). The syllable. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations, Part II* (pp. 337-383).
- Shattuck-Hufnagel, S. & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25(2), 193-247.
- Shatzman, K. & McQueen, J. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception and Psychophysics*. 68, 1-16.
- Shea, C. & Curtin, S. (2005). *Learning Allophones from the Input*. Poster presented at the 30th Annual Boston University Conference on Language Development. Boston, MA.
- Shillcock, R. (1990). Lexical hypotheses in continuous speech. In G. Altmann (Ed.), *Cognitive models of speech processing* (pp. 24–49). Cambridge, MA: MIT Press.
- Shoemaker, E. (2005). The Resolution of French liaison by native and non-native speakers. Unpublished Master's Thesis: The University of Texas at Austin.

- Shoemaker, E. (2006). *Durational differences in word-initial and word-final liaison consonants in French*. Unpublished manuscript, The University of Texas at Austin.
- Shoemaker, E. & Birdsong, D. (2008.) La résolution de la liaison par des locuteurs natifs et non-natifs. *Acquisition et Interaction en Langue Etrangère*, 27, 43-62.
- Singleton, D. & Ryan, L. (2004). *Language Acquisition: The Age Factor* (2nd ed.). Clevedon UK: Multilingual Matters.
- Siskind, J. (1996). A Computational Study of Cross-Situational Techniques for Learning Word-to-Meaning Mappings. *Cognition*, 61 (1), 39-91.
- Smith, A. (1998). French variable liaison: A proposed simplification. *Francophonie*, 17, 11–14.
- Smith, R. & Hawkins, S. (2000). Allophonic influences on word-spotting experiments. In A. Cutler, J. McQueen & R. Zondervan (Eds.) *Proceedings of the Workshop on Spoken Word Access Processes* (pp. 139-142). Nijmegen: MPI.
- Soto-Faraco, S., Sebastián-Gallés, N. & Cutler, A. (2001). Segmental and supra-segmental mismatch in lexical access. *Journal of Memory and Language*, 45, 412-432.
- Spinelli, E., McQueen, J. & Cutler, A. (2002). Resolution of liaison for lexical access in French. *Revue Française de Linguistique Appliquée*, VII-1, 83-96.
- Spinelli, E., McQueen, J. & Cutler, A. (2003). Processing Resyllabified Words in French. *Journal of Memory and Language*, 48, 233-254.
- Spinelli, E. & Meunier, F. (2005). Le traitement cognitif de la liaison dans la reconnaissance de la parole enchaînée. *Langages*, 158, 79-88.
- Spinelli, E., Welby, P. & Schaegis, A. (2007). Fine-grained access to targets and competitors in phonemically identical spoken sequences: the case of French elision. *Language and Cognitive Processes*, 22 (6), 828 – 859.
- Spivey, M. & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, 10, 281–284.
- Stevens, K. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111 (4), 1872–1891.
- Strange, W. (1995). Cross-language studies of speech perception: A historical review. In Strange, W. (Ed.) *Speech perception and linguistic experience: Issues in cross-language speech research*. (pp 3-45). Timonium, MD: York Press.
- Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S., Nishi, K. & Jenkins, J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, 26, 311-344.

- Strange, W., Verbrugge, R., Shankweiler, D. & Edman, T. (1976). Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America*, 60, 213-224.
- Strauss, T., Harris, H. & Magnuson, J. (2007). jTRACE: A reimplement and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods*, 39, 13-30.
- Stridfeldt, M. (2003). French liaison and segmentation of non-words by Swedish learners of French. *PHONUM* 9, 169-172.
- Stridfeldt, M. (2005). La perception du français oral par des apprenants suédois. Unpublished doctoral dissertation. Umeå University, Sweden.
- Summerfield, A. (1975). How a full account of segmental perception depends on prosody and vice versa. In A. Cohen & S. G. Neebboom (Eds.), *Structure and process in speech perception*. New York: Springer-Verlag,
- Summerfield, A. (1981). On articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 1074-1095
- Suomi, K., McQueen, J. & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language*, 36 (3), 422-444.
- Tabossi, P., Burani, C., & Scott, D. (1995). Word identification in fluent speech. *Journal of Memory and Language*, 34, 440-467.
- Tabossi, P., Collina, S., Mazzetti, M., & Zoppello, M. (2000). Syllables in the processing of spoken Italian. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 758-775.
- Tanenhaus, M., Magnuson, J., McMurray, B. & Aslin, R. (2000). No compelling evidence against feedback in spoken word recognition. *Behavioral and Brain Sciences*, 23(3), 348-349.
- Tanenhaus, M. & Spivey-Knowlton, M. (1996). Eye-tracking. Special issue: A guide to spoken word recognition paradigms. *Language and Cognitive Processes*, 11, 583-588.
- Thomas, A. (2004). Phonetic norm versus usage in advanced French as a second language. *International Review of Applied Linguistics*, 42, 365-382.
- Tranel, B. (1984). Closed syllable adjustment and the representation of schwa in French. In *Proceedings of the 10th annual meeting of the Berkeley Linguistic Society*. California: University of Berkeley.
- Tranel, B. (1987). *The Sounds of French: An Introduction*. Cambridge University Press.

- Tranel, B. (1995). Current issues in French phonology: Liaison and position theories. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 798-816). Cambridge, MA: Blackwell.
- Trehub, S. (1976). The Discrimination of Foreign Speech Contrasts by Infants and Adults. *Child Development*, 47, 466-472.
- Treiman, R. & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language*, 27, 87-104.
- Troubetzkoy, N. (1949). *Principes de phonologie*. Paris (1964): Klincksieck.
- Tsushima, T., Takizawa, O., Sasaki, M., Siraki, S., Nishi, K., Kohno, M., Menyuk, P. & Best, C. (1994). Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6-12 months: Language specific developmental changes in speech perception abilities. Paper presented at *International Conference on Spoken Language Processing, 4. Yokohama, Japan*.
- Turk, A. & Shattuck-Hufnagel, S. (2000). Word-Boundary-Related Duration Patterns in English. *Journal of Phonetics*, 28, 397-440.
- Tyler, L. & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics*, 34, 409-420.
- Ulrich, R. & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, 123, 34-80.
- Umeda, N. (1975). Vowel Duration in American English. *Journal of the Acoustical Society of America*, 58, 434-445.
- van der Lugt, A. (2001). The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics*, 63 (5), 811-823
- van de Weijer, J. (1998). Language Input for Word Discovery. Doctoral dissertation. *Max Planck Series in Psycholinguistics 9*.
- Van Wijnendaele, I. & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 616-627.
- Van Zon, M. (1997). Speech processing in Dutch: A cross-linguistic approach. Unpublished doctoral dissertation. Tilburg University.
- Van Zon, M. & de Gelder, B. (1993). Perception of word boundaries by Dutch listeners. *Proceedings of the 3rd European Conference on Speech Communication and Technology* (pp. 689-692). Berlin,.
- Vitevitch, M. & Luce, P. (1998). When words compete: levels of processing in perception of spoken words. *Psychological science*, 9 (4), 325-329.

- Vroomen, J. & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21 (1), 98-108.
- Vroomen, J. & de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 710–720.
- Vroomen, J. & de Gelder, B. (1999). Lexical access of resyllabified words: Evidence from phoneme monitoring. *Memory & Cognition*, 27, 413-421.
- Vroomen, J., Van Linden, S., de Gelder, B. & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory–visual speech perception: Contrasting build-up courses. *Neuropsychologia*, 45 (3), 572-577.
- Vroomen, J., van Zon, M. & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory & Cognition*, 24 (6), 744-755.
- Wallace, W., Stewart, M. & Malone, C. (1995). Recognition memory errors produced by implicit activation of word candidates during the processing of spoken words. *Journal of Memory and Language*, 34, 417-439.
- Warner, N., Jongman, A., Sereno, J. & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics*, 32, 251-276.
- Wauquier, S. (2006). Des ornithorynques et consonnes « doublement flottantes », Hommage à Pierre Encrevé, phonologue variationniste. Paper presented at *Faire signe, pour Pierre Encrevé*, 16-18 October.
- Wauquier-Gravelines, S. (1994). Segmentation lexicale en Français parlé. In *Proceedings of the XXXème Journées d'Etude sur la Parole* (pp. 517–522). Trégastel.
- Wauquier-Gravelines, S. (1996). *Organisation phonologique et traitement de la parole continue, Contraintes prosodiques et phonologiques de l'accès lexical*. Doctoral dissertation. Université Paris 7.
- Wauquier-Gravelines, S. (2003). Troncation et réduplication. Peut-on parler de gabarits morphologiques dans le lexique précoce? In Fradin, B., Dal, G., Hathout, M., Kerleroux, F., Roché, M. Plénat, M. (Eds.), *Les unités morphologiques. Silexicales 3*, Université de Lille II.
- Wauquier-Gravelines, S. (2004). Les liaisons dangereuses. Psycholinguistique et phonologie : une interface complexe. In Ferrand, L. & Grainger, J. (Eds.) *Psycholinguistique cognitive, Essais en l'honneur de Juan Segui*. Bruxelles: De Boeck Université, Collection Neurosciences et Cognition.
- Wauquier-Gravelines, S. & Braud, V. (2005). Proto-déterminant et acquisition de la liaison obligatoire en français. *Langages*, 158, 53-65.

- Watkins, M. & Peynircioglu, Z. (1983). On the nature of word recall: Evidence for linguistic specificity. *Journal of Verbal Learning and Verbal Behavior*, 22, 385–394.
- Weber, A. (2000a). Phonotactic and acoustic cues for word segmentation in English. In B. Yuan, T. Huang & X. Tang (Eds.), *Proceedings of the 6th International Conference on Spoken Language Processing*, (Vol. 3, pp. 782-785). Beijing, China: Beijing International Convention Center.
- Weber, A. (2000b). The role of phonotactics in the segmentation of native and non-native continuous speech. In A. Cutler, J. McQueen & R. Zondervan (Eds.), *Proceedings of the Workshop on Spoken Access Processes*, (pp. 143-146). Nijmegen, the Netherlands: Max Planck Institute for Psycholinguistics.
- Weber, A. (2001). *Language-specific listening: The case of phonetic sequences*. Doctoral Dissertation. University of Nijmegen (*MPI Series in Psycholinguistics*, 16.)
- Weber, A. & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50 (1), 1-25.
- Weber, A. & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America*, 119, 597-607.
- Welby, P. (2003). French intonational rises and their role in speech segmentation. In *Proceedings of Eurospeech: The 8th Annual Conf. on Speech Communication and Technology*, (pp. 2125–2128). Geneva.
- Werker, J. (1995). Exploring developmental changes in cross-language speech perception. In L. Gleitman & M. Liberman (Eds.), *Language, Vol 1*. D. Osherson (Ed.), *An invitation to cognitive science, 2nd edition* (pp. 87-106). Cambridge, Mass: MIT Press.
- Werker, J. & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.
- Werker, J. & Tees, R. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology*, 50, 509-35.
- Werker, J. & Tees, R. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, 46(3), 233-251.
- Whalen, D. (1991). Infrequent words are longer in duration than frequent words. *Journal of the Acoustical Society of America*, 90(4), 2311.
- Whalen, D. (1992). Further results on the duration of infrequent and frequent words. *Journal of the Acoustical Society of America*, 91(4), 2339–2340.

- Whalen, D., Best, C. & Irwin, J. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics*, 25, 501-528.
- Wright, R., Souza, P., Bor, S. & Reed, S. (2008). Consequences of regional accent differences for speech perception studies. *The Journal of the Acoustical Society of America*, 123(5), 3864.
- Yersin-Besson, C. & Grosjean, F. (1996). L'effet de l'enchaînement sur la reconnaissance des mots dans la parole continue, *L'année Psychologique*, 96, 9-30.
- Yip, M. (2004). Possible-Word constraints in Cantonese speech segmentation. *Journal of Psycholinguistic Research*, 33, 165-173.
- Zwitserslood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32, 25-64.
- Zwitserslood, P. & Schriefers, H. (1995). Effects of sensory information and processing time in spoken-word recognition. *Language and Cognitive Processes*, 10 (2), 121-137.