IMPACT OF PROSODIC POSITION ON VOCALIC SPACE IN GERMAN AND FRENCH

Cédric Gendrot, Kim Gerdes & Martine Adda-Decker

LPP-CNRS, Université de Paris 3, France
cgendrot@univ-paris3.fr; gerdes@univ-paris3.fr; madda@univ-paris3.fr

ABSTRACT

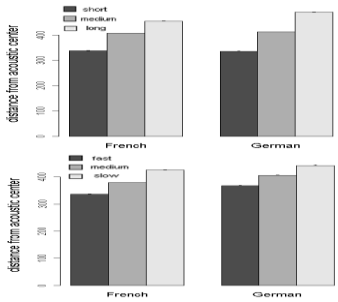
The relation between prosodic constituents of different levels (the prosodic hierarchy) and the realization of phonemes is investigated in French and German broadcast speech. Four levels of prosodic categories are defined and the boundaries are automatically derived using syntactic chunking and forced alignment. We show that the higher a vowel position within the prosodic structure, the more strengthened are its acoustic-prosodic correlates. Our results suggest that the influence of the prosodic hierarchy is slightly weaker in German, supposedly due to its lexical stress patterns.

Keywords: vowels, formants, prosody, strengthening, accentuation

1. INTRODUCTION

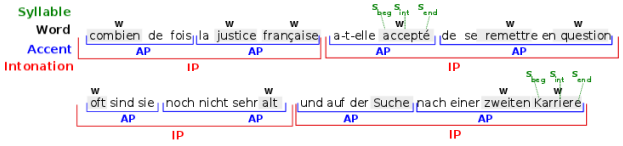
This work is part of a large scale study aiming at describing the variability of vowels in different languages. With the help of very large automatically segmented corpora, we were able to study a large number of contexts known to influence the realization of phonemes so as to quantify their influence and their interaction. In previous studies on broadcast speech [7], it was observed that longer vowels are considerably strengthened in terms of their formants compared to shorter vowels. They occupy a much larger acoustic space with acoustic values close to their canonical realizations. Formant values of vowels can give a good hint on their articulation. Indeed, F1 can assess for vowel aperture, notably for open and semi-open vowels, while F2 is more dependent on frontness/backness of the vowel. As for F3, it gives an indication of rounding. For example, a mid-open front vowel such as /ɛ/ will be more open and more anterior when strengthened/longer. Vowel duration, which changes with speaker's style and speech rate is known to influence vowels acoustic realizations [9]. In figure 1, speech rate was measured as the number of phones per second, excluding pauses. Three duration categories were considered: (1) slow (<11 ph/sec), (2) medium (11-16 ph/sec), and (3) fast (>16 ph/sec). The distance from acoustic center used in this figure is a measure to evaluate vowel strengthening as detailed in section 2.1. We show that, as noticed for vowel duration [7], the slower the speech rate, the more strengthened the vowels for French and German. Also, linguistic factors such as the surrounding phonemic context, the phoneme position in the syllable, the word, the syntagm or the utterance can be of great influence. The latter four are considered as prosodic constituents since specific intonation and duration patterns are expected and may serve demarcative functions. The realized prosodic constituents are supposed to be organized in a prosodic hierarchy, each constituent being embedded in a higher one [11].

Figure 1: Distance from acoustic center according to vowel duration (left) and speech rate (right).



It has been shown that the higher the level of the prosodic constituent, the longer and the more strengthened (in articulatory terms) the vowels are at the boundaries of these constituents (for French [4, 14]). This study aims at quantifying the influence of prosodic constituents in French and German on the spectral realization of vowels. The major difference with previous studies mentioned is the use of continuous speech rather than controlled read speech and the simultaneous study of all peripheral vowels. Productions of speakers are thus not directed. Data are more natural than those of read speech but their prosodic categories cannot be controlled apart from detecting them a posteriori. Four widely accepted categories for prosodic constituents were chosen: syllables, words, accentual phrases and intonational phrases (from lower to a higher level). These categories will be automatically detected and analyzed in their initial and final positions, i.e. at their boundaries. We only focus on peripheral vowels for an approximation of vocalic space used by vowels.

Figure 2: Illustration of the 4 investigated prosodic levels: syllable (S) distinguishing word beginning, word internal, word final positions, word W, accent phrase A and intonation phrase I. Only lexical content words (W) are considered for our analyses. An example is shown for French and German.



2. METHOD

2.1. Corpus and analysis

The corpora correspond to approximately 30 hours of speech for French (including about 500 men and 300 women) and 20 hours for German (350 men and 150 women). Both data sets correspond to broadcast news and documents. The French data were recorded and transcribed orthographically at the French CTA/DGA [5]. The German data correspond to ARTE documents of an FP5 project. Such speech cannot be described as fully spontaneous, but rather as prepared speech. The speech transcription system used for words and phonemes alignment is detailed in [1, 7]. Measurements of the first 4 formants (F1, F2, F3 & F4) were taken automatically with the Burg algorithm implemented in Praat [2]. Large threshold values were determined so as to filter out erratic values (approximately 4%). More details about this procedure can be found in [1]. So as to reduce the inter-speaker variation in our results, data were normalized by speaker according to the z-score Lobanov transform [10]. We hypothesize that vowels at the boundaries of higher prosodic constituents in French and German are strengthened when compared to lower prosodic constituents. To explore this hypothesis, we measure the acoustic space used by all vowels for each of the four prosodic levels (keeping initial and final positions apart). To this aim, we first compute an acoustic center using all vowels of the whole data set (around F1: 450Hz; F2: 1450Hz for both languages); then, for each vowel an Euclidean distance is computed with respect to the acoustic center. If the vowel's distance to the center makes a significant in- crease from one category to a higher one, then it is considered as strengthened. We bear in mind that this measure might be somewhat inappropriate as it is related to vowel centralization, which is only a secondary effect of vowel coarticulation. However, with all vowels moving away simultaneously from the acoustic centre, they necessarily get away from one another, thus favoring their phonemic identification (see Lindblom's theory of adaptive dispersion for an interpretation of this in the shape of vocalic systems).

2.2. Selection of prosodic categories

We describe below choices that were made to select the four analyzed prosodic categories (from a low to a higher level: syllables, words, accentual phrases and intonational phrases). Vowels in these categories are analyzed at their initial and final positions, i.e. at their boundaries.

Syllables were determined from phonemic segmentation. Syllabation rules [12] were used to segment the continuous string of phonemes. We did not take word delimitations into account for French but we did for German. Since it was difficult to consider initial and final syllable boundaries that would not overlap, we thus decided to consider for the syllabic level word syllables that are neither word final nor word initial, i.e. word internal.

Words boundaries were obtained from the manual transcription and automatically segmented by the alignment system. Vowels considered as initial are in fact among the first two phonemes of the word (for /a/, 'armée' [aʁme]: armée as well as 'partir'[paʁtiʁ]: to leave) and vowels considered as final were taken from the last two phonemes of the word (for /i/, 'partie' [paʁti]: party; 'partir' [paʁtiʁ]: to leave). Very similar results were obtained if we only considered absolute initial or final vowels but both for French and German, words starting with vowels are scarce.

The third level analyzed in our prosodic hierarchy is what we identified as accentual phrases (see (1) and (2) for an example). They were obtained from a syntactic chunking based on automatically established grammatical categories, according to the following rules:

(i) Nouns, prepositions and verbs were gathered with their closest surroundings (clitics, determinants, adjectives, etc.).

(ii) All sequences of words not defined by the previous rule.

Then, three merging rules were applied:

(a) Merging of any segment ending on an auxiliary or a modal with the following segment.

(b) Merging of any verbal segment with the following segment if the whole resulting chunk is less than seven syllables.

(c) Merging of any other sequence of segments of less than seven syllables.

Chunks generated by this algorithm may still have more than seven syllables, if the previous rules allow for it, for instance the chunk 'avec qui j'ai pu m'entretenir' (with whom I could talk) is very difficult to chunk further. The seven syllables rule applied here - taken after [16] - is a rough limitation to avoid overly long chunks. It may be supplanted in

future studies by a duration threshold which would be more suited to be adapted to different speaking styles, and more importantly, we should allow different values for French and German since words length in German is much longer in general. This category is meant to be as close as possible to the accentual phrase although we are aware not all these chunks are 'accented', i.e. bearing a generally rising tone and/or accompanied with a lengthening of the last syllable. We could also have combined this morphosyntactic information with prosodic information such as lengthening to make sure the accentual phrase was indeed one, but that would have been a somewhat circular procedure as the longest vowels are strengthened in the first place. Hence, this procedure aims at evaluating accentual phrases on a syntactic (underlying) level rather than based upon their prosodic characteristics.

The fourth investigated prosodic category is the intonational phrase. This was automatically detected with the help of pauses, the latter being taken into account when longer than 500ms only, on the account that intonational phrases are frequently preceded/followed by pauses. A distinction was first made between rising and falling contours with no significant differences. A summary of the obtained categories is detailed below.

3. RESULTS

For sake of space, all vowels are regrouped in the graphs but the analysis considered each vowel separately. We can notice at a first glance on Figure 3 (lower part) that the distance from the acoustic center increases with the level of the prosodic hierarchy, i.e. vowels occupy a much larger vocalic space in the highest prosodic category (IP), and this gradually for every analyzed level. More details are given in the discussion. Vowel durations were also compared revealing rising values with the level of the prosodic hierarchy for French. However German differentiates from French with the "word" category having higher values for initial positions. We hypothesize that this is due to accentuation. Indeed accentuation (not shown here), usually located in the words initial syllables, involves a longer duration and a strengthening of the vowel with higher dispersion.

No such differences were found between languages for final positions. As an illustration, Figure 4 shows the duration and dispersion values of vowels according to the syllable position (the syllable bearing the measured vowel is either beginning, final or intermediary in the word. We can see that French is characterized by a lengthened first, and mostly last syllable while less variation is observed for German with a slightly longer first and final syllable.

Figure 3: Duration (upper) and distance from acoustic center (lower) at initial positions according to the prosodic hierarchy (S: syllable, W: word, AP: accentual phrase, IP: intonational phrase).

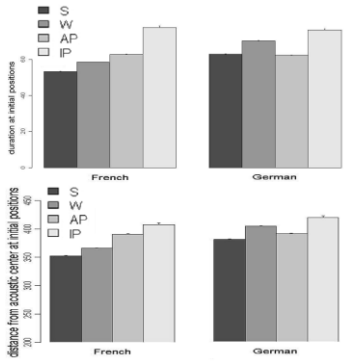


Figure 4: Duration (upper) and distance from acoustic center (lower) at final positions according to the prosodic hierarchy (S: syllable, W: word, AP: accentual phrase, IP: intonational phrase).

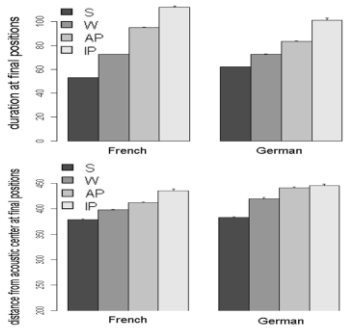
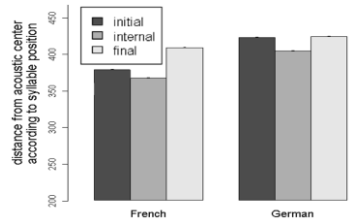


Figure 5: Distance from acoustic center according to syllable position in the word.



4. SUMMARY & PROSPECTS

As hypothesized, we observe a general prosodic hierarchy (from syllable to word, then accentual phrase and finally intonational phrase) based on spectral measurement results. It shows that the higher a vowel is in the prosodic hierarchy, the more strengthened it is.

However, as noticed by previous studies [4, 14, 15], all levels cannot be statistically distinguished from one another in a systematic manner. That is to say, some vowels are less variable according to the different prosodic positions analyzed.

For example for French, /i/ and /y/ were noticed to be less variable on the F1/F2 plan. A further

investigation on F3 and F4 movements showed that when moving up the prosodic hierarchy, /y/ is characterized by a narrowing of F2 and F3, while /i/ is characterized by a narrowing of F3 and F4, thus favoring their focal characteristics [13], notably more spreading for /i/ and more rounding for /y/. The point of interest in this result is the information it brings about the nature of strengthening. The hyperarticulation hypothesis [8] predicts that distinctive features of the phoneme will be enhanced when the phoneme is strengthened. This means that closed vowels will be even more closed, rounded vowels even more rounded, and so forth. Our data thus provide a strong hint for the hyperarticulation hypothesis.

Other vowels such as /ɔ/, /o/, and /u/ revealed variations that could be unexpected in some contexts. For example, we noticed that /ɔ/ has a non typical variation in (strictly) final positions with particularly high F2 values; as for /u/ and /o/, they are the least represented peripheral vowels (around 2% each for French) which may account for their somewhat less regular variations along the prosodic hierarchy. Also some vowels are frequently found in specific distributional contexts, including unaccented syllables for central vowels, which would explain reduction phenomena. The distribution for each vowel has to be further investigated in both languages.

Both initial and positions are characterized by strengthening on the higher levels of the prosodic hierarchy (accentual and intonational phrase), despite a large number of grammatical words at the beginning of these prosodic constituents.

Grammatical words could have been expected to be more hypoarticulated since they are more frequent lexicon items and don't carry as much information as lexical words. Indeed, abundant literature already exists about the linguistic information carried by words as well as their lexical frequency, and their implications in the articulation of these words (Lindblom and the Hypo- and Hyperarticulation theory as one mere example). The predictions emanating from these theories have been empirically tested on the acoustic realization of vowels (Wright, 2003, among others). We could thus have expected that boundaries of AP and IP might be more strengthened towards the end (mainly composed of lexical words) and not at the beginning (composed of grammatical words in majority). However, in preliminary analyses, we noticed that vowels formants were not subject to change when comparing grammatical with lexical words. We decided not to filter out these constituents with the additional assumption that in continuous speech, if the speaker does decide to signal boundaries to the listener, then he will do so whatever the grammatical category. The results confirm indeed this assumption, since it was measured that both initial and positions are characterized by strengthening on the higher levels of the prosodic hierarchy (group1 and group2), despite a large number of grammatical words at the beginning of these prosodic constituents One has to remember that beginnings of each accentual phrase or intonational phrase is also the end of a preceding one of the same level. As proposed by [3], these boundaries are moments of gesture slowering (pi-gesture) which favour hyperarticulation. It is thus fairly logical to observe the same phenomena at both initial and final positions of our analyzed prosodic levels, whatever the grammatical category involved.

5. REFERENCES

[1] Adda-Decker, M., Gendrot C., Nguyen N. 2008. Contributions du traitement automatique de la parole à l'étude des voyelles orales du fran cais. *TAL* 49(3), 13-46.

[2] Boersma, P., Weenink, D. 2009. *Praat: doing phonetics by computer* (Version 5.1.22) [Computer program]. Retrieved September 15, 2009. *http://www.praat.org/*

[3] Byrd, D., Krivokapic, J., Lee, S. 2006. How far, how long: On the temporal scope of phrase boundary effects. *JASA* 120, 1589-1599.

[4] Fougeron C. 2001. Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics* 29(2), 109-135.

[5] Galliano, S., et al. 2005. ESTER Phase II evaluation campaign for the rich transcription of French broadcast news. *Proc. of Interspeech* Lisboa, 1149-1152.

[6] Gauvain J.-L., et al. 2005. Where are we in transcribing French broadcast news? *Proc. of Interspeech* Lisboa, 1665-1668.

[7] 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. *Proc. of Interspeech* Lisboa, 2453-2456.

[8] de Jong, K.J. 1995. The supraglottal articulation of promi- nence in English, *JASA* 97, 491-504.

[9] Lindblom B. 1963. Spectrogaphic study of vowel reduction. *JASA* 35, 1773-1781.

[10] Lobanov, B. 1971. Classification of Russian vowels spoken by different speakers. *JASA* 49, 606-608.

[11] Nespor M., Vogel, I. 1986. *Prosodic Phonology*. Dordrecht: Foris.

[12] Pallier, C. 1994. *Rôle de la Syllabe dans la Perception de la Parole: Études Attentionnelles*. Ph.D. thesis, EHESS, Paris.

[13] Schwartz J.L., Boë L.J., Vallée, N., Abry, C. 1997. The dispersion-focalization theory of vowel systems. *J. Phon.* 25, 255-286.

[14] Tabain, M. 2003. Effects of prosodic boundary on /aC/ sequences: acoustic results. *JASA* 113, 516-531.

[15] Tabain, M., Perrier, P. 2005. Articulation and acoustics of /i/ in pre-boundary position in French. *JP* 33, 77-100.

[16] Vaissière, J. 1971. *Contribution à la Synthèse par Règles du Francais*. Ph.D. thesis, Université de Grenoble.