Vowel reduction in conversational speech in French: the role of lexical factors

Christine Meunier, Robert Espesser

▶ **To cite this version:**

Christine Meunier, Robert Espesser. Vowel reduction in conversational speech in French: the role of lexical factors. Journal of Phonetics, 2011, 39 (3), pp.271-278. 10.1016/j.wocn.2010.11.008 . hal-01205592

HAL Id: hal-01205592

https://hal.science/hal-01205592

Submitted on 25 Apr 2016

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

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

Vowel reduction in conversational speech in French: the role of lexical factors

Christine Meunier[a]* & Robert Espesser[a]

*Corresponding author: christine.meunier@lpl-aix.fr, Tel: (+33)442953639, Fax: (+33)4429537 88

[a]Laboratoire Parole et Langage, Université de Provence, CNRS, 5 avenue Pasteur - 13100 Aix-en-Provence, France

Abstract

In this study we investigate vowel reduction and the role of some lexical factors in the production of vowels extracted from a corpus of French conversations. Vowel durations and spectral quality are examined with respect to 1. their interaction in the corpus, 2. the position of vowels in words, and 3. word frequency and word category. The analyses are conducted on vowels produced by 16 speakers. Our study provides strong evidence that vowel reduction (decrease in durations and more centralized spectral values) affects most of the vowels in conversational speech. The results show that vowels in final syllables of words were less often reduced while the preceding ones show reduced durations and centralized formant values. Moreover vowels are more reduced in monosyllabic function words than in monosyllabic content words. Nevertheless, we did not find a clear effect of word frequency on vowel durations. Finally, our study shows that vowel reduction depends on several factors related to lexical properties (word category) and to prosodic properties (stress and final lengthening).

Keywords

Vowels, French, word frequency, word category, durations, spectral reduction, spontaneous speech.

1. Introduction

Speech reduction in conversational speech is currently a challenging topic, since its manifestation highlights the complex interaction and relationship between linguistic domains. It involves questions about the processes that lead speakers to produce speech in a non-canonical way. The main question is when (in the speech flow) and how (as which phonetic phenomena) speech is reduced. The answer to these questions would give us valuable information on the relationship between the phonetic, lexical, syntactic and pragmatic domains.

This paper is focused on vowel reduction, which we consider to be a phenomenon of undershoot related to unattained targets, generally involving shortened vowel duration leading to a reduced vowel space. This phenomenon affects vowel quality so that the values observed under certain conditions are often far removed from prototypical ones. A large study on French vowel reduction has been conducted by Gendrot & Adda-Decker (2005, 2007 and 2008). These researchers investigated the reduction of French vowel duration and spectral quality on a large French corpus (ESTER) which is composed of recordings of French radio broadcast news, in which speech is largely prepared and planned. They showed that short vowels have more central spectral values, confirming that spectral reductions are correlated with duration reductions (as in other languages). Moreover, they highlighted the effect of some lexical factors on vowel durations. Our study is based on a corpus of conversational speech, since conversational speech is known to be less controlled than broadcast news (Torreira & Ernestus, 2010) and provides specific phenomena (disfluences, unfinished sentences, hesitations, back-channels, overlapping speech, etc.). Consequently, we ask whether vowel reduction would be greater or qualitatively different in very casual speech. Our study provides specific data on vowel production in casual speech in French.

The occurrence of reduction in vowel duration and changes in the spectral quality of the vowel is influenced by a number of factors. Lindblom (1963) showed that vowel reduction occurs in languages with lexical stress, such as English and Swedish, and particularly affects unstressed vowels. It is generally agreed that French does not have lexical stress (Hirst et al. 2001), and that stress in French is a property of the phrase rather than the word. Nevertheless, several studies have shown that the end of the Accentual Phrase (a domain greater than the lexical unit) coincides with a word-final syllable (Dell, 1984, Fougeron, 2002). Since the end of an Accentual Phrase (AP) is the final syllable of a word, internal syllables should be more reduced than the final ones. Indeed, Adda-Decker et al. (2008) showed that vowel duration depended on syllable position within the word: long vowels were more frequent in final syllables whereas short vowels tended to appear in non-final syllables. Since speech in our corpus is less controlled and prepared, we asked whether vowel durations and quality would be affected in the same way.

Speech communication is a dynamic process in which speakers adapt their production so that their message can be understood by listeners (Lindblom 1990). Reduced speech therefore often appears when the linguistic message is highly predictable (Jurafsky et al., 2001, Hume, 2004). On the one hand, the predictability of words may be determined in part by their frequency of occurrence in speech. Several studies have shown that high frequency words are more often reduced than low frequency ones (Bybee, 2002). Pluymakers et al. (2005), for example, observed reduced durations in high frequency morphologically complex Dutch words. But word frequency also affects phonetic details such as voicing assimilation (Ernestus et al., 2006). On the other hand, predictability may also depend on word category. In a corpus of spontaneous conversations between American English speakers, Johnson (2004) observed that function words were deleted (or produced with forms that deviated from their citation form) more often than were content words. In a study on controlled speech in Dutch, van Bergem (1993) showed that both word stress and word category have an influence on the spectral quality of vowels: unstressed words and function words were more reduced. From this overview, it seems that the production of reduced forms, and particularly acoustic reduction, is strongly related to predictability and information in the speech flow. The more predictable a word is, the less phonetic information is necessary (Hume, 2004). With respect to vowel reduction, Adda-Decker & al. (2008) showed that short vowels are more often present in function words than in content words. In our analysis we quantify this difference and we analyze spectral quality according to both word categories (function word and content word).

Speech reduction and variation has received particular interest since studies on spontaneous speech have become more common. It has been shown that the amount of reduction in spontaneous speech is greater than expected (Johnson, 2004). Speaking styles differ in what reduction phenomena they demonstrate, depending on the degree of control in the speech. In our study, we aim to evaluate the amount of vowel reduction in very casual speech. We start with a general overview of vowel reduction in our corpus, and then present our analyses of the lexical factors.

2. Speech material and analyses

2.1. The Corpus of Interactional Data (CID)

The CID corpus (Bertrand et al. 2008) consists of eight hours of audio/video recordings of French conversational speech. Each hour is a recording of a conversation between two participants, speakers from the southeast region of France or long-term residents of the region. Five conversations are between two women and three are between two men. The conversations took place in a studio, and the participants wore head-mounted microphones, with each voice recorded on a separate track. The participants were given two topics to discuss: conflicts in their professional environment or funny situations that they had found themselves in. These were only suggestions however, and the participants were free to talk about any topic. As the speakers knew each other quite well, the resulting recordings are highly relaxed conversations.

The corpus was transcribed orthographically by two advanced phonetics students. This transcription was augmented with certain phonetic details (for example, particular pronunciations, such as je sais 'I know', canonically pronounced /ʒəsɛ/, but often pronounced [ʃɛ], were included in the transcription, as well as regional pronunciations). This initial transcription was then processed by a grapheme-to-phoneme converter (Di Cristo & Di Cristo 2001) and an aligner (Fohr et al. 2004) to get the time point of the segments with the speech signal.

The CID corpus is partially or fully annotated on several linguistic levels (segmental, prosodic, syntactic, discourse, gestural, etc.) allowing the analysis of the phonetic segments in interaction with the other linguistic levels. The results presented in this paper are based on the automatic alignment for all 16 speakers.

2.2. French vowels in the CID

The French vowel system includes 11 oral vowels and three or four nasal vowels, depending on the dialect. We are interested in global characteristics of vowel reduction, and our hypotheses do not at all depend on the distinction between oral and nasal vowels. The spectral analysis of nasal vowels is often not reliable, and we therefore chose to analyze only the oral vowels. French does not have a phonological distinction between long and short vowels. Oral vowels represent 39% of all the phonemes in the CID corpus (106,961 oral vowels out of 272,166 phonemes). Although 11 oral vowels are usually distinguished in French, the aligner recognizes only seven macro-classes of oral vowels, since it does not distinguish between mid-closed and mid-open vowels (see Table 1). We do not consider this to be problematic for our analysis, since these distinctions are not strictly phonological. The distinctions between mid-closed and mid-open vowels (/e/-/ɛ/, /ø/-/œ/, /o/-/ɔ/) could be considered as allophonic variation, since these vowels do not typically appear in the same syllable structures. The open vowels appear in closed syllables (*nerf* 'nerve' /nɛr/), while the closed vowels appear in open syllables (*nez* /ne/ 'nose'). For some cases, minimal pairs (*épée* /epe/ 'sword', *épais* /epɛ/ 'thick') can be found, but not in all regional accents (Durand & Lyche, 2004). The consequences for the analyses are that /e/, /o/ and /ø/ would show intermediate spectral values in our corpus.

Table 1 French oral vowels in IPA and vowels grouped by the aligner in the CID

IPA	a	e ɛ	o ɔ	ø œ ə	i	y	u
Aligner	a	e	o	ø	i	y	u
Number of vowels in the CID	25057 (23%)	34915 (33%)	7178 (7%)	15703 (15%)	13337 (12%)	6010 (6%)	4761 (4%)

The vowel /e/ is the most frequent vowel in our corpus followed by /a/, /ø/ and /i/. Together /e/, /a/ and /ø/ represent 71% of all the oral vowels. The least frequent oral vowels are /o/, /y/, and /u/. This distribution is quite similar to that described in literature on French (New et al., 2001), although in other studies, /a/ is typically the most frequent vowel. These differences between the vowel frequencies reported in the literature, and what is observed in the CID corpus can be explained by the merging of the mid-vowels in the automatic alignment. Most of the vowels are produced in a very few words (frequent monosyllabic function words), with 50% of the vowels realized in fewer than 70 different words. For example, 11% of the tokens of /u/ in the corpus appear in the function word *tout* 'every, all'.

2.3. Analyses

Vowel durations were computed from the boundaries placed by the aligner. Due to the parameterization of the aligner, the minimal duration was 24 ms and changed by a step of 8 ms. Formant frequencies were estimated by a linear prediction method (autocorrelation), with a Viterbi algorithm to select the best candidates by imposing frequency continuity constraints (ESPS package, Entropic, 1997). A total of 32,266 vowels shorter than 30 ms and 2,075 vowels longer than 300 ms were dropped from the analyses. These very short vowels often correspond to alignment issues caused by transcriptions in which phonemes not physically present in the signal were transcribed. Extremely long vowels often correspond to long hesitations in the production of certain function words (*mais* ‘but’, *alors* ‘so’, etc.). The exclusion of these very long vowels did not affect the analysis, though, since they were not within the scope of the study. As a result, in this first section, the analyses will be conducted on the 72,620 remaining oral vowels.

All the analyses were carried on using linear mixed effects models (LMEs) which provide a powerful tool for the analysis of grouped data. A mixed model incorporates both fixed effects, which are parameters associated with certain repeatable levels of experimental factors, and random effects, which are associated with individual experimental units sampled at random from a population (Pinheiro & Bates, 2000; Quené & Van den Bergh, 2004). Hence LMEs take into account the correlation of the observations within the same experimental unit. Moreover LMEs can handle unbalanced data. The question of degrees of freedom for these models continues to be a subject of some controversy in the literature. Hence a Monte Carlo Markov Chain technique (Baayen, 2008; Bates et al., 2008) is used to get p-values (MCMC p-values). The data analysis was carried on using the computing language R (R Development core Team, 2008)

3. Results

3.1. Durational and spectral values

The vowel durations extracted from our spontaneous speech corpus are extremely short compared with the durations of vowels produced in controlled speech (e.g., in the reading of nonsense CVC and VCV sequences by French speakers, Bartkova & Sorin, 1987, Table 2). The mean vowel duration was 73 ms. The longest vowel was /u/, followed by /ø/ and /i/. The shortest vowel was /a/ (Table 2).

Table 2: Vowel mean durations (ms) in the CID corpus in comparison with vowel mean durations in controlled speech (Bartkova & Sorin 1987)

<table>
<tr><th></th><th>a</th><th>e</th><th>ɛ</th><th>o</th><th>ɔ</th><th>ø</th><th>œ</th><th>ə</th><th>i</th><th>y</th><th>u</th></tr>
<tr><td>CID</td><td>65</td><td colspan="2">71</td><td colspan="2">73</td><td colspan="3">84</td><td>75</td><td>72</td><td>86</td></tr>
<tr><td>Bartkova & Sorin (1987)</td><td>177</td><td>180</td><td>175</td><td>186</td><td>170</td><td>186</td><td>185</td><td>130</td><td>170</td><td>167</td><td>170</td></tr>
</table>

The distribution of vowel durations was highly unbalanced: 50% of vowel durations were less than 60 ms. A high degree of reduction in vowel durations was thus observed in this corpus, in line with the results observed in another study on French (Adda-Decker et al., 2008).

We next investigated the relation between reduction in duration and spectral reduction. The table 3 shows the spectral values of the oral vowels of the CID. Male and female formant values are presented separately, along with a reference (Tubach, 1989). We present the mode (the higher mode, if there were two modes) of each formant value distribution, since the large amount of data allows the use of this robust measure. The number of measures across the different cases ranged from 196 (long /u/ male speakers) to 6655 (short /e/ female speakers). The vowels were split into three different duration-dependent data sets corresponding to short, intermediate and long vowels in the CID corpus: Set 1=30 ms to 50 ms, Set 2=51 ms to 120 ms, Set 3=121 ms to 300 ms.

Table 3: Female and male spectral values (F1, F2 and F3 in Hz) according to the three duration sets in the CID. Values in controlled speech (Tubach, 1989) are given for a comparison. The values in parentheses indicate that the second higher mode is used (and was selected after visual inspection of the distribution). The number of measurements in each set (female and male values) is also given.

		Number of measurements in the CID						
		a	e/ε	o/ɔ	ø/œ	i	y	u
female	30-50ms	4741	6655	1343	2423	2238	1028	799
	51-120ms	5395	6540	1727	1820	2964	1085	1201
	121-300ms	770	1782	382	1140	756	216	400
Male	30-50ms	3125	4088	860	1844	1474	691	365
	51-120ms	2615	3162	1160	1346	1699	1044	661
	121-300ms	406	926	231	816	433	196	274

		Fomant values – Female speakers									
	Vowel durations	a	e	ε	o	ɔ	ø	œ	i	y	u
F1	30-50ms	414	355		363		354		269	270	292
	51-120ms	521	392		380		383		294	294	293
	121-300ms	752	393		391		443		285	298	281
	Tubach (1989)	788	417	660	461	634	469	647	306	305	311
F2	30-50ms	1697	1912		1561		1737		2205	1915	(1015)
	51-120ms	1699	2037		1304		1678		2269	1939	968
	121-300ms	1639	2186		1015		1697		2484	1959	862
	Tubach (1989)	1503	2351	2080	855	1180	1605	1690	2456	2046	804
F3	30-50ms	2766	2818		2751		2736		2970	2702	2793
	51-120ms	2788	2815		2775		2718		3054	2604	2768
	121-300ms	2816	2831		2801		2651		3379	2563	2833
	Tubach (1989)	2737	3128	2954	2756	2690	2581	2753	3389	2535	2485

		Fomant values – Male speakers									
	Vowel durations	a	e	ɛ	o	ɔ	ø	œ	i	y	u
F1	30-50ms	418	304		307		300		228	246	255
	51-120ms	508	347		345		352		250	255	266
	121-300ms	618	357		377		387		251	269	259
	Tubach (1989)	684	365	530	383	531	381	517	308	300	315
F2	30-50ms	1526	1699		1263		1435		1851	1700	(1310)
	51-120ms	1464	1718		1033		1409		1955	1714	(900)
	121-300ms	1403	1774		919		1359		1978	1785	848
	Tubach (1989)	1256	1921	1718	793	998	1417	1391	2064	1750	764
F3	30-50ms	2446	2443		(2281)		(2378)		2610	(2355)	(2290)
	51-120ms	2457	2484		(2378)		2350		2742	(2289)	(2235)
	121-300ms	2442	2490		(2500)		2312		2848	(2233)	(2389)
	Tubach (1989)	2503	2644	2558	2283	2299	2235	2379	2976	2120	2027

Spectral reduction was observed in relation to duration variation. In Set 1 (short vowels) the vowel space is clearly less peripheral than in Set 3 (long vowels), with Set 2 being intermediate. The extent of the reduction depends on vowel identity and spectral parameters (F1, F2 or F3). Concerning F1, the most sensitive spectral reduction was the decrease of /a/ values, the most open vowel (Figure 1). F1 values of mid-vowels also decreased to a lesser extent. The F1 of closed vowels was not sensitive to reduction in duration. We thus observed a decrease of F1 values related to a duration decrease for mid open vowels, particularly for the open vowel /a/. The variations of F1 show a trend for speakers to produce short vowels with a less open jaw and/or higher (less low) tongue position. F2 values show a clear decrease for front vowels (/i/ and /e/), an increase especially for back vowels (/u/ and /o/) and, to a lesser extent, for central vowels (/a/ and /ø/). Thus the production of short vowels appears to be less peripheral (front and back vowels show more central values). The results show that the male vowel space is more reduced than female vowel space and tends to have greater reduction for back vowels than for front ones (figure 1). Note that, for both male and female speakers, the first three formants of the four vowels /a/, /e/, /ø/ and /o/ are extremely close for short durations (set 1, Table 3). This lack of acoustic separation may also suggest that these four vowels may be difficult for listeners to distinguish. A decrease in F3 is observed for the front vowels (/i/ and /e/) as duration decreases, while the other vowels show an increase of F3 for short vowels. F3 variations tend to show that, for short vowels, lip position may be less extreme (less rounded for rounded vowels and less spread for unrounded vowels).

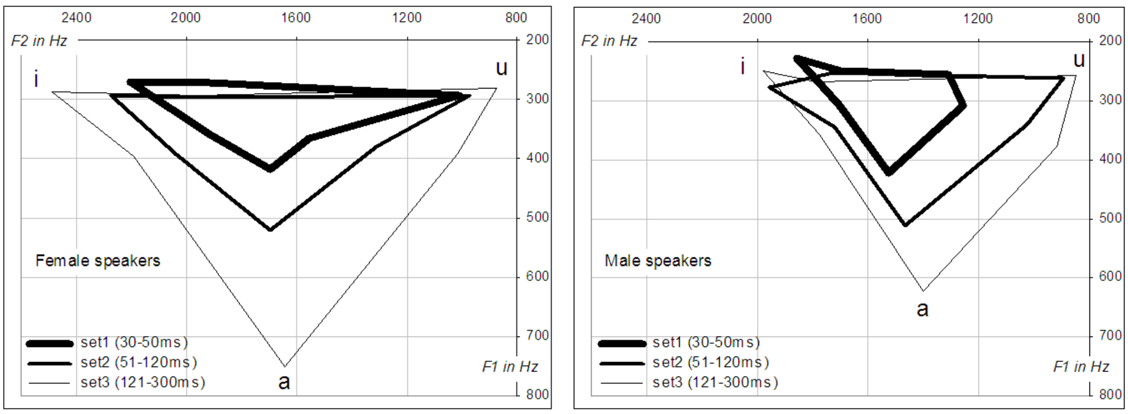

Figure 1: Vowel triangles for female (left) and male (right) speakers (mode of F1 and F2 values in Hz) according to duration sets (Set 1: 30-50ms, Set 2: 51-120ms, Set 3: 121-300ms).

These results are in agreement with those obtained in previous studies (Lindblom, 1963; Gendrot & Adda-Decker, 2005) and confirm the relationship between decrease in vowel durations and hypoarticulation. Although the comparison is not perfect (in Gendrot & Adda-Decker, 2005, only the mean formant values are given), it seems that the spectral reduction observed is greater in our corpus, especially for /a/ which is more reduced in the CID than in Gendrot & Adda-Decker's study. Moreover, our results provide different patterns of reduction in male and female speakers' productions.

3.2. Lexical factors

In this section, we discuss duration analyses conducted on all vowel data, since duration variation affects all vowels in the same way. Spectral analyses, however, were not conducted on all vowels for several reasons. First, vowel spectral data have to be analyzed separately for each vowel (the variation observed depend on spectral parameters) and this complicates the analyses and their interpretations. Second, spectral variation depends on vowel identity, and certain analyses may be irrelevant for particular classes of vowels (for closed vowels like /y/ and F2 for the central vowel /ø/, for example). Third, in this section, we will focus not on vowel-specific effects, but rather on lexical factors. We therefore carried out spectral analyses on the F1 of /a/, since we observed particularly striking spectral variation related to duration variation for this vowel (Figure 1). Furthermore, there are many tokens of /a/ since it is the second-most frequent vowel in our corpus. Therefore we considered the analysis of F1 of /a/ to be a good starting point to provide results of vowel spectral reduction according to lexical factors.

3.2.1. Within-word position effects

Several studies (Lindblom, 1963; van Bergem, 1993) have shown that vowels in unaccented syllables are more often reduced than those in accented syllables. This seems to be a particularity of languages with lexical stress. French does not have lexical prosodic characteristics (Hirst et al., 2001), and there is no lexical stress in the language. The location of stress is fixed at the word level, but the domain (the Accentual Phrase, AP) of its realization can be larger than the word (Jun & Fougeron, 2002 and references therein). Consequently, the final strong syllable (not including word-final schwa) of a word is realized with a longer duration only if it is the end of an AP (Dell, 1984). The AP may thus correspond to single word or to a group of words. Consequently, our hypothesis is that reduced vowels will appear in non-final syllables of words. Since we do not have a complete annotation of the APs in the CID corpus, vowels were analyzed according to their syllable position in the word. Each word was syllabified individually on the basis of its actual phonetic realization (for example, if the bisyllabic word *petit* 'small' /pə.ti/ was realized with the vowel /ə/ elided, the vowel /i/ was therefore embedded in the monosyllabic item /pti/). We considered five different sets of words, defined by word length (1, 2, 3, 4 and 5 syllables), and we examined the target vowels at each syllable position within the word. For example, /i/ is in second syllable in the three-syllable word *papillon* ‘butterfly’, and /y/ is in first syllable in the monosyllabic word *tu* ‘you’. The number of vowel tokens examined for each set is given in Table 4. Words with 6 or 7 syllables were disregarded, given their very low frequency in the corpus.

Table 4 Number of vowels according to word length (1, 2, 3, 4 and 5 syllables).

Number of syllables	1 *lit* 'bed'	2 *bateau* 'boat'	3 *papillon* 'butterfly'	4 *hippopotame* 'hippopotamus'	5 *caractéristiques* 'characteristics'
Number of vowel tokens	41,117 (57%)	20,734 (29%)	6,998 (10%)	2,444 (3%)	392 (1%)

For each set, we fitted a linear-mixed model (LME), with the logarithm of the vowel duration as the dependent variable. The fixed effect was the position of the vowel in the word, expressed as an *n*-level factor for a *n*-syllable word. Random intercepts were included for speakers.

Figure 2 shows estimated vowel durations in the different conditions. The results show that vowels of the final and penultimate syllables of words are longer than those of earlier syllables. As an example, for 3-

syllable words, the vowel of the second syllable (54 ms) was significantly longer than the vowel in first position (49 ms); similarly, the vowel in third position (68 ms) was significantly longer than the vowel in second position (54 ms). The significant p-values (MCMC p-values) were all less than 0.003, while the non-significant p-values were all greater than 0.46.

The observed duration increases were quite regular. The mean durations of the final syllable vowels were very similar (64-68 ms), regardless of the number of syllables in the word. The one exception was the set of 5-syllable words, for which the final syllable was longer (77 ms). The mean duration of vowels in monosyllabic words was similar to that of vowels in the final syllables of multisyllabic words.

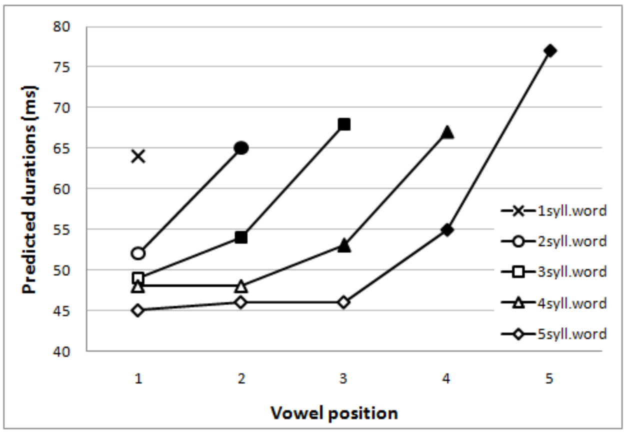

Figure 2: Vowel durations (ms) estimated by the model. Durations are plotted as a function of position of the vowel in the word and of word length. The four curves represent vowel duration according to the position of the vowel in words from 2 to 5 syllables long. The isolated unfilled circle represents monosyllabic words. Filled circles indicate that the value differs significantly from the preceding one.

In our study, the estimated duration of the final syllable is the mean duration across unstressed and stressed final syllables. Consequently, as the unstressed syllables are not lengthened, at least relative to the AP, it is important to interpret these lengthening values with caution. The real values of stressed vowel lengthening are thus probably longer than the values we mentioned. Our results are consistent with a recent description of French radio speech; Adda-Decker et al. (2008) observed that long syllables (>100ms) are more often in final syllables while short ones (<50ms) are more often in non-final ones. Moreover, we observed that vowel lengthening also affects the penultimate syllable as well. An explanation could be that the stress for word with final schwa would fall on the penultimate syllable (since schwa cannot be stressed). But a complementary study (where words with final schwa were excluded) shows the same trends. In fact, the lengthening of the penultimate syllable has been observed – but not quantified – in a previous study (Di Cristo, 1985). This may reflect a tendency for this lengthening phenomenon to cross syllable boundaries.

In addition to Adda-Decker's study, we analyzed formant values in order to observe if vowel quality is affected by duration variation within words, and thus by their position in the words. As noted above, we focused on spectral variations of the F1 of /a/. We eliminated the 5-syllable word set, since there were too few tokens (0.5% of the total) containing /a/ in the final syllable. A total of 16,664 /a/ tokens were analyzed (Table 5).

Table 5 Number of /a/ tokens according to word length.

	Number of syllables in word			
	1	2	3	4
Number of /a/ tokens	9,830 (59%)	4,809 (29%)	1,401 (8%)	624 (3.5%)

For each set, we fitted a linear-mixed model, with the logarithm of the F1 (in Hz) value as the dependent variable. The fixed effect was vowel position in the word expressed as an *n*-level factor for an *n*-syllable word. Random intercept terms were included for speakers.

The results are quite consistent with the lexical pattern observed for durations. F1 values are significantly higher for /a/ in final syllables (Figure 3), consistent with lower tongue position and/or a more open jaw, while F1 for /a/ in non-final syllables was lower, and thus farther from prototypical values for this open vowel (Table 1, F1). For each set, the F1 of the last vowel differs significantly from the preceding one. The significant p-values were all less than 0.001, while the non-significant p-values were all greater than 0.067. The vowel of the penultimate syllable was reduced like the earlier ones. The very small, although significant, duration differences observed between penultimate vowels and preceding ones were probably not sufficient for formant differences to emerge for this specific vowel (/a/).

We note a difference between reduction in duration and spectral reduction: F1 values for /a/ in monosyllabic words were similar to those for /a/ in non-final syllables, while durations of vowels in monosyllabic words were similar to those of vowels in final syllables. Consequently, vowels in monosyllabic words seem to have reduced spectral values but unreduced durations.

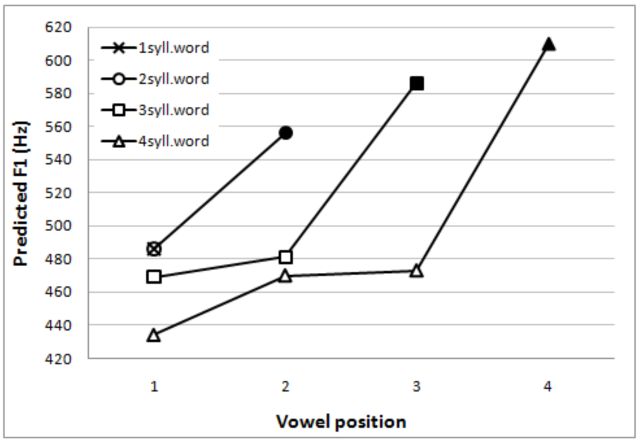

Figure 3: Estimated F1 of /a/ as a function of its position in the word and of vowel length. The three curves represent F1 according to the position of the vowel in words from 2 to 4 syllables long (F1 values for /a/ in monosyllabic words and in the first position of bisyllabic words are identical). Filled circles indicate that the value differs significantly from the preceding one.

The final lengthening can be interpreted as an effect of the final stress in the AP. This stress is marked on the final syllable (of the word), therefore the final syllable is more likely to be stressed and thus lengthened. Our results do not demonstrate that final vowels are never reduced since some of them might be reduced (within the AP domain) and some others are not (end of the AP). Nevertheless, our data shows that vowels in non-final syllables are reduced. This reduction affects the durations of vowels (as shown by Adda-Decker in a different corpus) but also vowel spectral quality, as the decrease of the F1 of /a/ in our study suggests.

3.2.2 Effect of word frequency and word category

The type of word used in conversation may also have an influence on how vowels are realized, according to the frequency of words and their informativeness. Frequency and informativeness are lexical factors which may influence phonetic reduction: when a word is highly frequent and also highly predictable, listeners may need less phonetic information to identify it. Similarly, words which are poorly informative (conjunctions, prepositions, etc.) and syntactically highly predictable in an utterance may be reduced without disturbing the listener's comprehension of the message. Indeed, Adda-Decker (2008) found that vowels in function words are more often short than vowels in content words.

To investigate this question, we analyzed vowel reduction according to two lexical factors: word frequency and word category. We computed word frequencies based on the CID corpus itself since frequencies computed from conversational speech might differ from those of written texts or more controlled speech (newspapers or broadcast news for example). Word frequency was defined as: the number of token occurrences/the total number of tokens in the corpus (about 110,000 tokens). We analyzed only monosyllabic words in order to control the variation in duration due to the position of the vowel within the word (see section 3.2.1."Within-word position effects"). Oral vowels in monosyllabic words represent 57% (41,592 vowels) of the total number of oral vowels in the corpus (72,620).

We distinguished two word categories: content words (nouns, verbs, adjectives and adverbs) and function words (prepositions, conjunctions, pronouns and auxiliary verbs). Content words have a higher semantic load than function ones. These two categories represent 84% of the monosyllabic productions. Other productions (interjections like *euh, ah, ben, hum*, etc.) were not taken into account. There were 25 monosyllabic structures in the CID corpus, however the three most frequent structures accounted for 81% of the monosyllabic words: CV (45%, n=15,708), V (24%, n=8,275) and CVC (12%, n=4,414). Moreover, the V structure was excluded from the analyses. The phonetic neighborhood of a vowel in a V syllabic structure is obviously unpredictable. We reasoned that a variable phonetic context and a possible resyllabification by speakers might cause specific variations of duration which are not relevant for our analysis. For instance, the sequence *il est parti* ‘he has gone’, produced /i.le.par.ti/ with /e/ in a CV syllable, but considered as a V syllable, since tokens are syllabified independently. Rare phonetic sequences, i.e. less than five occurrences in the corpus (about 1.8% of the vowels), were excluded. A total of 19,312 vowels distributed over 202 phonetic sequences remained (Table 6).

Table 6 number of vowels in the two word categories according syllabic structure.

	Content words	Function words	Total
CV	3,276	12,538	15,814
CVC	2,256	1,242	3,498

With this type of corpus, word occurrences cannot be controlled and are therefore highly unbalanced. For instance, the 4th quantile of the lexical frequency distribution of the CVC function words contained 380 tokens. But these 380 tokens were all occurrences of a single word type, *pour* /pur/ ‘for’. Similarly, considering the distribution of the lexical frequency of the CV function words, 95% of the tokens with a lexical frequency above the median is represented by only 6 word types. Models with lexical frequency taken as a continuous parameter could therefore be biased.

We therefore transformed lexical frequency to a 2-level factor (henceforth *dichotomized lexical frequency* or DLF). Within each of the four subsets of Table 6, this factor is equal to *H* (high) if the lexical frequency is above the median value of the subset, *L* (low) otherwise. We ran separate analyses for CV and CVC words, since there were too few different CVC function words (types) with a high lexical (token) frequency to safely analyze both the effects of lexical frequency and word category for the CVC set.

CV words

We ran a LME with the logarithm of the vowel duration as the dependent variable, the dichotomized lexical frequency (DLF) and the word category as fixed effects. Random intercepts were added to account for the variability across the 16 speakers and the 75 phonetic sequences. The interaction coefficient between the DLF and the word category was not significant (MCMC p-value= 0.16). A model without this interaction term showed that DLF has no effect (MCMC p-value = 0.9), and that the morpho-syntactic category is highly significant (β=-0.19; se=0.018; t=-10; MCMC p-value <0.001). The estimated vowel duration was 57 ms for the function words and 69 ms for content words (Figure 4).

CVC words

As mentioned above, the influence of lexical frequency on duration can only be analyzed for the content words. We first studied the influence of the lexical frequency on the content words, and then analyzed the word category effect.

We ran a LME with the logarithm of vowel duration as the dependent variable and the DLF as the single fixed effect. The random terms were the same as those in the CV model (there were 117 phonetic sequences). The DLF was significant (β=0.11; se=0.037; t=2.98; MCMC p-values < 0.01). The estimated vowel duration was 64 ms for the high frequency words, and 71 ms for the low frequency words. This difference is slightly below the value of the resolution of the aligner (8 ms), and therefore must be interpreted with caution.

We investigated the influence of the word category separately, with a LME with the single factor word category as fixed effect. The random terms were the same as those in the CV model (there were 124 phonetic sequences). The word category factor was clearly significant (β= -0.208; se=0.026: t=-7.9; MCMC p-value < 0.001). The estimated vowel duration was 56 ms for function words, and 71 ms for content words (Figure 4). These values are very close to those obtained for the CV words.

We then asked whether these duration differences would correspond to spectral differences. We then examined variations of the F1 of /a/. The influence of word frequency on the F1 of /a/ was not analyzed, since we did not observe an effect of lexical frequency on vowel duration. We therefore focused on the influence of word category on F1. There were 5,360 monosyllabic words with the /a/ vowel (2,545 function words, 2,815 content words).

An LME was estimated, with the logarithm of F1 as the dependent variable and word category as fixed effect. Random intercepts were added to account for the variability across the 16 speakers and the 31 syllabic structures. The effect of word category was significant (β= -0.111; se=0.0158; t=-6.96; MCMC p-values <0.001). The estimated F1 values appear in Figure 4.

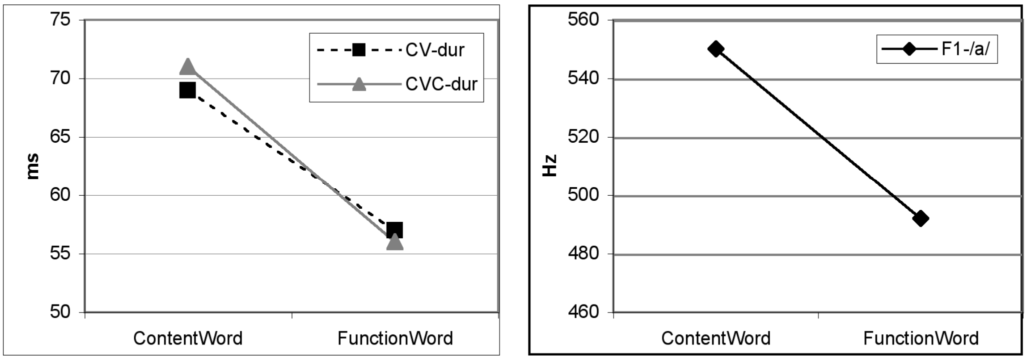

Figure 4: Estimated vowel durations (left in ms) and estimated F1 of /a/ (right in Hz) in monosyllabic words according to Word Category

To summarize, we did not find a clear effect of word frequency on vowel durations. A few words were very frequent in the corpus, and this complicated the analysis. But frequency and word category are highly related: most function words are very frequent, while most content words are rare. On the other hand, a clear effect of word category was found: content words show longer vowel durations and higher F1 of /a/, while function words show shorter vowel durations and lower F1 of /a/. Thus, as expected, function words are more reduced than content words. Nevertheless, once again several factors may interact: function words appear less often at the end of the AP in French (although they may occur in this position in spontaneous speech, as in *t'es où?* 'where are you?', *tu viens là?* 'are you coming here?'). Consequently, the reduction of function words may be attributed to their specific category but also to their position in the stress domain.

4. Discussion

Our goals were 1. to provide an overview of vowel reduction in casual speech in French and 2. to evaluate a possible effect of lexical factors on vowel reduction. The factors examined were the position of the vowel within the word, on one hand, and lexical frequency and word category, on the other.

Firstly, our study, based on conversational speech, confirms a recent study comparing French with other languages (Gendrot & Adda-Decker, 2007), and showing that vowel duration and spectral quality are highly related. Moreover, our study provides a precise quantification of vowel reduction and shows that spectral reduction is more pronounced in casual speech.

Secondly, the study provides clear evidence of a relationship between vowel position in the word on one hand, and duration and spectral reduction, on the other. Due to the Accentual Phrase, vowels in the final (and, to a lesser extent, penultimate) syllables of words may be lengthened, while earlier vowels have shorter durations. In another study on French speech, Adda-Decker et al. (2008) found similar results, except for penultimate syllables. Moreover, our results show that the spectral quality of vowels is also affected by the position of the vowel in the word: F1 values of /a/ are lower in non-final vowels, suggesting that non-final vowels are reduced in multisyllabic words. Previous studies show that unstressed vowels are reduced in languages with lexical stress like Swedish (Lindblom, 1963) or Dutch (van Bergem, 1993). We find the same results in French, although the language does not have lexical stress.

Thirdly, our results show that the production of French vowels depends on word category. Vowels produced in monosyllabic function words were more reduced than those produced in monosyllabic content words. These results are in agreement with those of van Bergem (1993) for Dutch vowels observed in controlled speech. As van Bergem notes, it is difficult to determine whether the production of reduced vowels in function words is due to the small "semantic load" of these words or to their high frequency of occurrence. He concluded that it might be due rather to high frequency because a frequency effect is also observed across other word categories in his study. Our results for French, however, are not in agreement with these conclusions, or with other previous results showing frequency effects (Pluymaekers et al., 2005; Ernestus et al., 2006). According to these studies, we expected that frequent words would be produced with reduced vowels in French, but our findings did not confirm an effect of frequency on vowel reduction. When words were relatively balanced across the different categories, no clear frequency effect was observed.

The analyses may be biased by the very high frequency of some function words in the CV and CVC structures, since spontaneous speech corpora are obviously not controlled. Our study concentrated on monosyllabic words, which are very frequent in the corpus (57% of the words of the corpus are monosyllabic, and the most frequent words are monosyllabic). It was impossible to analyze bisyllabic or longer words to compare function words and content words, since the number of function words is very low in these structures. There are only a very few monosyllabic types of function words, but these types are very frequent. At the other extreme, the number of types of content words is very high, but most of these types occur infrequently. How therefore should duration variations be interpreted? These variations might be due to the durations of a specific vowel if this vowel is present in a very frequent word, or to specific phonetic sequences (phonetic context) in monosyllabic words (for example, 75% of the /a/ tokens in monosyllabic function words appear in only two words). How can a very few words with many repetitions (tokens) be fairly compared to a great number of words with very few repetitions? In the same way, the effect of word category (vowels in function words are more reduced than those in content word) is hard to attribute to the single category factor, since function words appear more often in unstressed position. These are the limits of some analyses on natural speech. Some data are difficult to compare because they are too 'unbalanced' in natural speech. Studies on non-controlled speech (i.e. conversational or spontaneous speech) are obviously constrained by the real distribution of words in speech.

To summarize, we showed a significant effect of syllable position and word category on vowel reduction. Nevertheless, we must be careful not to over-interpret these effects, since the size of the variations observed seems to be only a small part of the general variation in the vowels of the corpus. Spontaneous speech is characterized by specific pragmatic and discourse functions. We hypothesize that a great part of vowel reduction may be conditioned by prosody and discourse constraints (Beyssade et al., 2004). We are particularly interested in investigating in more depth the prosodic and spectral characteristics of the vowels of function words. Some function words, such as *mais* 'but' or *et* 'and' have particular discourse functions (hesitation, emphatization, etc.), and we hypothesize that vowel reduction in spontaneous speech may be even more complex than a simple distinction between function words and content words.

References

Adda-Decker, M., Gendrot, C., Nguyen, N. (2008). Contributions du traitement automatique de la parole à l'étude des voyelles orales du français. *Traitement Automatique des Lang*ues, *49* (3).

Baayen, R. H. (2008). *LanguageR: Data sets and functions with Analyzing Linguistic Data: A practical introduction to statistic*s. R package version 0.953.

Bartkova, K., Sorin, C. (1987). A model of segmental duration for speech synthesis in French. *Speech Communication, 6 (3)*, 245-260.

Bates, D., Maechler, M. and Dai, B. (2008). *lme4: Linear mixed-effects models using S4 classes*. R package version 0.999375-27.

Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., Rauzy, S. (2008). Le CID. Corpus of Interactional Data. Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement Automatique des Langues, 49 (3)*.

Beyssade C., Delais-Roussarie, E., Doetjes, J., Marandin, J.M., Rialland, A. (2004). Prosody and Information in French. In *Handbook of French semantics*, F. Corblin & H. de Swart (eds). CSLI.

Blache, P. & Rauzy, S. (2008). Influence de la qualité de l'étiquetage sur le chunking: une corrélation dépendant de la taille des chunks . *Proceedings of Traitement Automatique des Langues Naturelles (TALN) Conference*, Avignon, 290-299.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261-290

Dell, F. (1984) L'accentuation dans les phrases en français. In *Forme sonore du langage : Structure des représentations en phonologie*, ed. by F. Dell, D.J. Hirst & J.R. Vergnaud, J.R, 65-122. Paris: Hermann.

Di Cristo, A. (1985). De la Microprosodie à l'Intonosyntaxe, *Thèse de Doctorat d'Etat*.. Publications de l'Université de Provence, Jeanne Lafitte eds, Marseille.

Di Cristo, A., and Di Cristo, P. (2001). Syntaix, une approche métrique-autosegmentale de la prosodie, *Traitement Automatique des Langues, 42*, no.1, 69–111

Durand, J. & Lyche, C. (2004). Structure et variation dans quelques systèmes vocaliques du français: l'enquête Phonologie et Français Contemporain (PFC). In *Variation et Francophonie*, A. Coveney & Sanders, C. (eds), Paris, L'Harmattan, 217-240.

Ernestus, M., Lahey, M., Verhees, F. & Baayen, R.H. (2006). Lexical frequency and voice assimilation. *The Journal of the Acoustical Society of America 120*, 1040-1051.

Fohr, D., Mella, O., Cerisara, C., and Illina, I. (2004). The automatic news transcription system: ANTS, some real time experiments. *Proceedings of Interspeech Conference*, 377–380.

Gendrot, C. & Adda-Decker, M. (2005). Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and german. *Proceedings of Interspeech Conference*, Lisbon, Portugal, 2453–2456.

Gendrot, C. & Adda-Decker, M. (2007). Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages. *International Conference of Phonetic Sciences*, Saarbrücken, August 2007, 1417-1420.

Hirst, D.; Di Cristo, A.; Nishinuma, Y. (2001) Prosodic parameters of French: A cross-language approach. In *Contrastive Studies of Japanese and Other Languages Series*, National Institute for Japanese Language, Tokyo, 7-20.

Hume, E. (2004). Deconstructing markedness: A Predictability-based Approach. *Proceedings of Berkeley Linguistic Society*.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (eds.) *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium.* Tokyo, Japan, 29-54.

Jun, S. & Fougeron, C. (2002) Realizations of accentual phrase in French intonation, *Probus* 14: 147-172.

Jurafsky, D., Bell, A., Gregory, M. and Raymond, W.D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. *Frequency and the emergence of linguistic structure*, edited by J.L. Bybee and P. Hopper (Benjamins, Amsterdam), pp. 229-254.

Lindblom B. (1963). Spectrographic study of vowel reduction, *The Journal of the Acoustical Society of America, 35*, 1773-1781.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the hyper- and hypospeech theory. *Speech Production and Speech Modelling*, Hardcastle and Marchal (eds.), Kluwer Academic Publishers, 403-439, 1990.M.

New B., Pallier C., Ferrand L., Matos R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE. *L'Année Psychologique, 101*, 447-462.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag.

Pluymakers, M. Ernestus, M., & Baayen, H. (2005). Lexical frequency and acoustic reduction in spoken Dutch, *Journal of the Acoustical Society of America*, 118, 2561-2569.

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

R Development Core Team (2008). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0,URL http://www.R-project.org.

Torreira, F. & Ernestus, M. (2010). The Nijmegen Corpus of Casual Speech. *Speech Communication*, 52, 201-212.

Tubach JP. (1989). *La parole et son traitement automatique*. Masson. Paris.

van Bergem, D. R. (1993) Acoustic vowel reduction as a function of sentence accent, word stress, and word class, *Speech Communication* 12, 1-23.