Prosodic boundary cues in German:

Evidence from the production and perception of bracketed lists

Caterina Petrone[1] , Hubert Truckenbrodt[2], Caroline Wellmann[4], Julia Holzgrefe-Lang[4],

Isabell Wartenburger[4] and Barbara Höhle[4]

[1]Laboratoire Parole et Langage, CNRS UMR 7309 and Aix-Marseille Université, Aix-en-Provence, France,

[2]Centre for General Linguistics (ZAS) Berlin, Germany,

[3]Humboldt-Universität zu Berlin, Germany,

[4]Cognitive Sciences, Dept. Linguistics, Universität Potsdam, Germany

Running Head: Prosodic boundaries in German

Corresponding author:

Caterina Petrone

Laboratoire Parole et Langage

5, av. Pasteur

13100 Aix-en-Provence

FRANCE

Email: caterina.petrone@lpl-aix.fr

Abstract

This study investigates prosodic phrasing of bracketed lists in German. We analyze variation in pauses, phrase-final lengthening and f0 in speech production and how these cues affect boundary perception. In line with the literature, it was found that pauses are often used to signal intonation phrase boundaries, while final lengthening and f0 are employed across different levels of the prosodic hierarchy. Deviations from expectations based on the standard syntax-prosody mapping are interpreted in terms of task-specific effects. That is, we argue that speakers add/delete prosodic boundaries to enhance the phonological contrast between different bracketings in the experimental task. In perception, three experiments were run, in which we tested only single cues (but temporally distributed at different locations of the sentences). Results from identification tasks and reaction time measurements indicate that pauses lead to a more abrupt shift in listeners' prosodic judgments, while f0 and final lengthening are exploited in a more gradient manner. Hence, pauses, final lengthening and f0 have an impact on boundary perception, though listeners show different sensitivity to the three acoustic cues.

1. Introduction

The hypothesis that prosodic-metrical structure plays a part in understanding spoken language has been shared in research in various adjacent fields. The postulated structure includes word stress and a domain of prosodic word that goes with it (see e.g. Hayes, 1995), as well as related assumptions about larger prosodic constituents and metrical prominence at higher prosodic levels. Many authors, including Kent and Netsell (1971), Fougeron and Keating (1997), and Cho (2005) have documented phonetic effects of the metrical side of this

2

representation (stress- or prominence-related), and phonetic effects of the prosodic edges of different height. The literature on prosodic phonology since Selkirk (1980) and Nespor and Vogel (1986) has analyzed the sensitivity of postlexical phonological phenomena to higher structure in terms a prosodic-metrical structure accessed by these phenomena and a mapping between syntax and prosodic-metrical structure. Since Selkirk (1996) and Truckenbrodt (1995), these relations between syntactic and prosodic structure are mostly viewed as violable requirements of Optimality Theory (Prince and Smolensky, 2004) that can be overridden in a language-specific fashion, by focus, by prosodic length constraints or by other requirements on the prosodic structure (see Truckenbrodt, 2007a and Selkirk, 2011). In addition, a lot of experimental work on intonation since Pierrehumbert (1980) anchored the elements of the intonation contours to the prosodic-metrical structure, with pitch accents associating to stressed/prominent syllables and edge tones to the edges of prosodic constituents (e.g., Gussenhoven, 2004; Féry, 2016). Jun (1998) raised the question whether the prosodic constituents postulated by the intonational literature are the same as the ones postulated in the literature in prosodic phonology. Converging with Selkirk and Shen (1990) and Hayes and Lahiri (1991), she concluded based on her Korean data that they are the same prosodic constituents.

In the current paper on German, two levels of this prosodic-metrical structure above the prosodic word are of interest. Separate phonological phrases (PPh) are typically found for the separate arguments and adjuncts of a clause. Their prosodic head carries an obligatory accent in English, Dutch, and German (Gussenhoven 1983, 1992; Selkirk 1984). Their dependency on the syntax has been construed as a systematic relation of PPhs to syntactic XPs in Truckenbrodt (2007b). At a higher level, boundaries of intonation phrases (IP) are obligatorily found between separate root sentences (Downing, 1970; Nespor and Vogel, 1986) or speech act CPs (Selkirk, 2011; Truckenbrodt, 2015), but they also frequently occur

at the end of long constituents, in presence of specific syntactic structures (e.g., paretheticals, sentential adverbs) or as a function of syntactic complexity (Watson and Gibson, 2004). Given the syntax-prosody mapping, many studies have investigated how speakers use prosody to disambiguate an ambiguous syntactic structure like coordinated (or "bracketed") lists of names such as *Lola or Mona and Lena.* Such constructions have more than one underlying syntactic structure since the three names can be grouped into two constituents at the same syntactic level but with different branching structure: *Lola (or Mona and Lena)* vs. *(Lola or Mona) and Lena*. The production of such bracketed lists was recently investigated by Wagner (2005) in English and by Féry and Kentner (2010) as well as Kentner and Féry (2013) in German. These studies have demonstrated that speakers use prosodic cues such as final lengthening, f0, and pauses to solve the surface ambiguity and thus to convey the information about the branching direction and the level of syntactic embedding of the construction.

In this paper, we want to contribute to a further understanding of results from experiments of this kind by evaluating the results from a production and a perception study conducted with speakers of German. The production experiment investigated prosodic cues (pauses, final lengthening, and f0) to bracketing across boundaries of different levels of the prosodic hierarchy: the prosodic word (PW), the phonological phrase (PPh) and the intonation phrase (IP).

Moreover, we investigated the specific perceptual impact of each of three main cues to bracketing, that is, final lengthening, f0, and pauses – an area that has hardly been studied in German. In a previous study, Kentner and Féry (2013) found evidence that German listeners are able to recover the prosodic structure of bracketed lists. However, their perception experiment was based on natural stimuli, hence it is impossible to isolate the contribution of each of the three cues in their study. In the current paper, we run three

perception experiments with stimuli in which final lengthening, f0, and pause duration were separately manipulated.

In the following sections, we first review literature relevant to the question of how pauses, f0, and final lengthening are employed to disambiguate ambiguous syntactic structures in both production (*§ 1.1.*) and perception (*§ 1.2.*). After describing our predictions (*§ 1.3.*), we turn to the presentation of the production experiment (*§ 2*) and of the three perception experiments (*§ 3*). The overall pattern of the results and their theoretical implications for the syntax-prosodic mapping are finally discussed (*§ 4*).

1. 1. Final lengthening, f0, and pauses in production

Prosodic boundaries are signaled by various phonetic and phonological cues, like phrase-final lengthening, initial strengthening, glottalization, silent pauses, and edge tones (e.g., Beckman and Edwards, 1990; Pierrehumbert and Talkin, 1992; Turk and Shattuck-Hufnagel, 2007). In this paper, we are interested in three of these cues: pauses (i.e., pause occurrence and duration), final lengthening (as conveyed by segmental duration) and intonation (as conveyed by f0). Final lengthening (also known as "pre-boundary lengthening", "boundary-related lengthening") refers to durational adjustments near the boundaries of the prosodic constituents. It is usually found at different levels of the prosodic hierarchy, with the amount of lengthening increasing from minor (e.g., at the end of a PW) to major prosodic boundaries (e.g., at the end of an IP). Final lengthening mostly concerns the segments immediately adjacent to the boundaries. For instance, in American English, the IP boundary is signaled by acoustic lengthening of the phrase-final syllable, vowel, or rhyme (Browman and Goldstein, 1992). Some work also suggest that final lengthening can extend to material preceding the final syllable, even though the magnitude of this effect is smaller than in the direct vicinity of

the boundary (see Turk and Shattuck-Hufnagel, 2007 for a review). For instance, prosodic prominence may be at work in determining the scope of final lengthening (Katskika, 2016). At a tonal level, edge tones are associated to prosodic constituents, such as the boundary tones for intonation phrases (Ladd, 1998). Moreover, prosodic constituents at a higher level of the hierarchy such as IPs might be marked by f0 reset (the suspension of downstep with a return to a higher height after an IP boundary) and upstep (the scaling of tones to a phrasal reference-line). Both phenomena are well documented in the literature (cf. Ladd, 1988; van den Berg, Gussenhoven, and Rietveld, 1992; Truckenbrodt, 2007a; Truckenbrodt and Féry, 2015). As for pauses, it has been attested cross-linguistically that the occurrence of a pause and its duration are positively correlated with the syntactic complexity (number of syntactic nodes) of the upcoming utterance (e.g., Grosjean, Grosjean and Lane, 1979; Cooper and Paccia-Cooper, 1980; Krivokapić, 2007 for English; Strangert, 1997 for Swedish). From a prosodic perspective, silent pauses often occur at IP boundaries (e.g., Downing, 1970; Nespor and Vogel, 1986).

As for German, Kohler (1983) found that, for laboratory speech, the last syllable of an IP receives much more lengthening (ranging from 87% to 176%) than the immediately preceding stressed syllable (ranging between 15% and 31%). Moreover, IPs are characterized by f0 declination and by a pause right after the boundary. In laboratory speech, prenuclear rises within the same intonation phrase undergo downstep, while the following intonation phrase is characterized by partial reset (e.g., Truckenbrodt, 2007b). Peters, Kohler, and Wesener (2005) investigated the acoustic cues that accompany the IP boundaries in the Kiel Corpus of German spoken language. They found final lengthening before the boundary in 66.2%, f0 marking (f0 reset or separating contour) in 74%, and pauses in 38.3% of the annotated IP boundaries. In relation to lists of names, Féry and Kentner (2010) and Kentner and Féry (2013), asked speakers to read sentences in which the grouping of the names was

made explicit by brackets. The grouping was further triggered by a context preceding the target sentence. They found that names preceding a strong prosodic boundary are signaled by final lengthening (measured as duration of a whole name) and upstep. On the other hand, when two names group into one syntactic constituent, the leftmost name is shortened and reduced in f0 to weaken the prosodic boundary between the two names. The observed phenomenon is explained in terms of the *proximity* principle. We will return to this principle in the discussion session.

To sum up, phonetic and phonological differences have been found in segmental duration (final lengthening), f0 (e.g., upstep, edge tones) and in pause occurrences at the prosodic boundaries. While pauses are a strong marker of IPs, final lengthening and f0 are used across different levels of the prosodic hierarchy. Our production study will focus on the production of bracketed lists. Differently from Féry and Kentner (2010) and Kentner and Féry (2013), speakers had to accomplish the task in presence of an addressee. It has been suggested in the literature that speakers in interaction have as one goal to produce utterances that allow particular addressees to recover the meanings they intend (Clark and Carlson, 1982; Clark and Murphy, 1982). Such an 'audience design' procedure might influence the prosodic patterns of the utterances, in that speakers more reliably produce disambiguating cues to syntactic ambiguity (see Speer, Warren, and Shafe, 2011 for a short review). Hence, it is possible that our speakers will enhance the prosodic contrasts in order for the groupings to be better recovered by the addressee.

1.2. Final lengthening, f0, and pauses in perception

Relatively few studies have been carried out on the perception of boundary strength. Listeners have been shown to exploit final lengthening, f0, and pauses as cues of phrasing,

though their use might vary across different prosodic constituents. For instance, Wightman, Shattuck-Hufnagel, Ostendorf and Price (1992) reported that only four prosodic constituents can be distinguished based on the lengthening of vowels in phrase-final syllables. However, a larger number of prosodic boundaries can be distinguished if other cues such as pauses or f0 are taken into account. In a study on Dutch, Swerts (1997) reported that Dutch listeners can distinguish six degrees of boundary strength (at a discourse level) by pause duration, showing that there is a significant association between longer pauses and stronger prosodic boundaries. On the other hand, the amount of f0 reset and type of boundary tones (high vs. low) are less strong cues for prosodic boundary perception. Proportions of boundaries detected in conditions where only pause was present (51%) were higher than in conditions where only f0 reset (17%) or boundary tone (9%) was present.

Scott (1982) focussed only on temporal cues, namely, final lengthening and pause. She used short English sentences that contained a list of three names separated by different conjunctions (e.g., "Kate and Pat or Tony will come"). Two alternative groupings of the noun phrases were possible, resulting either in a major prosodic boundary after the first or after the second noun. The stimuli were edited in a stepwise manner by systematically increasing or decreasing the amount of final lengthening and the pause duration. The combination of final lengthening and pause duration was found to be an effective boundary cue, increasingly shifting the listeners' decision towards the alternative phrasing. However, listeners also identified a prosodic boundary if only an extended pause was present. The impact of final lengthening as a sole cue, on the other hand, was not tested in the study.

As for German, Peters (2005) showed that pauses are salient cues of prosodic boundaries, which can override the perceptual saliency of other boundary cues. In Peters' (2005), German listeners were asked to judge the strength of a prosodic boundary between two syntactic phrases in an utterance. The boundary between the two syntactic phrases was

systematically manipulated with respect to the height of the f0 contour between the two phrases, final lengthening, and pauses. Three different degrees of final lengthening and of the f0 contour were applied and occurred as sole cues, in combination with each other or in combination with a pause. Results showed that perception of different degrees of f0 and final lengthening was possible only when no pause was present in the same stimulus. As soon as a pause was present, the effects of lengthening and f0 were masked.

Finally, Gollrad (2013) examined the contribution of duration, pauses and f0 cues in the perception of ambiguous syntactic structures in German. In clause-internal disambiguations, a Determiner Phrase (DP)/Noun Phrase (NP) was either a possessor to the preceding noun, prosodically anayzed as $(NP_1 (NP_2)_{PPh})_{PPh}$, or it was a separate object of the verb, prosodically analyzed as $(NP_1)_{PPh}(NP_2)_{PPh}$. An f0 rise preceding NP_2 in both conditions differed in height between the conditions in the productions she elicited. This height distinction turned out to play a minor role perceptually, as did small pauses preceding NP_2, while the duration of NP_1 had strong effects in cueing the prosodic distinction. In another series of experiments, the possessor case $(NP_1 (NP_2))$ was compared with a condition in which the two NPs were separated by a sentence break, prosodically an IP boundary $(...NP_1)_{IP}(NP_2...)$. Here a high offset, analyzed as a high tone, in $(NP_1 {}^H(NP_2))$ contrasted with a final low offset, analyzed as a low boundary tone, in $(... NP_1 {}_L)_{IP}(NP_2 ...)$. The productions also showed distinctions in lenght of NP_1 and clear pauses at the sentence-boundary. The perception-studies showed that the most decisive cue in this case were the f0 cues for the H vs. L tones. While the other two cues, duration of NP_1 and pause, are less decisive, they can still together disambiguate the structure without the f0 cues.

To summarize, the impact of phonetic cues on boundary perception varies both across languages and across levels of boundary strength. For German, the relative impact of pauses and f0 is more unclear across studies. In particular, Gollrad et al. (2013) indicate that the

boundary tone information play a major role in the perception of an IP boundary. However, from Gollrad et al. (2013), we cannot assess fine-grained perception differences across the three cues, since the stimuli were based on the mere presence vs. absence of the cues. Hence, in our study, we will look at the effects of the progressive manipulation of pause occurrences and duration, final lengthening, and f0 on both identification scores and reaction times.

1.3. Expectations

We employed sequences of three names *N1 or N2 and N3* that were bracketed in writing as in *Lola (or Mona and Lena)* or as in *(Lola or Mona) and Lena*. As a control condition we added sequences like *Lola (or Mona Urlena)* in which *Mona Urlena* is understood as the first and last name of a single person. The bisyllabic German conjunction *oder* 'or' always followed the first name and the monosyllabic conjunction *und* [ʊnt] 'and' followed the second name and was replaced with the syllable 'Ur- [ʊɐ̯] in the condition with the name.

Our expectations for the prosodic structure of the stimuli are illustrated in the following and explained below. For a given name with the syntactic structure [[Lola]$_{NP}$]$_{DP}$, we provide the lexical NP in a separate line from the functional DP-structure for reasons of space.[1]

---

[1] Note that, without any added effects of attempted disambiguation, the theory of the syntax-prosody mapping that we employ leads us to expect that the syntactic structures [Lola und Lena][oder Manu] in the Late condition and [Lola][und Lena oder Manu] in the Early condition could be mapped to the same prosodic structure. Namely, they should be constituted by three phonological phrases (Lola)(und Lena)(oder Manu) within one large intonation phrase. This is because in the more explicit structure [$_{DP}$ [$_{NP}$ Lola]] only the lexical NP, the inner constituent, is expected to trigger prosodic boundaries or stress and the

Late: *(N1 N2) N3*

[	]IP	[	]IP
(x)PPh	(x)PPh	(x)PPh	
(x)PW	(x)PW	(x)PW	prosodic structure
[[ ]DP oder [ ]DP]DP	und [ ]DP		syntactic DPs
[Lola]NP oder [Mona]NP	und [Lena]NP		syntactic NPs
( )			bracket in the stimuli

Early: *N1 (N2 N3)*

[ ]IP [	]IP		
(x)PPh	(x)PPh	(x)PPh	
(x)PW	(x)PW	(x)PW	prosodic structure
[ ]DP oder [[ ]DP	und [ ]DP]DP		syntactic DPs
[Lola]NP oder [Mona]NP	und [Lena]NP		syntactic NPs
( )			bracket in the stimuli

Word: control

[ ]IP [	]IP

---

additional higher groupings of two names e.g. [DP Lola und Lena] are not such NPs but functional DPs.

```
(  x  )PPh   (                          x    )PPh
(  x  )PW    (          x    )PW   (       x    )PW      prosodic structure
[      ]DP    oder [                      ]DP           syntactic DPs
[Lola]NP     oder [Mona           Urlena]NP            syntactic NPs
             (                              )           bracket in the stimuli
```

As for the production task, we expected that each lexical word (in our case names with penultimate word stress, N1, N2 and N3) corresponds to a prosodic word (Nespor and Vogel, 1986; Vigário, 2003; Prince and Smolensky, 2004; Selkirk, 2011). Furthermore we expected that each lexical noun phrase (NP) gives rise to a phonological phrase by a pattern predicted and found to occur cross-linguistically (Nespor and Vogel, 1986; Selkirk, 1986, 2011; Truckenbrodt, 2007b) and confirmed in extensive experiments by the occurrence of accents in German sentences (Truckenbrodt 2002, 2004, 2007a, Féry and Kügler, 2008). This would lead to one phonological phrase for each name in the Late and Early conditions. The Word condition allows comparing the level of constituency of the prosodic word to the immediately higher prosodic constituent, the PPh. In the Word condition, the sequence *Mona Urlena* presents a first and surname and thus constitutes one complex name within one noun phrase, for which we expect a single phonological phrase.

The conjunction should be grouped with the name following it (rather than the preceding name) reflecting a cross-linguistic pattern also attested in German (Pheby, 1981). The Early and Word conditions are supposed to show the same boundary type after the first name.

The relevant boundaries should be phonetically detectable in the following comparisons. The comparison of Early and Word condition for boundary cues on the second name should provide evidence for the phonological phrase boundary in the Early condition at the end of the second name. The additional intonation phrase boundary after the second name

in the Late condition should in turn be detectable in a comparison of the cues on the second name in the Early and the Late condition. The presence vs. absence of an intonation phrase boundary after the first name in the Early vs. Late condition would be evident in boundary cue distinctions on the first noun in these two conditions. Consequences of the grouping distinctions should be measurable in terms of the magnitude of final lengthening and f0 effects, correlated with boundary strength (IP > PPh > PW) and in terms of (optional) pauses at intonation phrase boundaries.

Concerning the perception experiments, our main aim is to test the use of the three prosodic cues for the distinction in bracketing between the Early and the Late boundary condition. Given that the contrast should be implemented at the IP level, f0 should play a major role (Gollrad, 2013). The relevance of the three acoustic parameters will be verified by running three perception experiments (identification tasks) in three separate groups of listeners for final lengthening, f0, and pauses, respectively. For each experiment, reaction times (RT) in the identification task will be measured (see, e.g., Chen, 2003; Niebuhr, 2007 for application of this paradigm to prosodic contrasts). Reaction times are supposed to reflect task difficulty (Massaro, 1987). Hence, they might offer hints about the processing of the three prosodic cues for the purpose of the syntactic disambiguation. Longer reaction times are supposed to indicate the presence of a category boundary at that point of the continuum, and they should reflect shifts in the identification score, obtained in the two-alternative forced-choice task.

2. Production study

2.1. Methods

2.1.1. Stimuli

We constructed stimuli for the three conditions as illustrated in the following. The brackets were part of the visual presentation of the stimuli and were to be produced recoverably by our speakers.

Late:	(Lola oder Mona) und Lena	*(N1 N2) N3*
Early:	Lola (oder Mona und Lena)	*N1 (N2 N3)*
Word:	Lola (oder Mona Urlena)	*N1 (N2 Ur-N3)*

Six expressions were used in each condition; three were generated by permuting the names given in the preceding example, three by using permutations of the additional names *Lilli*, *Manu*, *Nina*, which also have penultimate word stress. The permutations maintained the schemata for the Late condition: *(N1 N2) N3*, for the Early condition: *N1 (N2 N3)*, and for the Word condition: *N1 (N2 Ur-N3)*. They therefore also maintained the expected prosodic structures as illustrated in the Late, Early and Word conditions, respectively. In each condition, each name occurred once initially, once medially, and once in final position.

A pseudo-randomized list of stimuli was constructed. It contained each of these 18 bracketed sentences, which were repeated six times, for a total of 18 × 6 = 108 target expressions. All target sentences were preceded by the same context question (*Wer kommt?* "Who is coming?"). The pseudo-randomized list also included 180 fillers. The fillers also exhibited an early or a late boundary in a list of three elements (e.g., a late boundary after *Arzt* in: (*Die Krankenpflegerin oder der Arzt) (und der Lehrling) (essen in der Mensa)* vs. an early boundary after *Die Krankenpflegerin* in: (*Die Krankenpflegerin)(oder der Arzt und der Lehrling) (essen in der Mensa)* "The nurse or the doctor and the apprentice eat at the

cafeteria"). The fillers were preceded by an appropriate wh-question (e.g., *Wer isst in der Mensa?* "Who eats in the cafeteria?") and they were intended to add more variation at the segmental and prosodic level. A subset of the fillers included sentences contrasting with the target sentences of the Word condition. That is, they were characterized by a late position of the boundary: *(N1 or N2) Ur-N3*. This kind of filler sentences was only added to balance the sample size across the experimental conditions. As a consequence, they were not considered in the acoustic analysis.

In total, each list contained 108 target sentences plus 180 fillers, which were numbered from 1 to 288. Two versions of the list were created for presentation during the experiment, one with the brackets left in place, another one without the brackets. The lists were used in the reading task, in the way detailed in § *2.1.3*.

2.1.2. Participants

Fourteen monolingually raised German native speakers participated in the production study. They came from the Berlin-Brandenburg area and were university students in Potsdam or Berlin, between 20 and 30 years old (8 women). They were reimbursed for their time. After recordings, 12 additional students who did not participate in the production experiment accomplished a perceptual annotation of the produced tokens (see below).

2.1.3. Procedure

Recordings. The recordings with the speakers were made in an anechoic room of the speech lab at the Centre for General Linguistics (ZAS) in Berlin. The second author (a native speaker of German) and a participant (speaker) sat facing each other at a table. The speaker

had the bracketed list of stimuli in front of him/her, the experimenter had the list without brackets in front of him. A box placed on the middle of the table allowed them to see each other's faces, but neither of them could see the printed paper in front of the other person. As the experimenter had not been involved in the creation of the randomizations and the corresponding lists he was blind to the bracketing. The speaker was informed that the experimenter only had the list without brackets and that he/she should read the sequences in a way that would allow the experimenter to place the brackets into the right position. During the recordings, the experimenter read the number of each sentence and the context question. As an answer to the context question, the participant then read the bracketed sentence. The experimenter then added brackets to the sequence on his list, based on his auditory impression. After the experimenter had finished his annotation, he proceeded to read the next sentence number and the corresponding context question.

To ensure that the participants paid sufficient attention to the bracketing in their renditions and realized it in such a way that it would be recoverable for the experimenter, speakers could ask for a repetition of a sentence if they felt that their rendition was not natural. In case of a repetition, the experimenter read the context question again and the speaker read the bracketed sentence again.

The recording sessions lasted about 60 minutes. Two pauses were made, one after approximately 20 minutes and another one after approximately 40 minutes.

Perceptual annotation. The annotators were employed to assess whether a speaker produced the tokens in accordance with the expected bracketing. After a set of instructions, they were to listen to the recorded target sentences by headphones and annotate for each token either (N1 N2) N3-bracketing or N1 (N2 N3)-bracketing or N1 (N2 Ur-N3) on an unbracketed list of the target sentences in front of them. Filler sentences with (N1 Ur-N2) N3 structure were

also included in the corpus in order to balance the sample size, thus avoiding possible perceptual biases towards one or the other prosodic phrasing. The other fillers were not annotated. The annotators also had the choice to mark a rendition as unclear. Each annotator judged 3 out of 6 repetitions of each sentence of four speakers, that is, 216 (3*18*4) target sentences. The same repetition of the target sentence was judged by two annotators. The perceptual annotation lasted around 30 minutes.

The results show that all speakers successfully accomplished the task. That is, 1276 tokens (98.5%) were annotated with the original bracketing provided on the speaker's list by at least two out of three annotators (including the one from the experimenter during the recording session). Only 20 (1.5%) of the recorded tokens were judged as perceptually more ambiguous. All tokens were included in the following acoustic analyses.

Acoustic analysis. The token renditions were saved as separate files and processed with Praat (Boersma, 2001). The segmentation was conducted by the first author. It included the determination of the three cues of interest:

- *Final lengthening*. The duration of the final vowel in the first and second name of each token was measured. We chose the final vowels as cues of final lengthening as, cross-linguistically, final lengthening is more reliable in the syllable rimes than in their onset consonant (see Turk and Shattuck-Hufnagel, 2007 for a review). Indeed, we observed that the consonant preceding the vowels was minimally affected by prosodic boundary differences (6-9 ms). Creaky voice was defined according to Redi and Shattuck-Hufnagel (2001). Creaky voice was found in the last portion of the word-final vowels but its percentage (measured as a proportion of the whole vowel duration) was overall very low (mean values ranging from 2.5% to 17.5% per sentence). When creakiness occurred, segmental duration was computed as the sum of

17

the modal and the creaky portions of the vowels. Note that creakiness by itself can be a cue to boundary type; however, its separate contribution is not further investigated here (but see, e.g., Pierrehumbert and Talkin, 1992). Two out of 14 speakers (both men) showed creaky voice to such an extent that the majority of their recordings could not be reliably measured. Therefore, these two speakers were discarded from all further evaluations. As a consequence, our acoustic analyses are based on the recordings of 12 speakers (8 women).

- *f0*. The f0 contours were characterized by strong variability. In the majority of cases, a rising pattern was found on the first name (N1) and, separately, on the second name (N2). Each f0 rise started in the stressed syllable, and reached its maximum in the final syllable of the names. We measured the starting (corresponding to a Low target, L) and end point (corresponding to a High target, H) of such rises for both N1 (L1 and H1) and N2 (L2 and H2) as well as the utterance final L (L%). We did not measure the accent on the third name (N3) since it was always realized as a fall. The rise was defined as being of at least one semitone (thus above the just noticeable difference for f0 variation; cf. Knight and Nolan, 2006). In other cases, a high plateau spanned over the names. Hence, the end of the plateau was measured (resulting from the insertion of a H target). Finally, in fewer cases, the end of the names was marked by an f0 fall (resulting from the insertion of L target). In these cases, the end of the fall was measured. Some examples of the variability found for f0 contours are reported in Figure 1.

Based on Truckenbrodt (2007b), the phonological analysis of the f0 rises and high plateaus assumes a L*+H H- tonal pattern at the p-phrase, where L*+H is a pitch-accent and H- an edge tone. For f0 rises, the evaluation does not distinguish whether the f0 peak is determined by the H of the L+H or by the H-. The results on

18

German and the model of Truckenbrodt (2007b) provide a basis for treating these two as equivalent for the purposes at hand: First, at the end of an initial p-phrase or a downstepped p-phrase, the end of a L*+H rise and a following H edge tone of the p-phrase are of comparable height, either scaled initially to the same level or downstepped together, forming a horizontal plateau when they are separated by a syllable. They are scaled to the same abstract height in the model of Truckenbrodt (2007b) and in fact reanalyzed as a single, spreading H edge tone (Truckenbrodt, 2016). Second, at the end of a non-final intonation phrase, an upstep occurs in the recordings reported on there. Depending on speakers, this may occur on the edge tone of the intonation phrase or on the nuclear L*+H rise before that or both; speakers showing both had them at a comparable height and across the different speakers the upstep was found to be of comparable height to the initial peak. In our current material, we expected in many cases a condensed realization that does not allow for distinguishing the two H tones, i.e., the f0 rise. However, for all speakers, a condensed realization would contain the upstepped height and would be captured by our measurement of the maximal f0 in this part of the recording. At the same time, all speakers distinguished downstepped rises of non-final p-phrases from upstepped rises at the end of non-final i-phrases. Thus, in the intonation patterns investigated here, it can be assumed that the maximal f0 height of a rise is a cue to boundary strength regardless of which H tone in L+H H it is attributed to.

- *Pauses*. Silent pauses were annotated if a gap in the spectrogram was at least 20 ms long.

To sum up, we analyzed 108 target sentences x 12 speakers = 1296 utterances. Figure 6 shows an example of prosodic labeling.
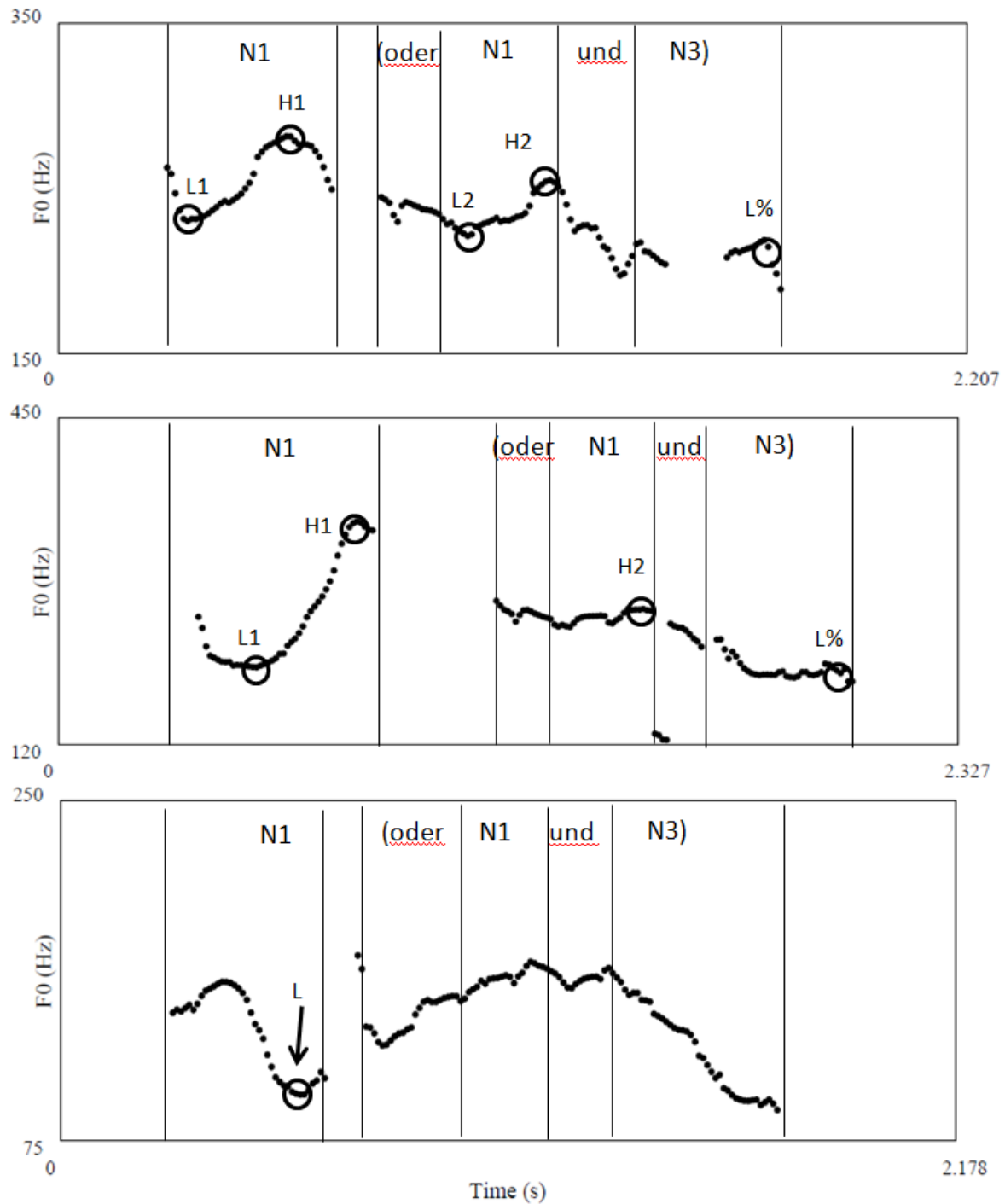
Figure 1. f0 curves for three tokens produced with an early boundary and characterized by a rising pattern both on N1 and N2 (top); by a rise on N1 and a high plateau on N2 (middle) and by a fall on N1 (bottom). Vertical

straight lines indicate word boundaries.Only tokens containing f0 rises or plateaus on N1 and N2 were intonationally annotated.

Statistics. The statistical significance of the results was tested using a series of linear mixed models by means of the R-environment (R Development Core Team, 2012, version 3.3.0, package R *lme4*). Mixed models have emerged as powerful tools for statistical analysis and they are preferable to more classical statistical analyses like repeated measures ANOVA for this kind of data (Quené and Van den Bergh, 2004; Baayen, 2008). The mixed models were separately run on the duration of the final vowels of N1 and N2 and on the f0 height at the end of N1 and N2 (either the f0 peak or the end of the high plateau). The mixed models included the factor BOUNDARY as fixed factor (which had as levels the three boundary conditions: Late, Early and Word). The Early condition was the intercept of the model (reference level), to which the other conditions were compared ("dummy" contrasts). For f0, the factor SEX was added since males have lower fundamental frequency than females.

We started the statistical analysis by fitting each model with all possible random effect components included (Barr, Levy, Scheepers, and Tily, 2013). Since the full models showed some over-parameterization (e.g., when the variance explained by a specific factor is close to zero), backward elimination based on principal component analysis and on likelihood-ratio tests was used to decide which components should be retained in the models (Bates, Kliegl, Vasishth, and Baayen, submitted). Principal component analysis of the mixed models enabled to compute the cumulative proportion of variance explained for the random components. Likelihood-ratio tests were run comparing full models (e.g., which contained a random component) with simpler ones (e.g., without that component). Models minimally included an intercept for WORD, SPEAKER and a slope BOUNDARY by SPEAKER. As for *p*-values, the standard functions in R to calculate linear mixed models do not provide them when the dependent variables are continuous (i.e., duration of pauses and final

lengthening; f0 peaks). For these variables, we employed bootstrapping methods with replacement from the original sample (number of sample = 500) to estimate *p*-values and 95% confidence intervals (CI) for each mixed model (function confint.merMod, Wald method). For the discrete dependent variable (pause occurrence), we report the *p*-value of the logit model. The cut-off point for significance was set at $p < 0.05$.

2.2. Results

2.2.1. Final lengthening

Figure 2 shows the lengthening of the final vowels. The duration of the vowel was higher immediately before the bracket. In fact, the average duration of the final vowel of N1 was 105 ms in the Late condition, 203 ms in the Early condition, and 200 ms in the Word condition. The duration of the final vowel of N2 was 179 ms, 96 ms and 106 ms in the Late, Early, and Word conditions, respectively.

The contrast between the Late and Early conditions was significant for the final vowels of N1 [β = -0.09, SE = 0.0076, t = -12.57, $p < .01$, CI = -0.11; -0.07] and N2 [β = 0.08, SE = 0.0052, t = 15.89, $p < .01$, CI = 0.07; 0.09]. As expected, the contrast between the Early and Word conditions didn't show any effects for N1. Contrary to our predictions, though, no differences between the Early and Word conditions were found for N2.

22
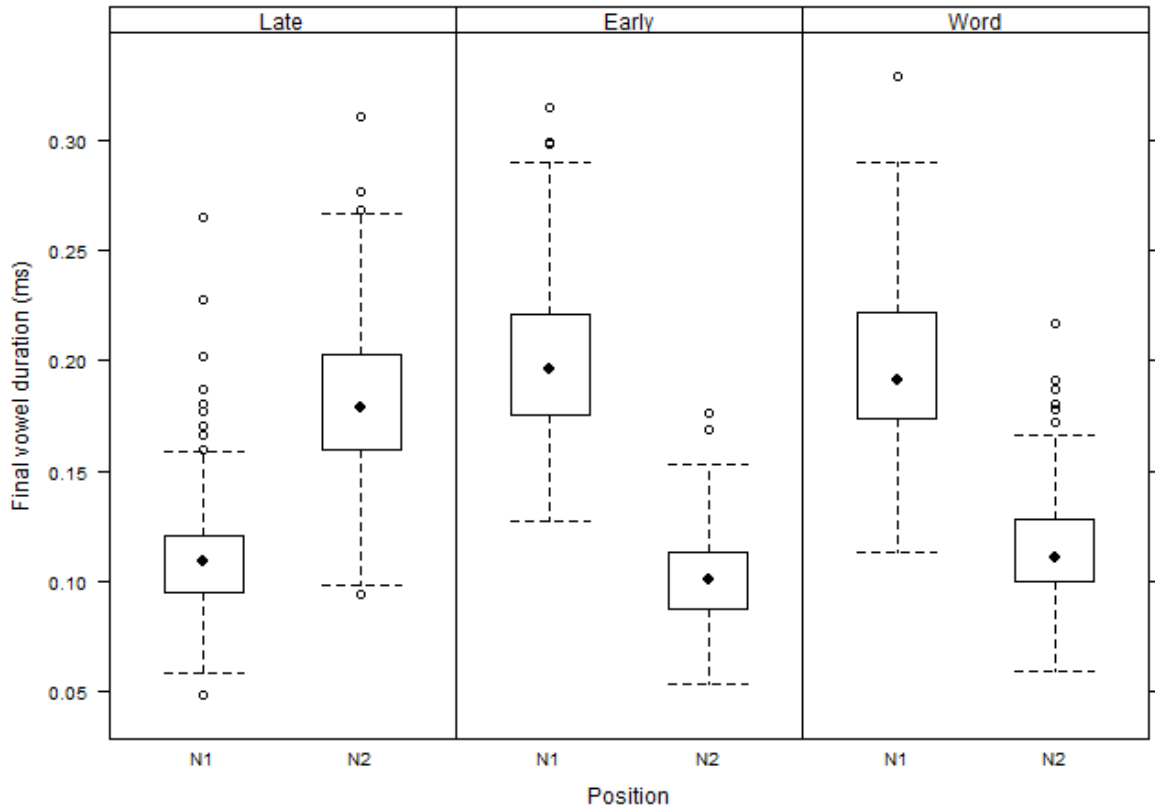
Figure 2. Boxplots for final vowel duration (y-axis) of N1 and N2 split across the three boundary conditions. The bracketing patterns are: Late = (N1 N2) (N3); Early = (N1) (N2 N3); Word = (N1) (N2 Ur-N3). Data are collapsed across speakers.

2.2.2. f0

A great variability was found for the realization of the f0 contour on the first (N1) and on the second (N2) names. Table 1 shows that the prosodic boundary at the end of N1 and N2 is predominantly signaled by a high f0 value (resulting either from a f0 rise or a high plateau) rather than a low f0 value (resulting from a pitch fall). This difference in frequency of occurrence is statistically significant on both N1 [$\beta = 1.54$, SE = 0.005, z = 27.33, p < 0.001] and N2 [$\beta = 1.86$ SE = 0.06, z = 28.89, p < 0.001].

Looking at the results by speaker, we can observe individual strategies for signaling prosodic boundaries. Therefore, individual data are plotted in Figure 3. This figure shows that

there is a clear preference across speakers for the production of single rises, with final f0 peaks on both N1 and N2. This is our expected f0 contour. A second contour is also possible, with the f0 stretch being realized as a high plateau spanning over the whole names. This configuration is optionally used by a few speakers (*fs*, *js*, *ts*). In either case, the prosodic boundary is marked by a H edge tone. The presence of a falling contour (marked by a final L edge tone) is also speaker-specific and frequently used only by speakers *as*, *jb*, *ls*. Finally, other f0 contours (e.g., rise-falls) are very rare.

The statistical analysis was performed on the f0 value at the end of the names (either the f0 peak of the final syllable or the f0 value the end of the high plateau). It confirmed a significant contrast between Late and Early condition both for N1 [β = -45.43, SE = 22.14, t = -2.05, p = .04, CI = -91.19, -0.83] and N2 [β = 40.88, SE = 14.49, t = 2.82, p = .008, CI = -12.86, 71.30].

The contrast between Early and Word boundary was not significant for N1 and for N2. As expected, SEX was significant (p < .01).

Due to the dominant use of the f0 rises, we focus here on tokens in which both N1 and N2 are realized with single f0 peaks. This allows a more straightforward evaluation of the contribution of f0 height on boundary strength, since we interpret the f0 peaks to be single instantiations of the H H- sequence in the L*+H H- tonal pattern. On average, the f0 peak was higher on N2 (243.2 Hz) than on N1 (231.2 Hz) in the Late condition, while the f0 peak was higher in N1 than in N2 in the Early (H1 = 266.2 Hz; H2 = 208.5 Hz) and Word (H1 = 268.5 Hz; H2 = 211.9 Hz) conditions. This indicates a trade-off relationship between the f0 peaks in N1 and N2. Graphical exploration revealed that, even when using f0 peaks, there is a large variation from speaker to speaker. Figure 4 shows the averaged normalized values for the five f0 measurement points (see § *2.1.3.*) in Late, Early and Word conditions for each speaker. The f0 normalization of the individual measurement values was obtained by a linear

24

transformation adapted by Truckenbrodt (2004). In the figure, the f0 values for each tone are rescaled between 0 (i.e., the minimum value within the pitch range of the speaker, corresponding to the L%) and 1 (i.e., the maximum value within the pitch range of the speaker, corresponding to the peak on N1). While some speakers (e.g., speaker *cv*) signal the difference between the Late and the other conditions solely by raising the f0 peak on N2, other speakers (e.g., *ts*) simultaneously decrease the height of the f0 peak on N1. In one speaker (*mw*), the Late condition is characterized by the raising of the f0 peak on N1, while there is no difference on the f0 peak on N2 across the boundary conditions.

Table 1. Frequency of the different edge tones in the corpus.

(%)	H		L	
	N1	N2	N1	N2
Late	98.6	79.2	1.4	20.8
Early	87.0	99.8	13.0	0.2
Word	84.5	99.3	15.5	0.7
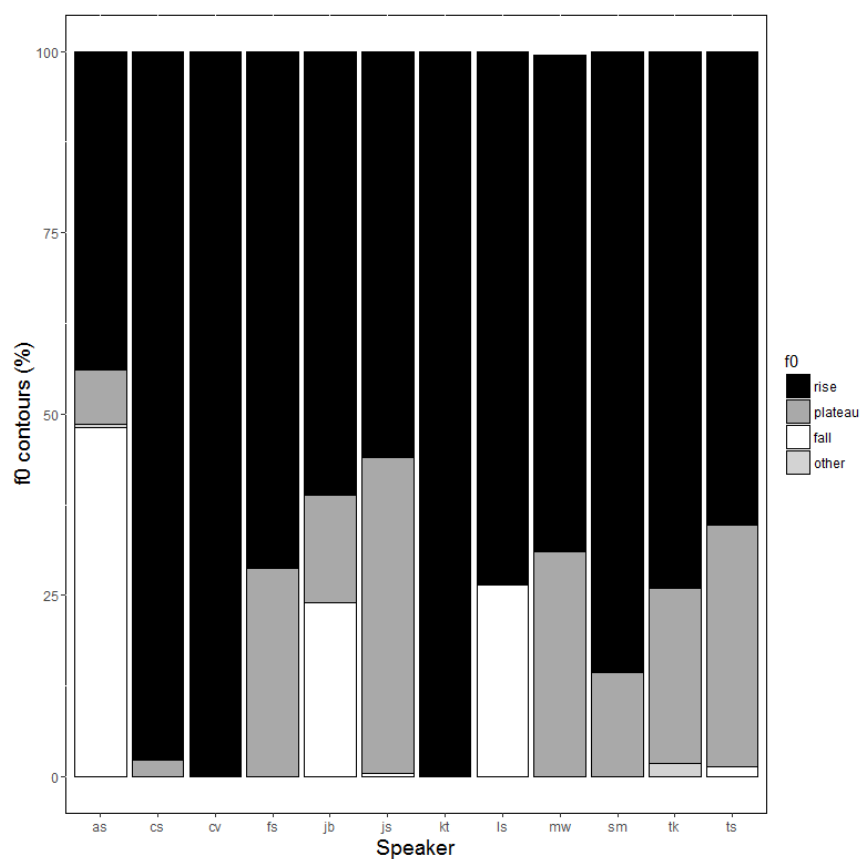
Figure 3. Distribution of the f0 contours (%) on the target names N1 and N2 (y-axis) separately for each

speaker (x-axis).
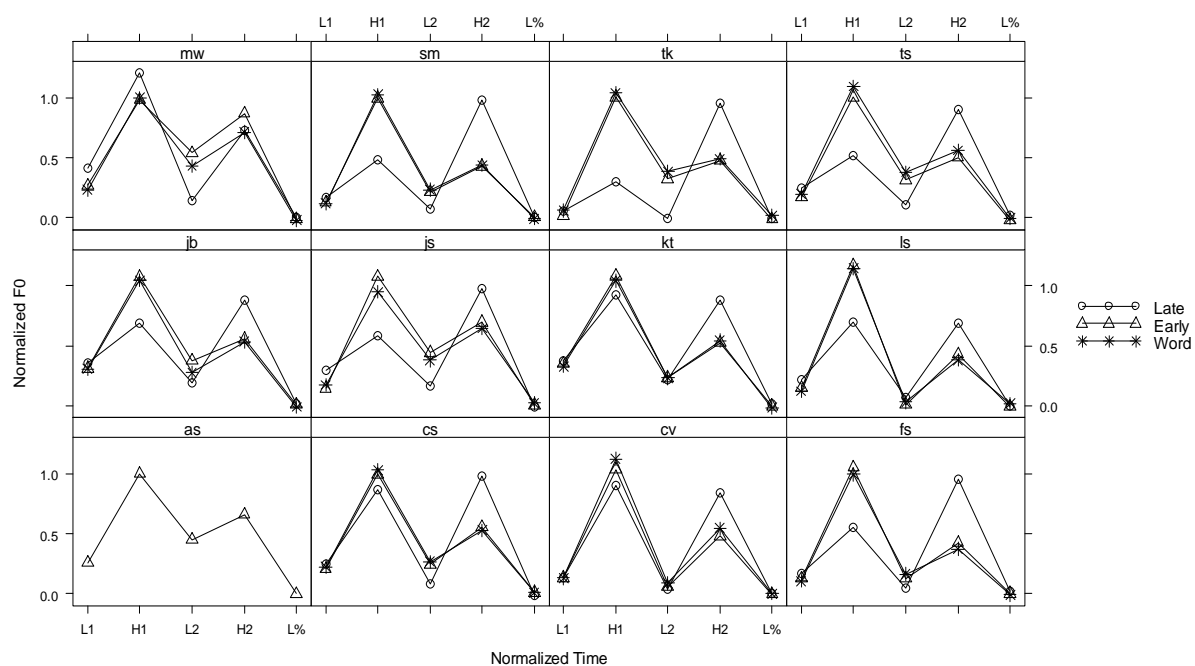
Figure 4. Normalized f0 (y-axis) across the five measurement points (x-axis) for the Early (triangles), Late

(circles) and Word (stars) boundary conditions. The bracketing patterns are: Late = (N1 N2) (N3); Early = (N1)

(N2 N3); Word = (N1) (N2 Ur-N3). The measurement points represent the f0 rises for both N1 (L1 and H1) and N2 (L2 and H2) as well as the utterance final L (L%). Results are separated by speakers. For speaker *as*, only data for the Early condition were analyzed since in the other two conditions the speaker systematically produced falling f0 movements instead of rising ones. Note that normalized values were used only for graphical exploration, while raw values were entered in the mixed models.

2.2.3. Pauses

Pauses occurred in 82.8% of the whole corpus and were mostly placed in the position of the bracket. Specifically, for the Late condition there were 336 pauses after N2 (77.7%) and 15 pauses after N1 (3.4%). For the Early condition, there were 350 pauses after N1 (81%) and only 4 after N2 (0.9%). For the Word condition, there were 348 pauses after N1 (80.5%), and 21 (4.8%) after N2. The number of pauses was significantly higher in the Early than in the Late condition after N1 [$\beta = 4.77$, SE = 0.29, z = 16.46, p < 0.001]. No differences were found between the Early and Word conditions. On the other hand, the number of pauses after N2 was significantly higher in the Late than in the Early condition [$\beta = 5.92$, SE = 0.51, z = 11.49, p < 0.001].

In addition to the distribution of pauses, Figure 5 shows differences in the duration of the pauses depending on the presence of a bracket. The pause after N1 and N2 was much longer when it occurred in presence of a bracket than in absence of a bracket. In fact, the mean pause duration in the position of the bracket was 119 ms (Late) and 174 ms (Early) and 192 ms (Word). For the few pauses occurring at no-bracket positions, the mean pause duration was 53 ms in the Late condition, 56 ms in the Early condition, and 57 ms in the Word condition. The statistical analyses showed that the contrast in pause duration between Late and Early conditions is significant for the pause after N1 [$\beta = -0.121$, SE = 0.20, t = -

6.06, p < .01, CI = -0.16; -0.08] and after N2 [β = 0. 081, SE = 0.02, t = 2.84, p < .01, CI = -0.02; 0.13]. We also found a small but significant contrast between Early and Word condition for N1 [β = 0.017, SE = 0.005, t = 3.21, p < .01, CI = 0.007; 0.028]. However, we doubt that this result has any functional relevance: as noted above, pauses after N1 in Word condition were a mere 18 ms longer than pauses after N1 in the Early condition.

We tentatively take pauses to be a cue of IPs but not of lower prosodic constituents (Nespor and Vogel, 1986). For example, in read speech of German sentences, Truckenbrodt (2002:85, Fn.8) found pauses regularly (for three speakers) or frequently (for a fourth speaker) at the boundary of two coordinated clauses–a canonical position of intonation phrase boundaries; clause-internal pauses at the positions of the many more PPh-boundaries were rare and exceptional. In the production results of Gollrad (2013), the pauses between sentences (intonation phrases) lasted on average 406 ms, while the clause-internal pauses that she measured had averages of 18 ms, 28 ms, 40 ms, 104 ms, and 39 ms, depending on the case. If pauses specifically cue IP-boundaries, the distribution of pauses found in the corpus further suggests that the brackets in our stimuli were produced as IP boundaries, at least in the large majority of cases.
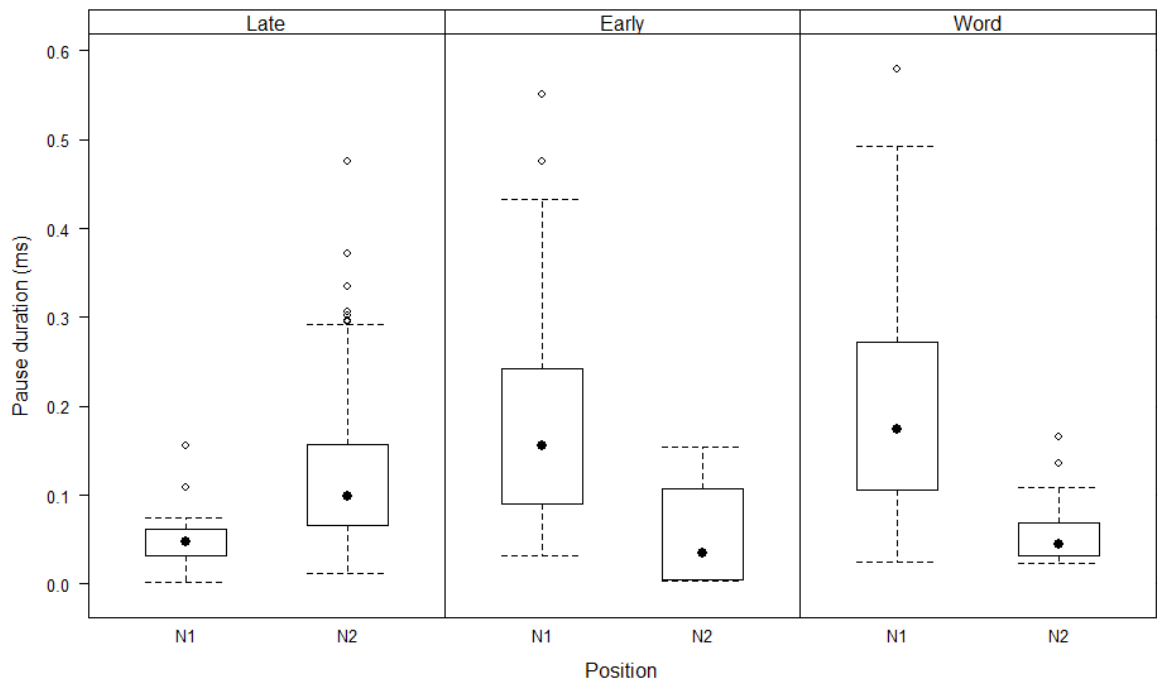
Figure 5. Boxplots for pause duration (y-axis) at the position after N1 and after N2 split across the three boundary conditions. The bracketing patterns are: Late = (N1 N2) (N3); Early = (N1) (N2 N3); Word = (N1) (N2 Ur-N3). Data are collapsed across speakers.

To sum up, the production experiment showed that differences in bracketing might be signaled by the simultaneous use of final lengthening, f0 and pauses. While pauses mainly occur at the edge of major prosodic constituents (such as IPs), final lengthening increases with boundary strength. As for f0, speaker-specific behavior appears to play a major role in the use of f0 cues to bracketing. The most common strategy is to produce single f0 rises, with the difference across boundary conditions being mainly signaled by the raising/lowering of the f0 peak on N1 and N2.

3. Perception study

We tested the perceptual relevance of the three acoustic parameters by running three perception experiments in three separate groups of listeners for final lengthening (10 listeners), f0 (12 listeners), and pause (13 listeners), respectively.

Each experiment consisted of a two-alternative forced-choice identification task in conjunction with measurements of reaction times. For each acoustic parameter, we created 2 continua of 10 manipulation steps, one created from a late boundary stimulus and the other one from an early boundary stimulus (i.e., a total of 20 stimuli). The natural stimuli were also included in the perception experiments, and their identification score was employed to set a criterion for listener inclusion in the data analysis. Since listeners correctly identified the natural stimuli, they were all assumed to reliably perform the task. The manipulations are described in detail in the following sections.

3.1. Methods

3.1.1. Bases for stimuli creation

To obtain base stimuli for the generation of the manipulated stimuli to be used in the perception experiments, we made additional recordings with a naive female German speaker who had particularly little creaky voice in her productions. She came from the Berlin-Brandenburg area and she was 25 years old. To minimize creakiness, proper names at N1 and N2 position ended with a high vowel (/i/) instead than a low one (/a/). In fact, glottalization phenomena seem to occur particularly with low vowels (Żygis, Brunner, and Moisik, 2012). From these productions we chose two natural stimuli (see Figure 6), one from the Early ("Early base") and one from the Late ("Late base") condition as bases for the following manipulations using PSOLA (*Pitch Synchronous Overlap and Add*, Carpentier and Moulines,

1990). Table 2 reports the acoustic values for the duration of the final vowels of N1 and N2 (for final lengthening), for the f0 peaks on N1 and N2 and for the duration of the pauses immediately following the two nouns for these two stimuli.
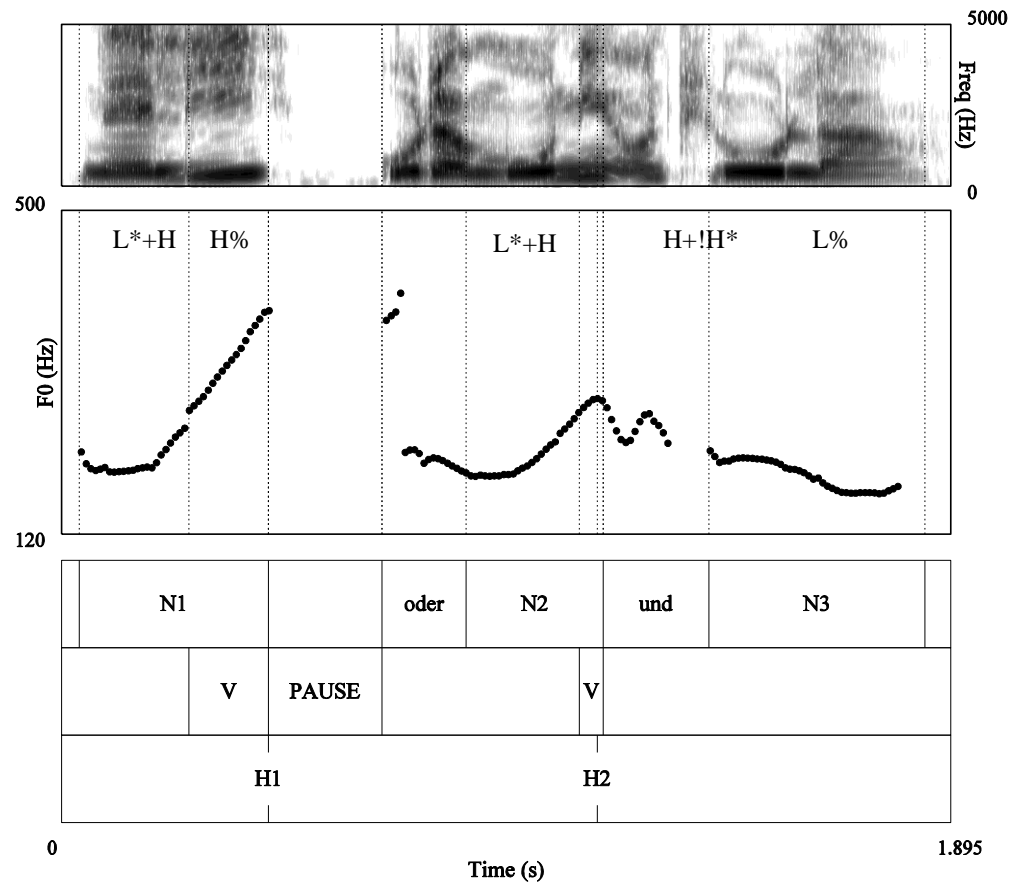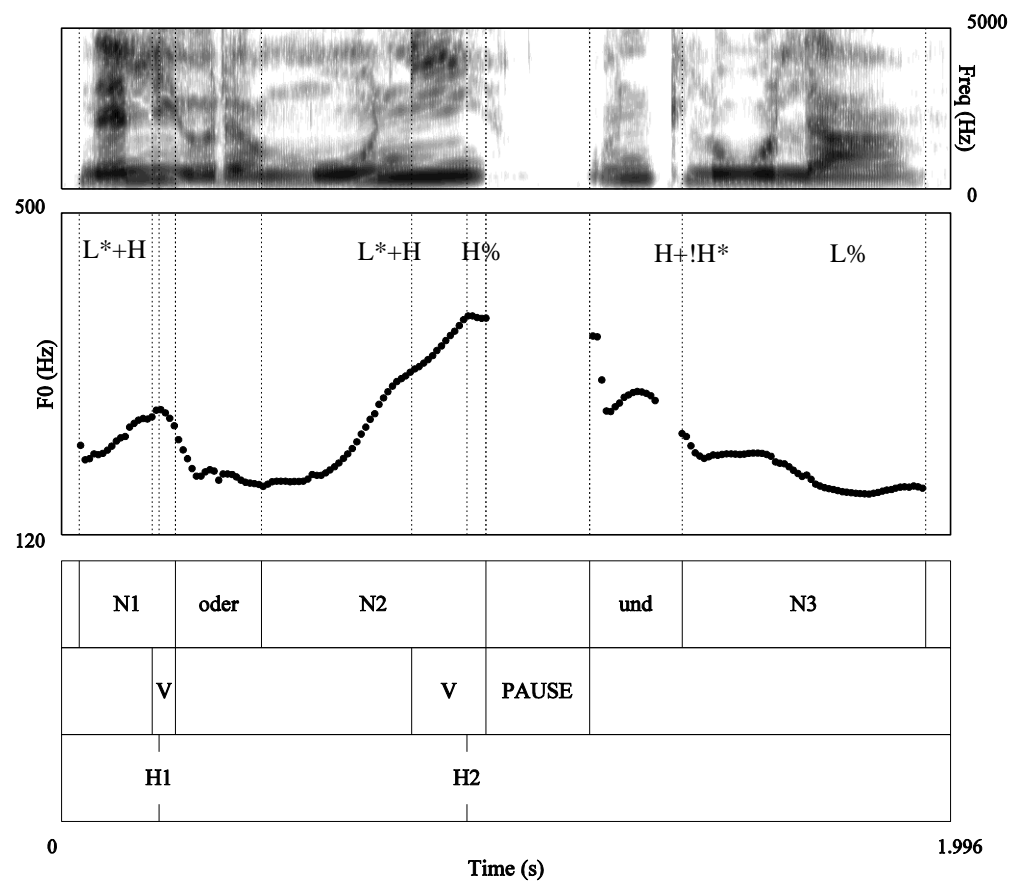
Figure 6. Spectrogram, f0 curve, tonal labeling and textgrid for stimuli with early (upper) and late (lower) boundary. Textgrids contain the labeling of the words (tier 1), that of the final vowels (tier 2) and of the f0 peaks (tier 3) of N1 and N2. The dashed lines indicate segmental boundaries and f0 peak temporal localization.

Table 2. Acoustic values in the natural stimuli chosen as bases for the perception manipulations. The diacritic "#" indicates that the pause was absent.

	Final vowel duration (ms)		f0 peak (Hz)		Pause duration (ms)	
	N1	N2	N1	N2	after N1	after N2
Late base	50	166	268.4	379.3	#	230
Early base	169	50	382.6	279.1	235	#

3.1.2. Stimuli: Final lengthening

In the first perception experiment, both pauses and f0 were neutralized to investigate the impact of final lengthening alone on boundary perception. Any pauses in the stimuli were deleted to neutralize this cue. To avoid a cue given by f0, it was flattened to an evenly falling contour that ranged from the utterance initial f0 value (218 Hz) to the utterance final f0 (178 Hz) of the speaker, averaged over the two stimulus bases. This neutralized the f0 cue as much as possible. We then applied length manipulations to the final vowels of N1 and of N2. Final lengthening was manipulated by creating a continuum of ten duration steps (from step 1 to step 10) from the Early base. We used steps of 15 ms for shortening the final vowel of N1 and of 14 ms for lengthening the final vowel of N2 in order to get values similar to the original productions at fixed steps of the continuum (step 2 and step 9, see Table 3). At the one end of the continuum, the final vowels of N1 and N2 were 181 ms and 45 ms long, respectively. At the other end of the continuum, the final vowel of N1 was 46 ms long and the final vowel of N2 was 171 ms long.

We performed the same manipulations on the Late base to neutralize pauses and f0 and created a 10 step final lengthening continuum from it in the same way and with the same step sizes as for the Early base. That is, we lengthened the final vowel on N1 and shortened the final vowel of N2 from step 10 to step 1. As a consequence, manipulations of both bases had a shorter vowel on N1 and a longer vowel on N2 at earlier steps of the continuum (e.g., step 1). On the other hand, the manipulations had longer vowel on N1 and shorter vowel on N2 at later steps of the continuum (e.g., step 10). Again, the duration manipulations at steps 2 and 9 were similar to the values of the natural productions. Prior to these manipulations, the conjunctions *oder* "or" and *und* "and" had been cross-spliced from the Late into the Early

base to minimize differences between the bases that might come from domain-initial strengthening (see, e.g., Keating, Cho, Fougeron, and Hsu, 2003).

Even though pauses and f0 cues to boundaries had been reduced from the base, there might be other acoustic parameters such as spectral tilt and length of other segments which remained in the base stimuli and which could be employed by listeners as secondary cues to bracketing. This is deemed to be harmless for the experimental design. However, the unknown effects of these remaining cues made it advisable to use stimuli from two bases as described.

To sum up, we worked with 2 bases, 10 steps of the manipulation and 7 repetitions of each stimulus. Across the 10 listeners, we obtained $140 \times 10 = 1400$ observations.

Table 3. Steps 1-10 of the final lengthening manipulation (ms) of the vowels at the end of N1 and N2. The values are separately shown for the Early and Late bases.

Steps of final lengthening manipulation (ms)										
Early base	1	2	3	4	5	6	7	8	9	10
Final vowel of N1	181	166	151	136	121	106	91	76	61	46
Final vowel of N2	45	59	73	87	101	115	129	143	157	171

Steps of final lengthening manipulation (ms)										
Late base	1	2	3	4	5	6	7	8	9	10
Final vowel of N1	180	165	150	135	120	105	90	75	60	45
Final vowel of N2	45	59	73	87	101	115	129	143	157	171

3.1.3. Stimuli: f0

In the second perception experiment, we investigated the impact of the height of the f0 peaks on boundary perception and hence used stimuli with neutralized final lengthening and pause cues. As base for the manipulation of f0, we chose two stimuli from the first perception experiment (manipulated for the lengthening of the final vowel) that had been judged by the participants as being perceptually ambiguous between early and late boundary interpretation. Specifically, for the Early base, we employed the stimulus at the step 6 of the final lengthening manipulation ("Early base - final lengthening step 6"); for the Late base, we employed the stimulus at the step 4 of the final lengthening manipulation ("Late base - final lengthening step 4"; see results in *§ 3.2.1*). Each of them served as basis for an f0 peak manipulation in 10 steps. These are shown in Figure 7.
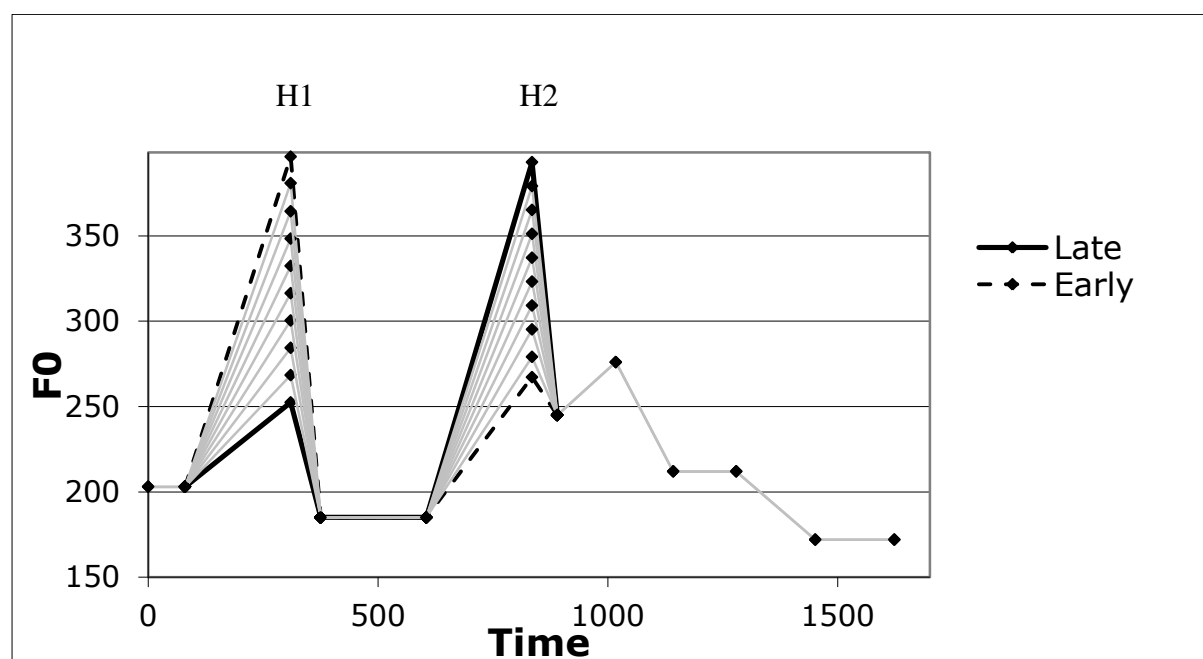
Figure 7. Stimuli of the f0 peak perception experiment. The f0 contour was obtained by the interpolation of 12 points (black). Points 3 and 6 (see text) correspond to the f0 peaks on N1 (H1) and N2 (H2), respectively. The dashed black line traces the highest peak on N1 that was combined with the lowest peak on N2 ("early boundary"). From there, the height of H1 was decreased and the height of H2 increased in steps, as indicated by

grey lines and dots. At the other end of the continuum, traced by a solid black line, the peak height on N1 was minimal and that of N2 was maximal ("late boundary").

First, differences in tonal alignment (i.e., the temporal synchronisation with the segmental or syllabic string) between the two bases were neutralized. This was done by choosing intermediate alignment values between those of the Early and Late bases. As shown in Figure 7, the f0 contour was obtained by the interpolation of 12 points: 1) at utterance onset; 2) at the f0 minimum in N1 (L1); 3) at the f0 peak in N1 (H1); 4) at the end of the fall from H1; 5) at the f0 minimum in N2 (L2); 6) at the f0 peak in N2 (H2); 7) and 8) at the start and end of the f0 rise at the conjunction "und"; 9) and 10) at the start and end of the high plateau; 11) at f0 minimum at the end of the fall from the plateau and 12) at the utterance end. The region from one point to the next one was interpolated by straight lines. Points (2-3) and (4-5) reproduced the L*+H accents in the first two names. Points (7-10) reproduced the H+!H* accent realized on the third name (cf. Grice, Baumann, and Jagdfeld, 2009). Apart from points at H1 and H2 location (i.e., points 3 and 6), the other points were kept at the same f0 value for all stimuli. Specifically, the f0 values at points (1, 2, 4, 5, 7, 8) were intermediate between values from the Early and Late base stimuli. The nuclear H was characterized as a high plateau, whose starting and end points had the same f0 value.

Following these initial modifications, the height of f0 peaks on N1 and N2 were simultaneously manipulated. For "Early base - final lengthening step 6", we progressively lowered the value of the f0 peak on N1 and increased that of the f0 peak on N2 in ten steps of f0 manipulation (steps 1-10). The step size was 16 Hz for the f0 peak on N1 and 14 Hz for that on N2 in order to get values similar to the original productions at fixed steps of the continuum. The f0 manipulations at steps 2 and 9 corresponded to values similar to those found in the natural productions. Similarly, for "Late base - final lengthening step 4", we progressively increased the f0 value of the peak on N1 and decreased that of the peak on N2

from step 10 to step 1. The step sizes were the same as for the other base. Then, at step 1 stimuli from both bases had higher f0 peak on N1 and lower f0 peak on N2, while at step 10 stimuli from both bases had lower f0 peak on N1 and higher f0 peak on N2. The f0 peak values for each manipulation step are shown in Table 4.

To sum up, we worked again with 2 bases, 10 steps of the manipulation, and 7 repetitions of each stimulus. As 12 listeners took part in the second perception experiment, this yielded a total of 1680 observations.

Table 4. Steps 1-10 of the manipulation (Hz) of the f0 peaks in N1 and N2. The values are separately shown for the Early and the Late bases.

Steps of F0 peak manipulation (Hz)										
Early base	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
F0 peak in N1	396.9	380.9	364.9	348.9	332.9	316.9	300.9	284.9	268.9	252.9
F0 peak in N2	265.1	279.1	293.1	307.1	321.1	335.1	349.1	363.1	377.1	391.1

Steps of F0 peak manipulation (Hz)										
Late base	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
F0 peak in N1	396.4	380.4	364.4	348.4	332.4	316.4	300.4	284.4	268.4	252.4
F0 peak in N2	267.2	281.2	295.2	309.2	323.2	337.2	351.2	365.2	379.2	393.2

3.1.4. Stimuli: Pauses

The third experiment aimed at investigating the perception of the pause cue when f0 and lengthening cues are perceptually ambiguous. From the results obtained in the f0 experiment,

we were able to choose "Early base - final lengthening step 6-f0 peak step 3" and "Late base - final lengthening step 4-f0 peak step 4" both being neutral between Early and Late judgments. From these, we created ten pause manipulations each (steps 1-10). For the Early base, at one end of the continuum, N1 was followed by a pause of length 243 ms, and N2 was not followed by a pause. Each step of the continuum decreased the length of the pause after N1 and increased the length of a pause after N2. At the other end of the continuum, N1 was not followed by a pause and N2 was followed by a silent pause of length 243 ms. Analoguously, for the Late base we progressively increased the pause after N1 and decreased the pause after N2 from step 10 to step 1. Step size was of 26 ms in order to get pause durations similar to the natural productions at fixed steps of the continuum. The values of pause duration of the manipulated stimuli are provided in Table 5.

To sum up, in the same manner as in the preceding two experiments, 10 manipulations of 2 bases were run with 7 repetitions and 13 listeners (1820 observations).

Table 5. Steps 1-10 of the manipulation (ms) of the pauses after N1 and N2. The values are separately shown for the Early and the Late bases.

Steps of pause manipulation (ms)										
Early base	1	2	3	4	5	6	7	8	9	10
Pause after N1	243	216	189	162	135	108	81	54	27	0
Pause after N2	0	27	54	81	108	135	162	189	216	243

Steps of pause manipulation (ms)

38

Late base	1	2	3	4	5	6	7	8	9	10
Pause after N1	234	208	182	156	130	104	78	52	26	0
Pause after N2	0	23	49	75	101	127	153	179	205	231

3.1.5. Participants

Thirty-five monolingually raised German native speakers (aged between 20 and 30 years) participated in the perception study and were reimbursed for their time. They came from the Berlin-Brandenburg area and were university students in Potsdam or Berlin. Specifically, 10 listeners (7 women) participated in the final lengthening experiment, 12 listeners (10 women) took part in the f0 perception experiment, and 13 listeners (11 women) in the pause perception experiment.

3.1.6. Procedure

The experimental setup was the same for all three perception experiments. A computer presentation was prepared in which the 20 stimuli occurred seven times in pseudo-randomized order, for a total of 140 presentations. During the experiment, each stimulus was auditorily presented to the listener, together with a visual display of two bracketings of the same sentence. Half of the listeners were presented with the early bracketing on the left, and the late bracketing on the right of the screen. For the other half the presentation positions of the two bracketings were vice versa. The listeners had to report which bracketing better matched their auditory impression of the sentence by pressing one of two keys on the keyboard that corresponded to the bracketing presented on the left or the right of the screen. They were instructed to carefully listen to the stimuli and to press the keys as quickly as

possible. Listeners did not receive any feedback on their answers. The experiment started with five practice trials followed by seven blocks of twenty pseudo-randomized test trials. A listening session lasted on average 15 minutes. All three perception experiments were run by means of the DMDX software (Forster and Forster, 2003).

To analyze these results statistically, a series of logit models with mixed effects was run separately on the identification score (i.e., the late boundary judgments) for the pause, final lengthening, and f0 peaks manipulations. Moreover, a series of quadratic models with mixed effects was run separately on reaction times for the three acoustic parameters. Reaction times were measured from stimulus onset. Prior to the statistical analysis, RT values were logarithmically transformed to correct for anormal distribution.

The logit models are based on binomial distributions (z-scores, Generalized Linear Model, GLM). This allows modelling binary decisions like two-alternative forced-choices (Baayen, 2008). The mean 50% crossover boundary was also measured, since it is taken to be the location of the category boundary shift between the early and late judgments. The crossover boundary of the logistic functions fitted to the responses was calculated separately for the Early and Late bases, for each of the three prosodic cues. As for reaction times (RTs), we expected longer reaction times at the category boundary (where identification might be more ambiguous between the early and the late bracketing) than at the endpoints of the continua (where identification of the bracketing should be less ambiguous). As a consequence, reaction time values across manipulation steps should be described by an upward parabola whose inflection point would coincide with the location within the continuum corresponding to the category boundary shift. If this is true, the reaction time function should be more accurately modelled through a non-linear regression than through a linear one, that is, a quadratic polynomial regression (Baayen, 2008). A quadratic term was thus introduced to model the nonlinear relation between the RT values and the manipulation

steps (note that, because of such expectations from categorical perception, possible linear effects of STEPS on RT are uninteresting for our research questions and will be reported in the Appendix).

The logit and quadratic models were applied here to test the effects of the factors STEP (1-10) and BASE (Early vs. Late) on identification scores (expressed in terms of late boundary judgment scores) and reaction times, respectively. The factor STEP was entered as a numerical variable for each prosodic cue and centered at the mean value. LISTENERS and REPETITION were included as random intercepts. REPETITION was considered as a random term since it was not controlled in a systematic way during the experimental session (e.g., stimuli were presented in a random order and at irregular temporal intervals). The models had maximal random structure. Backward elimination of non-significant variables (such as for interactions among fixed factors or random slopes) was applied. A likelihood-ratio test compared the more complex model to the simpler one, in order to check if removals were warranted.

Finally, the results of the late boundary judgments and reaction times from the three experiments on final lengthening, f0, and pauses were statistically compared through mixed models. This is aimed at revealing the relative strength of each cue as well as possible differences in task difficulty. Note that the fact that different listeners participated in the three experiments (nested design) was taken into account in the models. For continuous dependent variables (reaction times), we employed bootstrapping methods with replacement from the original sample (number of sample = 500) to estimate p-values and 95% confidence intervals (CI) for each mixed model. For discrete dependent variables (the late boundary judgments), we report the p-value of the logit models. The cut-off point for significance was set at $p < 0.05$.

3.2. Results

3.2.1. Final lengthening

Figure 8a-b shows the percentage of late boundary judgments and reaction times (ms) for the manipulations of final lengthening, plotted separately for the stimuli derived from the Late base and for the stimuli derived from the Early base. Stimuli with more lengthening on the final vowel of N2 and, at the same time, less lengthening on the final vowel of N1 are plotted further to the right on the x-axis. As can be seen in Figure 8a, these stimuli trigger the perception of the strongest boundary after N2, which is plotted higher on the y-axis. Inversely, stimuli containing more lengthening on the final vowel of N1 and, at the same time, less lengthening on the final vowel of N2 plotted further to the left, are perceived as having the strongest boundary after N1. This effect of STEP (i.e., the final lengthening manipulation) is statistically significant [$\beta = 0.29$, SE = 0.08, z = 3.52, p < 0.001]. It should be also noted from Figure 8a that the transition phase from a clear identification of early to a clear identification of late boundaries is quite broad. That is, for the Late base, it spans stimuli at steps 3-5. Similarly, for the Early base, it spans stimuli at steps 5-7.

The plot also shows that the perceptual shift from the late to the early boundary judgments is modulated by the base stimuli in spite of the neutralizations we applied (i.e., the two lines in Figure 8a do not overlap). For instance, the stimuli created from the Early base show only 12% late boundary judgments at step 1 and they reach only 60% late boundary judgments at step 10. On the other hand, the stimuli created from the Late base show an overall higher percentage of late boundary judgments, with the score being already at 45% at step 1 and 76% at step 10 (see also Appendix, Table A.4). This effect of BASE is significant

[β = 0.76, SE = 0.12, z = 6.30, p < 0.001]. The results confirm that there are remaining boundary cues in the bases that were not captured by our neutralization procedure. This is an expected effect that is orthogonal to our research question. There was no significant interaction of BASE with STEP.

As for the 50% crossover boundary, it was different across the bases. Specifically, the inflection point of the estimated logistic function was between steps 6 and 7 (estimated mean= 6.8) in the Early base and between steps 3 and 4 (estimated mean= 3.4) in the Late base. These results are in line with the observed means in Figure 8a. As a consequence, we selected stimuli 6 and 4 as bases for the following F0 manipulations, respectively for the Early and Late bases (see above).

Figure 8b shows the mean reaction times across manipulation steps. We expected longer reaction times for manipulation steps in vicinity to category boundary shifts than at the other manipulation steps. However, the data did not confirm such predictions. In the Early base, there was a gradual but non linear (but non linear) reaction time increase up to step 6. In the Late base, two visible reaction time peaks were found, a smaller one at step 2 and a bigger one at step 5. The mixed models showed, for both bases, that the quadratic term for the factor STEP did not reach significance. This means that RT differences across the continuum cannot be described by an upward parabola (whose inflection point, corresponding to a RT peak, was expected at the category boundary shift).

In sum, these results show that final lengthening is a perceptually relevant cue for the perception of prosodic boundaries in German and that the magnitude of this effect varies with base type. However, the late boundary judgments indicate a slow perceptual change from one end to the other end of the continua, with no clear local maxima in the reaction times.

(a)   **(b)**

43
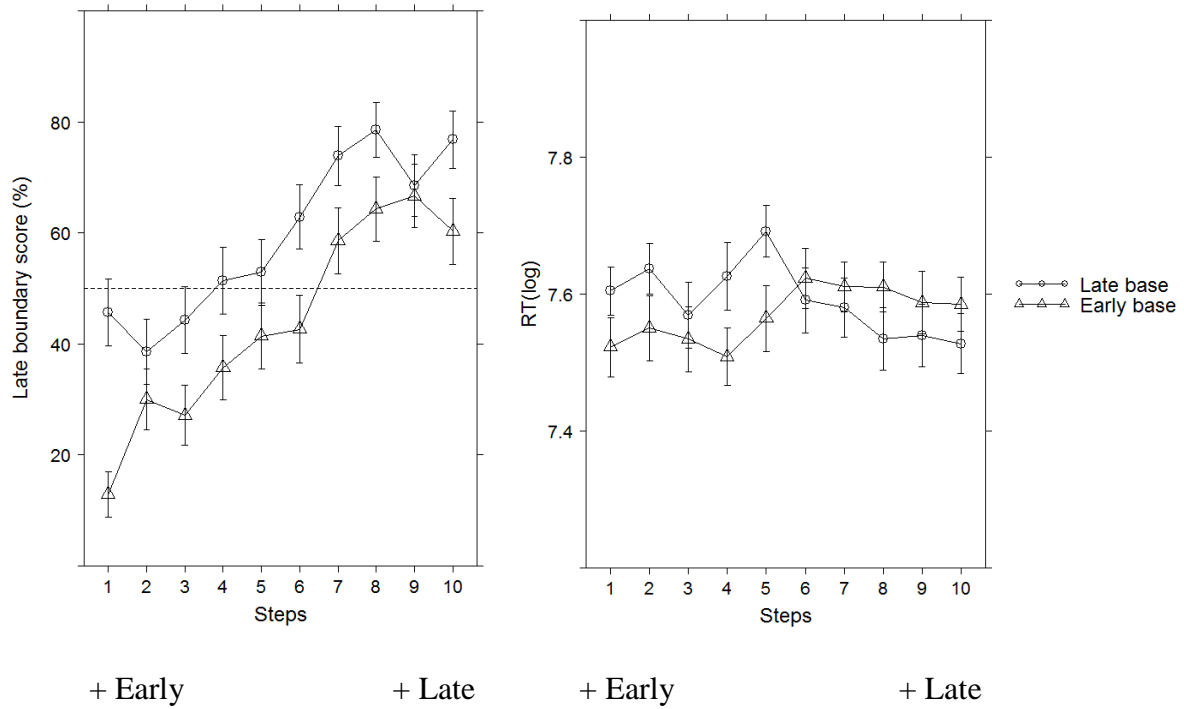
Figure 8. Means and standard errors for late boundary judgments (Figure 8a, left panel) and reaction times (Figure 8b, right panel) across the duration manipulation steps. Step 1 corresponds to stimuli with relatively more lengthening on the final vowel of N1 and relatively less lengthening on the final vowel of N2, thus matching acoustic findings for the early boundary. Step 10 has the opposite pattern, thus corresponding to the late boundary. The results are separately shown for the Late (circles) and the Early (triangles) base stimuli. The dashed line in the left panel corresponds to the chance level.

3.2.2. f0

The results for the perception of f0 are plotted in Figure 9a (for the late boundary judgments) and 9b (for RT), again separated for the two bases. Here, relatively higher f0 at the end of N1, with a relatively lower f0 at the end of N2, is plotted towards the right on the x-axis. The corresponding perception of the largest boundary after N1 is plotted higher on the y-axis. Similarly to the results for lengthening, Figure 9a shows that the percentage for late boundary judgments increases from step 1 to step 10 (see also Appendix, Table A.5). However,

44

judgments seem to be biased towards late boundary judgments, independent of the base type. For instance, at step 1 the mean score already reaches 39% for the Early base and 53% for the Late base stimuli. At step 10, the score is around 84% for both base stimuli. As a consequence, the response curves in Figure 9a are very shallow, with slow perceptual changes from Early to Late identification across the steps of manipulation. The effect of STEP (i.e., the f0 manipulation) on the perception of the position of the largest boundary is significant, meaning that stimuli with a relatively low f0 peak on N1 and a relatively high peak on N2 are perceived with a late boundary [$\beta = 0.26$, SE = 0.08, z = 3.06, p = 0.002]. The effect of BASE varies across the continuum being smaller at steps located at the edge of the continuum than at the steps in the middle of the continuum. The main effect of BASE was also significant [$\beta = 0.59$, SE = 0.28, z = 2.09, p = 0.036]. There is no interaction of the factors BASE and STEP.

The measurement of the 50% crossover boundary indicates that the inflection point of the estimated function was between steps 3 and 4 (estimated mean = 3.1) for the Early base. This is congruent with Figure 9a, in which the stimuli at steps 3 and 4 were the closest at 50%. The stimulus 4 at step 4 was chosen as basis for the third experiment on pause perception. For the Late base, the 50% crossover boundary was at step 1 (estimated mean = 0.99). However, graphical comparison between the estimated and observed mean scores suggested that this value was overestimated for the Late base. As a consequence the stimulus at the following step 3 was chosen as basis for the perception experiment on pauses (see above).

Figure 9b shows the reaction time data across manipulation steps. For the Late base, the gradual changes in identification are accompanied by a flat course of reaction times. For the Early base, reaction times are relatively short at the edges of the continuum (e.g., RT = 2.07 s at step 1; RT = 2.04 s at step 7) while a single reaction time peak is found at step 5 (RT

= 2.30 s). However, differently from our expectations, mean reaction times were shorter at the stimuli which were closer to the crossover boundary region (step 3 = 2.24 s.; step 4 = 2.13 s). The statistical analysis confirmed that the quadratic term for STEP was significant for the Early base [β = -1.04, SE = 0.27, t = -3.73, p = .002, CI = -1.53; -0.49]. The presence of an interaction STEP by BASE indicates that the effect of STEP in the Late base was significantly different from its effect in the Early base [β = 1.06, SE = 0.39, t = 2.93, p = .004, CI = 0.39; 1.94]. Indeed, when releveling the model (by changing the reference level for BASE to the Late Base), the effect of STEP is not significant, indicating that, differently from the Early base, the quadratic term cannot be applied for the Late base.

These results show that the increase of an f0 rise is likely to trigger the perception of a prosodic boundary in bracketed lists. Results from the identification task and reaction time measurements suggest a gradual transition from one end to the other end of the phonetic continua. Compared to the results of final lengthening, the amount of the late boundary judgments is relatively high already at very early steps of the continuum. For the Early base stimuli, such a high score might be due to a perceptual interaction between final lengthening and f0 cues. Remember that for the f0 experiment, we chose base stimuli which have been judged ambiguous as for final lengthening in the first perception experiment. As a consequence, the duration of the final vowels in N1 and N2 was the same as at duration manipulation step 6 for the Early base stimuli ("Early base - final lengthening step 6"), which is two steps further than the duration of the same vowels for the Late base stimuli ("Late base - final lengthening step 4"). In the first experiment, the f0 was flattened over the course of the sentence. In the second experiment, it is possible that, for the Early base stimuli, the addition of an f0 rise on N2 would have further increased the percept of a late boundary, already at early steps of the continuum. For the Late base stimuli, remaining boundary cues in the base which were not neutralized by our manipulation could have further biased the judgments

towards the late boundary interpretation. We will return to the issue of relevant prosodic constituents signaled by f0 in the discussion session.
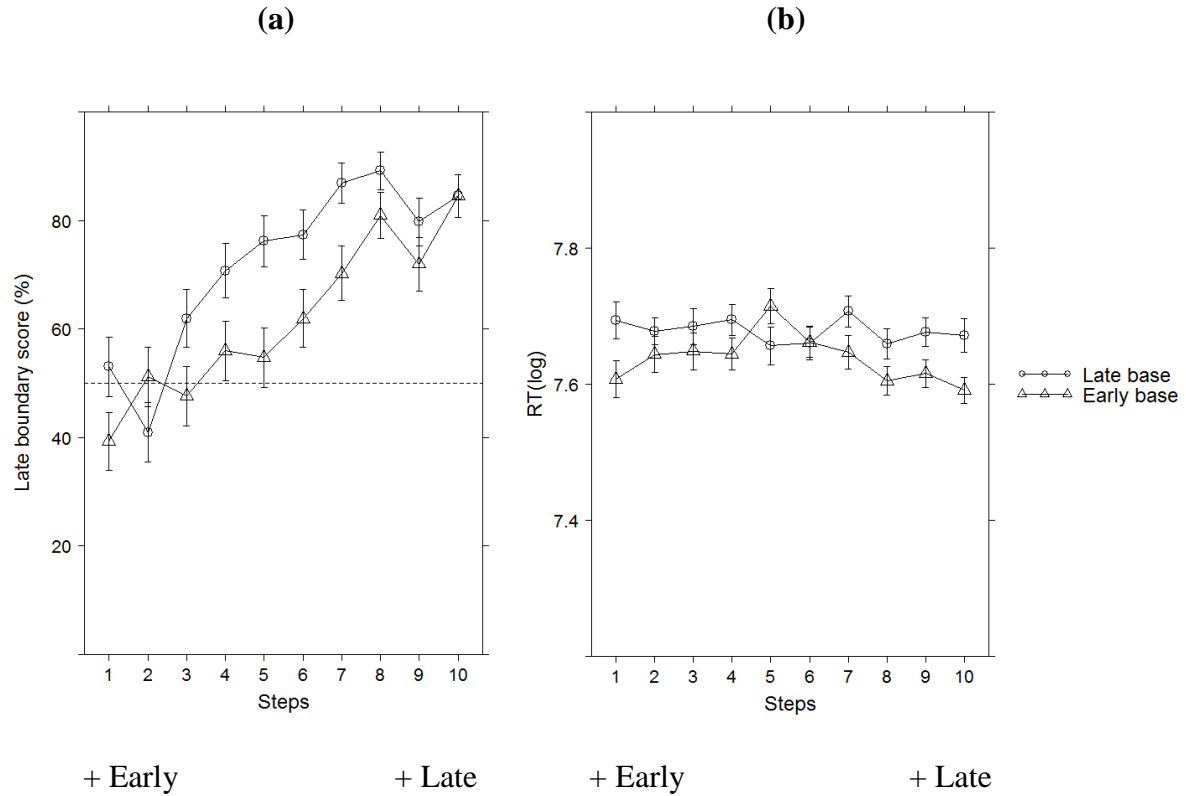
Figure 9. Means and standard errors for the late boundary judgments (Figure 9a, left panel) and reaction times (Figure 9b, right panel) across the f0 peak manipulation steps. Step 1 corresponds to stimuli with relatively higher f0 peak on N1 and relatively lower f0 peak on N2, thus matching acoustic findings for the early boundary. Step 10 has the opposite pattern, thus corresponding to the late boundary. The results are separately shown for the Late (circles) and the Early (triangles) base stimuli. The dashed line in the left panel corresponds to the chance level.

3.2.3. Pauses

Figure 10a-b shows the results of the perception experiment concerning the role of pauses in triggering boundary perception. Relatively longer pauses after N1, in combination with relatively shorter pauses following N2, are plotted further to the right on the x-axis. It can be seen that the percentage of late boundary judgments increases as a function of pause manipulation, with scores between 12.5% (i.e., for the stimulus from the Early base at pause manipulation step 1) and 87.5% (i.e., for the stimulus from the Late base at step 10; cf Appendix, Table A.6). Unlike the results of final lengthening and f0 manipulation, the response curves for pauses are very steep. The effect of STEP (i.e., the pause manipulation) is statistically significant [$\beta = 0.46$, SE = 0.09, z = 4.81, p < 0.001]. There is no significant effect of BASE. This is further confirmed by the analysis of the crossover boundary, which is similar across the two bases (estimated means: 6.9 for the Early base and 6.8 for the Late base). However, the interaction of BASE with STEP was significant [$\beta = 0.15$, SE = 0.04, z = 3.38, p < 0.001]. This interaction is due to the fact that, at early steps of the continuum, stimuli from the Early base scored slightly higher than stimuli from the Late base, while at later steps of the continuum, the difference between the two bases is much smaller. Results from the identification task thus show that pause duration is a perceptually relevant boundary cue.

Figure 10b shows RT data across manipulation steps. For both bases, the stimulus at step 6 (which is in the region of the estimated crossover boundary, i.e. 6.8 for the Early base and 6.7 for the Late base) coincides with an increase in mean reaction time (RT = 2.40 s). The statistical analysis confirmed the parabolic shape of the RT function [$\beta = -1.04$, SE = 0.25 t = -4.06, p < .01, CI = -1.54, -0.57], with no difference across the two bases. Hence, the patterns for the late boundary judgments and RT suggest that the shift from the late to the early judgments are more abrupt when pause is manipulated, while the shift is more gradual when final lengthening and f0 are manipulated.

(a) **(b)**

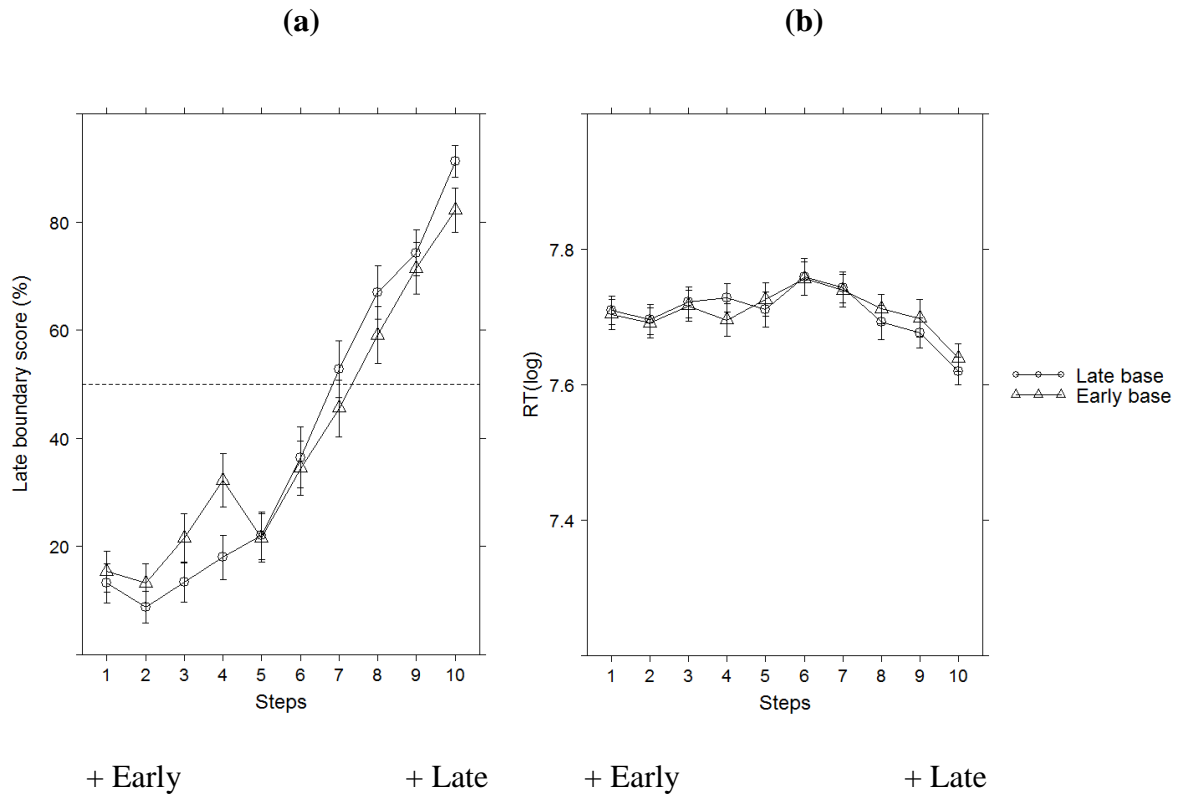+ Early   + Late   + Early   + Late

Figure 10. Means and standard errors for the late boundaryjudgments (Figure 10a, left panel) and reaction times (Figure 10b, right panel) across the pause manipulation steps. Step 1 corresponds to stimuli with a long pause after N1 and no pause after N2, thus matching acoustic findings for the early boundary. Step 10 has the opposite pattern, thus corresponding to the late boundary. The results are separately shown for the Late (circles) and the Early (triangles) base stimuli. The dashed line in the left panel corresponds to the chance level.

3.2.4. Comparison of the three perception experiments

A comparison in the score of the late boundary judgments and in reaction times was also run between the three experiments. The percentage of late boundary judgments (averaged across STEP and BASE) was higher for the f0 manipulation than for the final lengthening [β = -0.97, SE = 0.19, z = -5.01, p < 0.001] and pause manipulations [β = -1.04, SE = 0.18, z = -5.65, p <

0.001]. However, RT values were not significantly different across the three cues manipulations.

To sum up, the three perception experiments investigated the isolated contribution of final lengthening, f0, and pauses on the perception of bracketed lists. We found that the three cues are systematically employed by German listeners, though the magnitude of the effect differs across the three manipulations. Specifically, differences in late boundary judgments across the continuum were larger for the pause and final lengthening manipulations than for the f0 manipulations (where the percentage for late boundary is relatively higher, already at early steps of the continuum). Moreover, the response functions of the identification tasks are characterized by gradual perceptual changes from the early to the late boundary for final lengthening and f0. The response function is more abrupt for pauses. Similarly, for reaction times, RT peaks at the category boundary are found only for pause perception, but not for final lengthening and f0 perception. We will return on this issue in the discussion session. Finally, base type effects for the final lengthening and f0 manipulations suggest the presence of other auditory cues beyond the three acoustic properties investigated here. Such base type effects are absent for the pause manipulation, suggesting that pause is a strong marker to bracketing, overriding any secondary cue left in the stimuli.

We now turn to theoretical implications of the results of the production and perception experiments.

4. Discussion of the experimental findings

4.1. Production study

The three boundary cues investigated in this paper, final lengthening, f0, and pauses, were robustly produced in the production experiment presented here. This confirms findings of other production experiments on German including Gollrad et al. (2013), Féry and Kentner (2010), and Kentner and Féry (2013), which were also based on laboratory speech. Moreover, it confirms the corpus results of Peters et al. (2005), where these cues were shown to be robustly present in spontaneous speech.

In what follows, we aim at comparing the expected prosodic structures (as illustrated in *§ 1.3.*, see above) with the prosodic structures which were actually realized in the three experimental conditions *§ 4.1.1, § 4.1.2)*. We propose that deviations from the expected patterns might be due to the specific nature of the experiment. Following the literature, such task-specific effects can be captured by the *proximity* principle (Féry and Kentner, 2010; Kentner and Féry, 2013).

4.1.1. Evidence for intonation phrase boundaries

We believe that it is sensible to analyze the prosodic divisions that our speakers employed for the bracketed divisions as IP-boundaries, since about 80% of the brackets to be conveyed by our participants in the production experiment were cued by pauses. Moreover, the names at the bracketed position received increased final lengthening and higher f0 peaks (which is evidence of an H% intonation phrase edge tone. The high frequency of the occurrence of pauses (in comparison, for example, to the 38% of IP boundaries accompanied by pauses in Peters et al., 2005) may well be task-specific. While Peters et al. (2005)'s study was based on spontaneous speech, our task gave the speakers reasons to mark the IP boundaries in a maximally recoverable way, which is a possible incentive to maximize the use of cues for indicating the relevant boundaries.

4.1.2. The absence of internal phonological phrase boundaries

There are several indications in our data that suggest that the majority of our speakers deviated from the prosody we had hypothesized in one respect. It seems that they did not employ PPh boundaries at IP-internal positions. It seems instead that both PPhs and IPs were recruited for conveying the bracketing that was part of the task of the speakers. We think that the realized patterns for Late, Early and Word conditions are as follows. In these structures, the absence of a boundary that we expected based on the standard syntax-prosody mapping is illustrated by crossing out:

Late: *(N1 N2) N3*

[		]$_{IP}$	[		]$_{IP}$	
(~~x)$_{PPh}$ (	x	)$_{PPh}$	(	x	)$_{PPh}$	
(x)$_{PW}$ (	x	)$_{PW}$	(	x	)$_{PW}$	prosodic structure
[[ ]$_{DP}$ oder [		]$_{DP}$]$_{DP}$ und [		]$_{DP}$		syntactic DPs
[Lola]$_{NP}$ oder [Mona]$_{NP}$			und [Lena]$_{NP}$			syntactic NPs
(		)				bracket in the stimuli

Early: *N1 (N2 N3)*

[	]$_{IP}$	[			]$_{IP}$	
(x)$_{PPh}$ (		~~x)$_{PPh}$ (	x	)$_{PPh}$		
(x)$_{PW}$ (	x	)$_{PW}$ (	x	)$_{PW}$	prosodic structure

[   ]$_{DP}$ oder [[   ]$_{DP}$ und [   ]$_{DP}$]$_{DP}$ syntactic DPs

[Lola]$_{NP}$ oder [Mona]$_{NP}$ und [Lena]$_{NP}$ syntactic NPs

(   )   bracket in the stimuli

Word: control

[   ]$_{IP}$ [   ]$_{IP}$

(x   )$_{PPh}$ (x   )$_{PPh}$

(x   )$_{PW}$ (x   )$_{PW}$ (x   )$_{PW}$ prosodic structure

[   ]$_{DP}$ oder [   ]$_{DP}$ syntactic DPs

[Lola]$_{NP}$ oder [Mona   Urlena]$_{NP}$ syntactic NPs

(   )   bracket in the stimuli

The first clear indication of the absence of the expected PPh boundaries is the absence of acoustic difference at the crucial position between the Early and the Word conditions. That is, neither the final lengthening results nor the f0 results gave any indication of a stronger boundary after N2 (*Mona* in our example) in the Early condition relative to the Word condition.

The second clear indication relates to tonal height in the results of the f0 rises of many of the speakers. Earlier experiments on German (Grabe, 1998; Truckenbrodt, 2002, 2004, 2007a; Féry and Kügler, 2008) recorded sequences of PPhs without a bracketing task. Results showed that the first f0 peak in the first PPh was either the highest peak of the utterance or at least not much lower than any later peak. In the present experiment, tonal height results of speakers *cs* and *cv* (see Fig. 4) confirm these previous findings: The first peak is high. The second peak is lower (downstepped) if there is no boundary immediately following it. The

second peak is comparable in height to the first if a boundary immediately follows it (and if the boundary is not final). However, as can be seen in Figure 3, many of the speakers in our experiment (*sm*, *tk*, *ts*, *ib*, *is*, *fs*) produced a very low first peak followed by a considerably higher second peak in the Late condition. A possible interpretation is that the first peak does not correspond to the strongest stress of a PPh but only to word-stress instead. This would mean that the bracketed constituent of *(Lola oder Mona)(und Lena)* was realized as a single PPh, with its strongest stress on *Mona* (rightmost in the PPh as expected). This is compatible with the observation that these speakers also produced fairly low peaks in medial position in the Early condition (*((Lola)(oder Mona und Lena))*) and that these were as low as in the Word condition.

Thus, we think that, in both conditions, our speakers suppressed the PPh boundary internal to the intonation phrase relative to the phrasing that would be obtained by a standard syntax-prosody mapping. Such a deviation from expectations can be due to task-specific constraints. In our experiment, speakers' attention was drawn to the bracketing distinctions. Conscious and successful efforts were made to disambiguate the bracketing in a way that is recoverable for the addressee, in that speakers assigned a felicitous prosody to the intended syntactic structure. This is in line with the idea that a certain communicative situation requires more robust prosodic cues to prosodic boundaries which would enhance the intended linguistic message ("audience design", cf. Speer et al. 2011). Crucially, in our experiment, maximization of contrasts is achieved not only by the addition of major prosodic constituents at syntactic junctures, but also to the deletion of prosodic constituents that would otherwise be assigned.

Our experiment does not allow teasing apart to what extent the effects of prosody for syntactic disambiguation depend on the grammar and task-specific constraints, as we did not manipulate situational factors such as differences in task demands. However, similar results

were obtained by Féry and Kentner (2010) and Kentner and Féry (2013). In their study, they also found that the task of rendering a bracket around two names reduced the prosodic cues after the first of the two names, relative to an unbracketed rendition. Their baseline was an unbracketed list of names, for which they found a high initial peak with following downstep. Bracketing of the first two names as in [(N1 conj N2) conj N3 (...)] showed an initial peak considerably reduced in height. Féry and Kentner (2010) and Kentner and Féry (2013) formulate a principle of *proximity* that has the effect of "weakening" a prosodic boundary on a left branch. The *proximity* principle states (Kentner and Féry, 2013):

The prosodic boundary at the terminal constituent x is weakened if the following terminal constituent y is the sister of x or dominated by the sister of x – unless x is immediately dominated by the root node of the domain under consideration.

We think that our results are in accordance with this principle, especially concerning the first part of the formulation. Our results allow us to locate the reasons behind this observation and to assess possibly task-specific effects on the standard syntax-phonology mapping. We formulate the reasons behind *proximity* as follows:

To emphasize the presence of a constituent X (e.g. in rendering brackets in the stimuli of an experiment around the constituent), a speaker may render X as a phonological phrase, even if the syntax-prosody mapping would otherwise assign the words in X to two (or more) phonological phrases.

The scheme in (4) illustrates the conflict between the standard syntax-prosody mapping and the task-specific effect.

(4)   Expected by syntax-prosody mapping: Proximity:

(   )(   )Iph   (   )(   )IPh

(   )(   )(   )PPh   (✂   )(   )PPh

(Lola)(oder Manu)(und Lena)PW   (Lola)(oder Manu)(und Lena)PW

(   )(   )Iph   (   )(   )IPh

(   )(   )(   )PPh   (   )(✂   )PPh

(Lola)(oder Manu)(und Lena)PW   (Lola)(oder Manu)(und Lena)PW

Goal: Each name (DP) is a   Goal: Brackets from experimental

separate PPh.   task identical to PPh.

In sum, it seems to us that many of our speakers have realized the task of conveying a given

bracketing by deleting internal PPh boundaries, as captured by the *proximity* principle.[2]

---

[2] While our prosodic analysis is compatible with recent theories like Selkirk (2011), in which
the intonation phrase still plays an important role, the suggestions of Selkirk (2011) also
freely allow smaller phonological phrases inside of larger phonological phrases. Notice that
this does not seem to affect the analytical choices in the structures we investigated, since
there do not seem to be PPh boundaries around the individual names when two names were
elicited as bracketed together.

4.2. Perception study

4.2.1. Relevance of the three prosodic cues

A major finding is that the bracketing distinction between the Early and the Late conditions is triggered by each of the three cues separately. However, f0, final lengthening, and pauses do not affect the early vs. late boundary judgments in the same way. While the manipulations of final lengthening and pauses at different temporal locations shifted the judgments from early to late interpretation, f0 peak manipulations biased listeners' judgments towards a late boundary interpretation even at early steps of the phonetic continua. Hence, contrary to Gollrad (2013), f0 appears to be a less reliable cue for IP boundary perception. In addition, pauses are used in a categorical fashion for the IP boundary perception in that their manipulation led to abrupt shifts in listeners' judgments concerning the location of the IP boundary.

The perception results can be partly interpreted by taking into account the production data. In the production study, the f0 was the most variable cue across speakers. The brackets could be realized with different f0 contours (e.g., f0 rises, high plateaus, or falling configurations), depending on the speakers' choice. Despite this variability, only the most frequent contour (f0 rises) was selected for the perception manipulation. On the other hand, pauses and final lengthening occurred more systematically and more invariantly at the brackets without being constrained by a particular f0 contour. As a consequence, assessing the role of f0 in boundary perception is less straightforward. Because of the variability in production, listeners might adopt different strategies when computing f0 information than when computing temporal information (for instance, they might be more sensitive to probabilistic or statistical properties of the f0 contours).

The joint interpretation of identification and reaction times for final lengthening and f0 peaks manipulation indicates that the shift between early and late boundary is gradual, especially for f0 and final lengthening. In those cases, the identification function is generally shallow and no clear RT peaks are consistently realized at the category boundaries. This is different from our expectations, by which RT maxima should correspond to abrupt changes in identification. For pause manipulation, the abrupt changes in the response function of the identification task are accompanied by longer RT values in the middle of the continua (inverted u-shaped response).

What reasons might there be for a gradual transition from one end to the other end of the f0 and final lengthening continuum in the identification task, with the absence of clear local maxima or minima in the reaction times? We would like to give two hypotheses.

One possible reason emerges from our analysis of the production experiment in terms of both IP boundaries and PPh boundaries. Let us consider this in terms of the F0 continuum in Figure 7. The continuum over the structure [N1 oder N2 und N3] started with a high N1 and a low N2 and ended with a low N1 and a high N2. This turns out not to be just the addition or deletion of a single prosodic boundary. With a fuller understanding of the structures, there are a number of boundary changes in the course of the continuum. A high N1 and a low N2 may correspond to [(N1)][(oder N2 und N3)], with round brackets for PPh and square brackets for IP. From this end of the continuum, moving to a low N1 and a high N2 may correspond to moving to: [(N1 oder N2)][(und N3)]. This involves the following boundary changes: (a) insertion of a PPh boundary after N2, (b) deletion of a PPh boundary after N1, (c) insertion of an IP boundary after N2, (d) deletion of an IP boundary after N1. These changes need not all be located at the same point of the continuum. If they are at all more or less discrete points in the continuum, they may be distributed across different points of the continuum, with their effects showing considerable overlap. This reasoning alone

might account for the gradual nature of the transition in the identification task and the absence of local maxima or minima in the reaction times across the continuum.

Another possible reason concerns the nature of the prosodic boundaries, and it emerges from the literature on the categoriality vs. gradiency of prosodic boundary perception (see Krivokapić, 2007 for a review). We know that listeners might compute the magnitude of the prosodic boundary relative to other boundaries in the utterance (e.g., Clifton, Carlson and Frazier, 2002; Snedeker and Casserly, 2010). However, it is not clear yet whether relative boundary strength is computed in a categorical or gradient manner. On one hand, if differences between prosodic boundaries (IP, PPh, PW) are discrete and categorical, acoustic changes in their production should be perceived in a categorical fashion, i.e., around a limited set of clusters of phonetic values (one for each prosodic boundary; cf. Schafer, 1997). On the other hand, empirical studies found that the perception of relative boundary strength changes gradually along with continuous acoustic cues (e.g., Wightman et al., 1992; Swerts, 1997; Wagner and Crivellaro, 2010) indicating a gradient prosodic hierarchy. In our perception study, the base stimuli had only one IP-internal prosodic boundary and hence only two parsing decisions might be possible, either [N1][oder N2 und N3] or [N1 oder N2][und N3] where the square brackets stands for IP. Listeners might have computed relative strength to locate the IP boundary, for example, they were judging the strength of the late boundary in relation to the strength of the early boundary. When the cues were manipulated, the parsing decision was less clear. Continuous acoustic variation within the IP category was thus taken into account by listeners. Crucially, the way of dealing with such a variation depended on the specific cue. Listeners gave more categorical responses for pause manipulation (i.e., disregarding continuous variation); however, they exploited f0 and final lengthening cues in a more gradient manner. This difference is in line with our acoustic results, where: (1) pauses are used in a more categorical way to signal the presence/absence of an IP boundary; (2) final

lengthening is used for different levels of boundary; and (3) f0 is the most variable cue, being characterized by strong speaker-specific differences. Moreover, it might account for the finding in the literature that pause overrides the perceptual salience of the other cues (e.g., Peters, 2005) since categorical differences among prosodic cues might have stronger effects than gradient ones.

Future experiments might help in deciding among these two hypotheses. For instance, the first hypothesis could be addressed by manipulating the phonetic space which covers only two prosodic structures (e.g., between the prosodic word and the phonological phrase; and between the phonological and the intonational phrases). If the effects of the phonetic cues are categorical, this would mean that gradual transitions from early boundary perception to late boundary perception in the current experiment might be due to the fact that there are several categorical changes along the continua. The second hypothesis might be addressed through a study in which listeners are exposed to a larger range of prosodic boundaries, to verify whether they perceive them in small or large number of clusters of phonetic values. The two hypotheses can be complementary. Indeed, it has been proposed that, though prosodic hierarchy might involve a few distinct prosodic categories, there might be more structurally different prosodic boundaries which are perceived in a gradient manner (e.g., Krivokapić, 2007).

Finally, in our study separate manipulations of final lengthening, f0, and pauses were necessary to understand the role of single cues to boundary perception. However, this procedure may be an oversimplification of the complex perceptual integration of phonetic cues. Hence, in future experiments, the gradual manipulation of combined cues will be crucial to test the impact of cue weighting, that is, the relative importance of one cue (e.g., final lengthening) as compared to another one (e.g., f0 peaks).

4.2.2. Prosodic disambiguation and on-line prosodic processing

In this paragraph, we would like to address the question whether the separately presented acoustic parameters are in fact processed as cues to prosodic boundaries by the listeners. Very recently, we investigated the processing of lists of names in German by means of event-related brain potentials (ERPs) experiments. In what follows, we will briefly compare results from these experiments with the current study.

We already know that the processing of IP boundaries can be assessed at the brain level by ERPs. Here, the presence of a specific ERP component, the *Closure Positive Shift* (CPS), indexes the processing of intonational phrasing (Steinhauer, Alter and Friederici, 1999). Using natural stimuli that were similar to our unmanipulated stimuli, Holzgrefe, Wellmann, Petrone, Truckenbrodt, Höhle and Wartenburger (2013) found that listeners show a CPS at the position of the bracket, thus giving neural evidence for the presence of an IP boundary (it should be noted though that the CPS was elicited only for the Late condition, for which listeners had prosodic information available to make a real-time interpretation of such breaks as IP boundaries).

In another series of experiments, Holzgrefe, Wellmann, Petrone, Truckenbrodt, Höhle and Wartenburger (2016) found that a CPS is triggered in German by a combination of the two cues final lengthening and f0 height, though not by only final lengthening and not by only f0 height (for compatible results using the head turn preference paradigm with eight-month-old children, see Wellmann, Holzgrefe, Truckenbrodt, Wartenburger, and Höhle, 2012). It is therefore concluded that pause, albeit being a very salient cue, is not necessary for IP perception. The discrepancy with our current results can be due to a difference in the kind of manipulations which were applied. In our current experiments, we simultaneously manipulated pauses, f0 and final lengthening at two different locations in the sentences, that is, at N1 and at N2. Thus, our listeners could rely on information about each of these cues,

which was more distributed over the utterance. On the other hand, Holzgrefe et al. (2016) manipulated duration and f0 only at the position of the (late) IP boundary, that is, only at N2. One could hypothesize that, because of this local manipulation, f0 and final lengthening were not robust enough to alone convey the difference across the two boundary conditions.

Finally, it would be interesting in the future to verify whether pauses alone- which proved to be perceptually relevant in our experiment - may trigger full boundary processing at the brain level. Previous studies have shown that a CPS also occurs in the absence of pauses (Steinhauer et al., 1999; Pannekamp, Toepel, Alter, Hahne, and Friederici, 2005; Holzgrefe et al., 2016), which is in line with our current results testing only single, but temporally distributed boundary cues.

5. Conclusions

This study was aimed at investigating the role of final lengthening, pauses and f0 cues on the production and perception of prosodic boundaries of different strengths.

In line with the literature, results in production show that pauses are employed to signal intonation phrase boundaries, while final lengthening and f0 are employed across different boundary strengths. Differently from expectations from the standard syntax-prosodic mapping, no phonetic differences were found between prosodic word and phonological phrase boundaries (e.g., Nespor and Vogel, 1986; Vigário, 2003; Prince and Smolensky, 2004; Selkirk, 2011; Truckenbrodt 2002, 2004, 2007a-b). This is interpreted in terms of task-specific constraints. Speakers were able to map the correspondence between IPs and syntactic constituents, and the correspondence between PPhs and syntactic constituents (Selkirk 2011). It seems that the task of rendering a particular structure leads not only to the addition of major prosodic constituents, but also to the suppression of prosodic constituents that would otherwise be assigned.

In perception, our results show that, when presented as single (but temporally distributed) cues, final lengthening, f0, and pauses each affected listeners' judgments. Listeners gave categorical responses for pauses while a more gradual transition was found when using f0 and final lengthening cues. This difference accounts for some findings in the literature that the pause is a strong cue for boundary perception (e.g., Peters, 2005), though it is not necessary to perceive a prosodic boundary (see § *4.3.2*).

Taken together, our findings suggest that (1) the syntax-prosody mapping is influenced by task-specific constraints; (2) distributed phonetic properties are both produced and perceived for disambiguating ambiguous syntactic structure; and (3) pauses, final lengthening, and f0 have a different effect on boundary production and perception. Future work will be crucial to investigate how these three acoustic cues are weighted in the production and perception of prosodic boundaries across different levels of the prosodic hierarchy.

APPENDIX

Table A.1. Statistical results for the fixed effects of the final linear mixed model of the production experiment on final lengthening. Data are reported for the duration of vowels on N1 (a) and N2 (b). Significance codes: p < 0.001 (***), p < 0.01 (**), p < 0.05 (*).

(a)

Fixed effects	β	SE	t-Value
Intercept (Early)	0.20	0.0075	26.59**
BOUNDARY(Late)	-0.09	0.0076	-12.57**
BOUNDARY(Word)	0.008	0.0076	-0.35

(b)

Fixed effects	β	SE	t-Value
Intercept (Early)	0.09	0.0054	17.72**
BOUNDARY(Late)	0.08	0.0052	15.89**
BOUNDARY(Word)	0.008	0.0052	1.68

Table A.2. Statistical results for the fixed effects of the final linear mixed model of the production experiment on f0. Data are reported for f0 height (f0 rises and plateaus) on N1 (a) and N2 (b). Significance codes: p < 0.001 (***), p < 0.01 (**), p < 0.05 (*).

(a)

Fixed effects	β	SE	t-Value
Intercept (Early)	297.04	15.88	18.707**

Fixed effects	β	SE	t-Value
BOUNDARY(Late)	-45.43	22.14	-2.05*
BOUNDARY(Word)	-1.47	22.14	-0.06
SEX	95.82	24.39	-3.92**
BOUNDARY(Late):SEX	9.77	35.64	0.27
BOUNDARY(Word):SEX	4.22	35.60	0.11

(b)

Fixed effects	β	SE	t-Value
Intercept (Early)	237.66	10.30	23.06**
BOUNDARY(Late)	40.88	14.49	2.82**
BOUNDARY(Word)	-2.53	14.49	0.17
SEX	-34.17	22.41	-1.52
BOUNDARY(Late):SEX	-34.17	22.41	-1.52
BOUNDARY(Word):SEX	-3.40	31.52	-0.10

Table A.3. Statistical results for the fixed effects of the final linear mixed model of the production experiment on pauses. Data are reported for the duration of pauses after N1 (a) and N2 (b). Significance codes: $p < 0.001$ (***), $p < 0.01$ (**), $p < 0.05$ (*).

(a)

Fixed effects	β	SE	t-Value
Intercept (Early)	0.16	0.02	7.53**

Fixed effects	β	SE	t-Value
BOUNDARY(Late)	-0.12	0.02	-6.06**
BOUNDARY(Word)	0.01	0.005	3.21**

(b)

Fixed effects	β	SE	t-Value
Intercept (Early)	0.02	0.03	0.96
BOUNDARY(Late)	0.08	0.02	2.84**
BOUNDARY(Word)	0.01	0.03	0.57

Table A.4. Mean score (%) and standard error (SE) of the late boundary judgments across the steps 1-10 of the final lenghtening manipulation. Dara are shown separately for Early (left) and Late (right) base stimuli.

Early base			Late base		
Steps	Mean	SE	Steps	Mean	SE
1	8.9	3.8	1	39.2	6.5
2	25.0	5.8	2	35.7	6.4
3	25.0	5.8	3	41.0	6.6
4	30.3	6.1	4	48.2	6.7
5	39.2	6.5	5	53.5	6.7
6	44.6	6.7	6	69.6	6.1
7	60.7	6.5	7	82.1	5.1
8	73.2	5.9	8	83.9	4.9
9	66.0	6.3	9	73.2	5.9

| 10 | 66.6 | 6.4 | 10 | 83.9 | 4.9 |

Table A.5. Mean score (%) and standard error (SE) of the late boundary judgments across the steps 1-10 of the f0 peak manipulation. Dara are shown separately for Early (left) and Late (right) base stimuli

Early base			Late base		
Steps	Mean	SE	Steps	Mean	SE
1	31.7	5.9	1	45.1	6.3
2	44.4	6.3	2	37.0	6.1
3	39.6	6.2	3	60.3	6.2
4	53.9	6.3	4	70.9	5.8
5	53.9	6.3	5	74.6	5.5
6	61.9	6.1	6	77.7	5.2
7	69.8	5.8	7	87.3	4.2
8	82.5	4.8	8	87.3	4.2
9	70.9	5.8	9	79.3	5.1
10	84.1	4.6	10	84.1	4.6

Table A.6. Mean score (%) and standard error (SE) of the late boundary judgments across the steps 1-10 of the pause manipulation. Dara are shown separately for Early (left) and Late (right) base stimuli

Early base			Late base		
Steps	Mean	SE	Steps	Mean	SE
1	12.5	4.4	1	14.2	4.7
2	12.5	4.4	2	7.14	3.4
3	20.7	5.6	3	18.1	5.2
4	36.3	6.5	4	16.6	5.1
5	26.4	6.1	5	28.5	6.0
6	43.6	6.7	6	28.3	6.2
7	43.6	6.7	7	57.1	6.6
8	54.7	6.9	8	67.8	6.2
9	69.6	6.1	9	69.6	6.1
10	76.7	5.6	10	87.5	4.4

Table A.7. Statistical results for the fixed effects of : (a) the final mixed logit model (for late boundary judgments) and (b) the final linear mixed model (for reaction times) of the perception experiment on final lengthening. Significance codes: $p < 0.001$ (***), $p < 0.01$ (**), $p < 0.05$ (*).

(a)

Fixed effects	β	SE	z-Value
Intercept (Early)	-0.41	0.12	-3.19**
STEP	0.29	0.08	3.52***
BASE	0.76	0.12	6.30***

	-0.06	0.05	-1. 06
STEP:BASE			

(b)

Fixed effects	β	*SE*	*t*-Value
Intercept (Early)	7.56	0.14	52.00**
STEP (linear)	0.77	0.35	2.15*
STEP (quadratic)	-0.14	0.35	-0.39
BASE	0.01	0.01	1.08
STEP (linear):BASE	-1.88	0.50	-3.73**
STEP (quadratic):BASE	-1.11	0.50	-0.23

Table A.8. Statistical results for the fixed effects of : (a) the final mixed logit model (for late boundary judgments) and (b) the final linear mixed model (for reaction times) of the percepti on experiment on f0 peaks. Significance codes: $p < 0.001$ (***), $p < 0.01$ (**), $p < 0.05$ (*).

(a)

Fixed effects	β	*SE*	*z*-Value
Intercept (Early)	0.51	0.24	2.05*
STEP	0.26	0.08	3.06**
BASE	0.59	0.28	2.09*
STEP:BASE	0.01	0.04	0.32

(b)

Fixed effects	β	SE	z-Value
Intercept (Early)	7.63	0.03	239.43**
STEP (linear)	-0.43	0.33	-1.31
STEP (quadratic)	-1.04	0.27	-3.73**
BASE	0.04	0.01	2.59**
STEP (linear):BASE	0.18	0.39	0.47
STEP (quadratic):BASE	1.16	0.39	2.93**

Table A.9. Statistical results for the fixed effects of : (a) the final mixed logit model (for late boundary judgments) and (b) the final linear mixed model (for reaction times) of the percepti on experiment on pauses. Significance codes: $p < 0.001$ (***), $p < 0.01$ (**), $p < 0.05$ (*).

(a)

Fixed effects	β	SE	z-Value
Intercept (Early)	-0.86	0.19	-4.37***
STEP	0.46	0.09	4.81***
BASE	-0.28	0.20	-1.37
STEP:BASE	0.15	0.04	3.38***

(b)

Fixed effects	β	*SE*	*z*-Value
Intercept (Early)	7.71	0.03	202.97**
STEP (linear)	-0.27	0.33	-0.81
STEP (quadratic)	-1.04	0.25	-4.06**
BASE	-0.004	0.01	-0.34
STEP (linear):BASE	-0.50	0.36	-1.39
STEP (quadratic):BASE	-1.13	0.36	-0.38

References

Baayen, H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R.* Cambridge: Cambridge University Press.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Kliegl, R., Vasishth, S., and Baayen, R. H. (submitted). Parsimonious mixed models. Retrieved from: arXiv:1506.04967.

Beckman, M. E., and Edwards, J. (1990). Lengthening and shortening and the nature of prosodic constituency. In J. Kingston, and M. E. Beckman (Eds.), *Papers in Laboratory Phonology I* (pp. 152–178). Cambridge: Cambridge University Press.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341–345.

Browman, C.P., and Goldstein, L. (1992). Articulatory Phonology: An Overview. *Phonetica*, *49*, 155–180.

Carpentier, F., and Moulines, E. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9***,** 453–457.

Cho, T. (2005). Prosodic influences on consonant production in Dutch: effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, *33*, 121–157.

Chen, A. (2003). Reaction time as an indicator to discrete intonational contrasts in English. *Proceedings Eurospeech*, 97–100. Retrieved from: http://www.isca-speech.org/archive/archive_papers/eurospeech_2003/e03_0097.pdf

Clark, H. H., and Carlson, T. B. (1982). Hearers and speech acts. *Language*, *58*, 332–373.

Clark, H. H., and Murphy, G. L. (1982). Audience design in meaning and reference. In J.-F. Le Ny and W. Kintsch (Eds.), *Language and comprehension*, 287–299. Amsterdam, The Netherlands: North-Holland.

Clifton C., Jr., Carlson K., and Frazier L. (2002). Informative prosodic boundaries. *Language and Speech*, *45*, 87–114.

Cooper, W. E., and Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Downing, B. (1970). *Syntactic Structure and Phonological Phrasing in English*. Ph.D. Thesis, University of Texas.

Féry, C. (2016). *Intonation and Prosodic Structure.* Cambridge: Cambridge University Press.

Féry, C., and Kentner, G. (2010). The prosody of embedded coordinations in German and Hindi. *Proceedings of Speech Prosody*, 5th International Conference 2010, (Chicago, IL: USA).

Féry, C., and Kügler, F. (2008) Pitch accent scaling on given, new and focused constituents in German, *Journal of Phonetics*, *36*, 680–703.

Fougeron, C., and Keating, P.A. (1997). Articulatory Strengthening at Edges of Prosodic Domains. *Journal of the Acoustical Society of America*, *101*(6)**,** 3728–3740.

Forster, K., and Forster, J. (2003). A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments and Computers*, *35*(1), 116–126. Computer program retrieved at: http://www.u.arizona.edu/~kforster/dmdx/dmdx.htm

Gollrad, A. (2013). *Processing cue weighting in sentence comprehension: Processing German case ambiguous structures*. Ph.D. Thesis, Universität Potsdam.

Wagner M., and Crivellaro S. (2010). Relative prosodic boundary strength and prior bias in disambiguation. In *Proceedings of Speech Prosody*, 5th International Conference 2010, (Chicago, IL: USA).

Grabe, E. (1998). *Comparative intonational phonology: English and German.* Ph.D. Thesis, Universiteit Nijmegen.

Grice, M., Baumann, S., and Jagdfeld, N. (2009). Tonal association and derived nuclear accents–The case of downstepping contours in German. *Lingua*, *119*(6), 881–905.

Grosjean, F., Grosjean, L., and Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, *11*, 51–58.

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics*, *19*, 377–417.

Gussenhoven, C. (1992). Sentence accents and argument structure. In Ignacio M. Roca (Ed.), *Themantic Structure: Its Role in Grammar* (pp. 79–106). Berlin/New York: Foris.

Gussenhoven, C. (2004). *The phonology of tone and intonation.* Cambridge: Cambridge University Press.

Hayes, B. (1995). *Metrical stress theory: principles and case studies.* Chicago: The University of Chicago Press.

Hayes, B., and Lahiri, A. (1991). Bengali intonational phonology. *Natural Language and Linguistic Theory*, *9***,** 47–96.

Holzgrefe, J., Wellmann, C., Petrone, C., Truckenbrodt, H., Höhle, B., and Wartenburger, I. (2013). Brain response to prosodic boundary cues depends on boundary position. *Frontiers in Psychology*, *4*, 421. doi: 10.3389/fpsyg.2013.00421

Holzgrefe-Lang, J., Wellmann, C., Petrone, C., Räling, R., Truckenbrodt, H., Höhle, B., and Wartenburger, I. (2016). How pitch change and final lengthening cue boundary perception in German: Converging evidence from ERPs and prosodic judgments. *Language, Cognition and Neuroscience*, *31*, 904–920.

Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, *15*, 189–226.

Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics,* 149–181.

Keating, P., Cho, T., Fougeron, C., and Hsu, C. (2003). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, and R. Temple (Eds.), *Papers in Laboratory Phonology VI: Phonetic Interpretation* (pp. 143–161). Cambridge: Cambridge University Press.

Kent, R.D., and Netsell, R. (1971). Effects of stress contrasts on certain articulatory parameters. *Phonetica*, *24*, 23–44.

Kentner, G., and Féry, C. (2013). A new approach to prosodic grouping. *The Linguistic Review*, *30*(2), 277–311.

Knight, R.-A., and Nolan, F. (2006). The effect of pitch span on intonational plateaus. *Journal of the International Phonetic Association*, 36(1), pp. 21–38.

Kohler, K. J. (1983). Prosodic boundary signals in German.*Phonetica*, *40*, 89–134.

Krivokapić, J. (2007). *The planning, production, and perception of prosodic structure*. Ph.D. Thesis, University of Southern California.

Ladd, D.R. (1988).Declination 'reset' and the hierarchical organization of utterances. *Journal of Acoustical Society of America*, 530–544.

Ladd, D. R. (1998). *Intonation phonology*. (2nd edition). Cambridge: Cambridge University Press.

Massaro, D.W. (1987). *Speech perception by ear and eye: a paradigm for psychological inquiry*. Hillsdale, NJ: London.

Nespor, M., and Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Niebuhr, O. (2007). The signalling of German rising-falling intonation categories - The interplay of synchronization, shape, and height. *Phonetica*, 64, 174–193.

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., and Friederici, A. D. (2005). Prosody-driven sentence processing: An event-related brain potential study. *Journal of Cognitive Neuroscience*, *17*, 407–421.

Peters, B. (2005). Weiterführende Untersuchungen zu prosodischen Grenzen in deutscher Spontansprache. In K. J. Kohler, F. Kleber, and B. Peters (Eds.), *AIPUK. Prosodic structures in German spontaneous speech (AIPUK 35a)* (pp. 203–345). Kiel: IPDS.

Peters, B., Kohler, K.J., and Wesener, T. (2005). Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache. In K.J. Kohler, F. Kleber, and B. Peters (Eds.). *Prosodic structures in German spontaneous speech (Arbeitsberichte des IPDS d. Uni Kiel 35a)* (pp. 143–184). Kiel: IPSD, Kiel University.

Pheby, J. (1981). Phonologie: Intonation. In K.E. Heidolph, W. Flämig, and W. Motsch (Eds.), *Grundzüge einer deutschen Grammatik* (pp. 839–897). Berlin: Akademie-Verlag.

Pierrehumbert, J.B. (1980). *The Phonology and Phonetics of English Intonation.* Ph.D. Thesis, Massachusetts Institute of Technology.

Pierrehumbert, J., and D. Talkin (1991). Lenition of /h/ and glottal stop. In G. Docherty, and D. R. Ladd (Eds.), *Papers in Laboratory Phonology II* (pp. 90–117). Cambridge, UK: Cambridge University Press.

Prince, A., and Smolensky, P. (2004). *Optimality Theory: constraint interaction in generative grammar.* Oxford: Blackwell.

Quené, H., and Van Den Bergh, H. (2004). On multi-level modelling of data from repeated measures designs: A tutorial. *Speech Communication*, *43*, 103–121.

Redi, L., and Stefanie Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, *29*, 407–429.

R-Development, C.-T. (2012). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.

Scott, D.R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America*, *71*(4), 996–1007.

Selkirk, E. (1980). Prosodic Domains in Phonology: Sanskrit Revisited. In M. Aronoff, and M.-L. Kean (Eds.), *Juncture: A Collection of Original Papers* (pp. 107–129). Saratoga, CA: Anma Libri.

Selkirk, E. (1984). *Phonology and Syntax: the Relation between Sound and Structure*. Cambridge, Mass: MIT Press.

Selkirk, E. (1986). On Derived Domains in Sentence Phonology. *Phonology Yearbook*, 3, 371–405.

Selkirk, E. (1996). The prosodic structure of function words. In J. L. Morgan, and K. Demuth (Eds.). *Signal to Syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–213). Mahwah, NJ: Lawrence Erlbaum.

Selkirk, E. (2011). The syntax-phonology interface. In J. Goldsmith, J. Riggle, and A. Yu (Eds.), *The handbook of phonological theory, 2nd edition* (pp. 435–484). Oxford: Blackwell.

Selkirk, E., and Shen, T. (1990). Prosodic domains in Shanghai Chinese. In S. Inkelas, and D. Zec. *The phonology-syntax connection* (pp. 313–337). Chicago: The University of Chicago Press.

Speer, S. R., Warren, P., and Schafer, A. J. (2011).Situationally independent prosodic phrasing. *Laboratory Phonology*, *2*(1), 35–98.

Snedeker J., and Casserly E. (2010). Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes*, *25*, 1234–1264.

Steinhauer, K., Alter, K., and Friederici, A.D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*(2), 191–196.

Strangert, E. (1997). Relating prosody to syntax: Boundary signaling in Swedish. In G. Kokkinakis, N. Fakotakis, and E. Dermatas (Eds.), *Proceedings of the 5th European Conference on Speech Communication and Technology* (pp. 239–242).

Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America*, *101*, 514–521.

Truckenbrodt, H. (1995). *Phonological phrases: their relation to syntax, focus, and prominence.* Ph.D. Thesis, Massachusetts Institute of Technology.

Truckenbrodt, H. (2002). Upstep and embedded register levels. *Phonology*, *19*, 77–120.

Truckenbrodt, H. (2004). Final lowering in non-final position. *Journal of Phonetics*, *32*, 313–348.

Truckenbrodt, H. (2007a). The syntax-phonology interface. In P. De Lacy (Ed.), *The Cambridge handbook of phonology* (pp. 435–456). Cambridge: Cambridge University Press.

Truckenbrodt, H. (2007b). Upstep on edge tones and on nuclear accents. In C. Gussenhoven, and T. Riad (Eds.), *Tones and tunes. Volume 2: Experimental studies in word and sentence prosody* (pp. 349–386). Berlin: Mouton.

Truckenbrodt, H. (2015). Intonation phrases and speech acts. In Marlies Kluck, Dennis Ott and Mark de Vrieds (Eds.), *Parenthesis and ellipsis. Cross-linguistic and theoretical perspectives* (pp. 301–349). Berlin: Walter de Gruyter.

Truckenbrodt, H.(2016). Intonation in der Lautsprache: Prosodische Struktur. In U. Domahs, and B. Primus (Eds.), *Handbuch Sprachwissen: Laut – Gebärde – Buchstabe* (pp. 106–124). Berlin: De Gruyter.

Truckenbrodt, H., and Féry, C. (2015). Hierarchical organization and tonal scaling. *Phonology*, *32*, 19–47.

Turk, A., and Shattuck-Hufnagel, S. (2007). Multiple targets of phrase final lengthening in American English words. *Journal of Phonetics*, *35*, 445–473.

van den Berg, R., Gussenhoven, C., and Rietveld, T. (1992). Downstep in Dutch: Implications for a model. In G. J. Docherty, and D. R. Ladd (Eds.), *Papers in Laboratory Phonology II* (pp. 335–367). Cambridge: Cambridge University Press.

Vigário, M. (2003). *The prosodic word in European Portuguese.* Berlin: Mouton de Gruyter.

Wagner, M. (2005). *Prosody and recursion.* Ph.D. Thesis, Massachusetts Institute of Technology.

Wagner M., and Crivellaro S. (2010). Relative prosodic boundary strength and prior bias in disambiguation. In *Proceedings of Speech Prosody*, 5th International Conference

2010, (Chicago, IL: USA). Retrieved from:

http://speechprosody2010.illinois.edu/papers/100238.pdf

Watson, D., and Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, *19*(6), 713–755.

Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., and Höhle, B. (2012). How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology*, *3*, 580. doi: 10.3389/fpsyg.2012.00580

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and P. J. Price (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, *91*, 1707–1717.

Żygis, M., Brunner, J., and Moisik, S. (2012) The affinity of low vowels and glottal stops:an articulatory model and a perceptual investigation. Poster presented at the *13th Conference on Laboratory Phonology*. Universität Stuttgart, 27-29.07.2012.