Prosodic hierarchy and spectral realization of vowels in French

Cédric Gendrot and Kim Gerdes

cgendrot@univ-paris3.fr, kim@gerdes.fr

Laboratoire de Phonétique et Phonologie, Université Paris3 Sorbonne Nouvelle, CNRS UMR7018

Abstract :

The aim of this study is to relate spectral realization of vowels and prosodic hierarchy in continuous speech. The IRISA speech alignment system is used and formant values of oral vowels are automatically measured in a total of 500,000 segments from around 30 hours of journalistic broadcast speech in French.

The link between the duration of vowels and their spectral realization (through their formant values) has been validated for a long time since Lindblom (1963), i.e. the longer the vowels the more hyperarticulated they are. Similarly, the relation between prosodic constituents of different levels (the prosodic hierarchy) and the duration of phonemes close to these boundaries has been demonstrated (for French, Fougeron 2001; Tabain 2003ab).

In this study we show that the level of prosodic constituent influences the spectral realization of vowels. Although significant differences can't be established for all levels and phonemes, we observe a general prosodic hierarchy (from syllable to word, then accentual phrase and finally intonational phrase) based on spectral measurement results, showing that the higher a vowel is in the prosodic structure of French, the more hyperarticulated it is.

1. Introduction

1.1. General comments and recalling previous results

This work is part of a larger scale study aiming at describing the variability of French vowels. 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 precisely their influence and their interaction.

In previous studies we observed that spectral realization of vowels was greatly influenced by their duration in French (Gendrot & Adda-Decker 2005; Gendrot & Adda-Decker to appear), i.e. longer vowels were considerably hyperarticulated compared to shorter vowels. In other words, the longer the vowels, the larger the acoustic space they will occupy, being thus more and more distinct from one another (figure 1). If we consider the result from an opposite view, the gap between the measured mean formant values and reference formant values is inversely proportional to vowel duration: a tendency to reduction for vowels of short duration clearly emerged. This result was measured for French as well as for seven other languages (German, English, Mandarin Chinese, European Portuguese, Spanish, Italian and Arabic in Gendrot & Adda-Decker 2007).

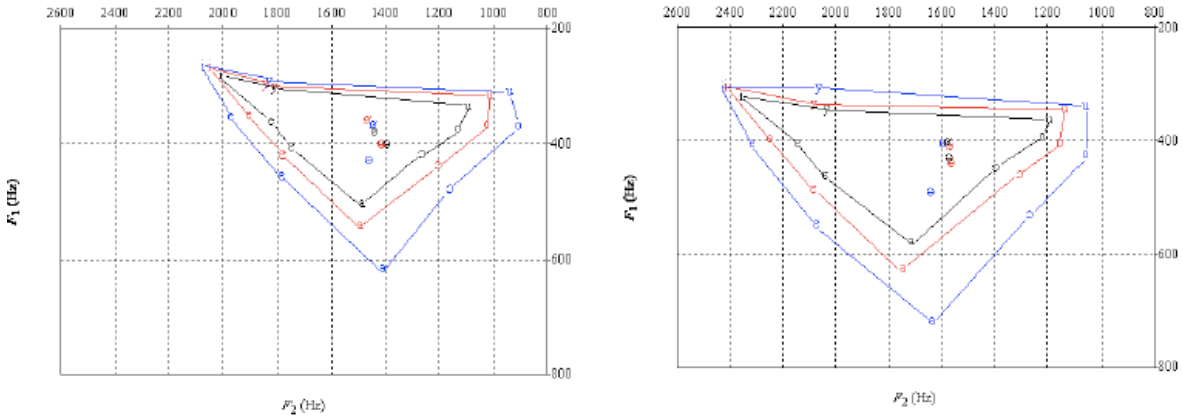


Figure 1: Measured mean average values of F1 and F2 for French vowels according to their duration. By ascending order from inside to outside (black [30 - 50ms], red [60 - 80], blue [90 - 110]). Male speakers on the left, female speakers on the right.

The link between spectral realization of vowels (through their formant values) and their duration has been validated for a long time by Lindblom (1963) and many others since then. It has been also shown that reduction of short vowels as can observed on figure 1 is not a mere 'centralization' of the acoustic space, but rather coarticulation: short vowels take the acoustic characteristics of consonants which draws them for most contexts towards the centre of the acoustic space. The relation between acoustic realization and articulation of vowels can be estimated from the formant measurements (Stevens 1997; Vaissière 2007 and references therein). Indeed, F1 can assess for vowel aperture, notably for open and semi-open vowels, while F2 is more dependant on frontness/backness of the vowel. As for F3, it gives a good indication of rounding, especially for vowels /y/, /ø/ and /œ/. Vowels /ø/ and /œ/ are not the object of this study as they are central vowels (for more details see Fougeron et al. 2007 or Bürki et al. to appear); we will only focus on peripheral vowels here for an approximation of vocalic space used by vowels. Variations of vowels /y/ (and /i/) are also further detailed in other studies as their different realisations may be more dependent on F3 movements (Gendrot et al. 2008). Our interest goes mainly towards the variations in terms of F1 (correlated to opening/closing) and F2 (roughly correlated to frontness/backness)

Reasons for variation in vowel duration are multiple. Factors such as speaker's style and speech rate can of course greatly influence, but 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 too. The four units lastly mentioned are considered as prosodic constituents since specific intonation and duration patterns produced by speakers may serve a demarcative function: 'they indicate the occurrence of the boundaries of words and phrases and, presumably, make it easier to identify such grammatical units in the stream of speech […] they are boundary signals that reinforce the identity and syntagmatic unity of words and phrases' (Encyclopedia Britannica).

The realized prosodic constituents are considered as being organized in a prosodic hierarchy, each constituent being embedded in a higher one: this is the strict layer hypothesis as developed by Nespor and Vogel (1986) for example. In this thread, the relation between prosodic constituent boundaries of different levels and the duration of phonemes adjacent to these boundaries has been demonstrated. It has been shown that the higher the level of the prosodic constituent, the longer the vowels are at the boundaries of these constituents (for

French, Fougeron 2001; Tabain 2003). In this study we aim at showing that the level of prosodic constituents in French in turn influences the spectral realization of vowels. Four unanimously accepted categories for prosodic constituents were chosen, from a low to a higher level: syllables, words, accentual phrases and intonational phrases. These categories will be automatically detected and analyzed in their initial and final positions, i.e. at their boundaries.

1.2. Previous findings and hypotheses

These positions have been more rarely investigated in terms of articulation and especially spectral measurements. Focusing on French, Tabain used EMMA (ElectroMagnetic Midsgaittal Articulatograph) on constituent final positions while Fougeron used EPG (ElectroPalatoGraph) on initial positions. Their results showed that for a phoneme in an initial or final position of a prosodic constituent, the higher the constituent in the prosodic hierarchy, the more strengthened or hyperarticulated the phoneme tends to be (see also Cho 2005, for American English). They also showed that this strengthening is not necessarily correlated with duration (Keating et al. 2004). Our aim in this study is to replicate these results on continuous speech rather than controlled read speech - and for all vowels - but with the use of acoustic measurements only. The same spectral measurements (i.e. formants of vowels) used in our previous studies will be effected on these positions. We want to show that vowels at the boundaries of higher prosodic constituents in French are hyperarticulated when compared to lower prosodic constituents. So as to do this, we will evaluate the acoustic space used by all vowels for each of these prosodic levels (keeping initial and final positions apart). According to the hypotheses previously developed, we expect to observe an enlarging of the vocalic space when going upwards in the prosodic hierarchy, i.e. from syllable, to word, then accentual phrase and finally intonational phrase. So as to quantify these differences, we will calculate the dispersion (euclidean distance) from the acoustic centre as measured on the whole data (Gendrot & Adda-Decker 2007; Bradlow 1996); if the vowel is moving away significantly from the acoustic centre (F1: 450 Hz; F2: 1450 Hz), then it is considered as hyperarticulated. We bear in mind that this measure is inappropriate in some ways as it is related to the measurement of vowel centralization, which as detailed previously 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 favouring the phonemic identification (see Lindblom's theory of adaptive dispersion for an interpretation of this in the shape of vocalic systems).

In the following sections, we will summarize the procedures and methods used in our previous studies. In a first place, the corpus used and its automatic segmentation will be described. Secondly, the measurements will be detailed, including the precautions used to avoid detection errors.

2. Method

2.1. Corpus and automatic segmentation

The corpus corresponds to approximately 30 hours of speech (roughly 500 men and 300 women) mainly extracted from broadcast news of France Inter and France Info, recorded and transcribed orthographically at the French CTA/DGA (Galliano et al. 2005). It corresponds to radio and TV journalistic shows: articulation, without being emphasized, remains quite distinct, so that speech can be understood by a broad audience. Such speech cannot be described as fully spontaneous, but rather as prepared speech: only few hesitations,

repetitions, and word fragments are observed and syntactic structures often remain close to written language. Hypo- and hyperarticulation vowel phenomena, which we are interested in throughout this study, are undoubtedly less present here than in more conversational-style spontaneous speech.

The IRISA ('Institut de Recherche en Informatique et Systèmes Aléatoires') speech transcription system was used for corpus alignment. Orthographical transcriptions being known a priori the alignment system is used to locate phone boundaries, to choose among potential pronunciation alternatives (in particular 'liaisons' and schwas), and to discard silences, breath and other noise segments. Context independent phone models are used for alignment. Whereas context-dependent (e.g. triphones) acoustic models produce better transcription performances (i.e. a lower word error rates), context-independent acoustic models are more reliable for phone boundary location. For technical reasons, the segmentation resolution is limited to 10ms and the minimum duration of a segment is 30ms. Labelling thus produced is not a phonetic, but rather a phonological or phonemic labelling (corresponding in most cases to standard word pronunciations). Formant measures then allow to evaluate the variations observed in the acoustic realisation of phonemes.

2.2. Automatic formant extraction

Formant extraction makes use of the Burg algorithm implemented in the PRAAT software (Boersma & Weenink 2009). The detection of amplitude peaks is determined in a band lower than 5 KHz for male speakers and lower than 5.5 KHz for females. Measurements were taken respectively at 1/3, ½, 2/3 of the vowel segment, and then averaged to provide a single value. The interpretation of the extracted amplitude peaks as formants can raise controversy on a considerable number of segments: noises, too high fundamental frequency (voice of women and children), nasality... Two methodological safeguards are applied to prevent from errors: (i) nasal vowels were excluded from the study. (ii) amplitude peak values are filtered in order to reject erratic items, with respect to the acoustics of the vocal tract. For each vowel, upper and lower formant value limits are given for the first three formants (Gendrot & Adda-Decker 2005 for the values): if one of the formants lies outside the specified ranges, the corresponding vowel segment is rejected. Formant ranges were chosen in a broad way. A hundred visual checks for each vowel were carried out in order to reject as "errors" only severe formant detection problems and not the 'deviating' values which might be due to contextual assimilation effects, to prosody or to speaker's characteristics for example. After this filtering, approximately 20000 vowels out of the 500000 oral French vowels were rejected (4% of segments rejected). The major part of these rejections corresponds to segments of very short duration (12000 of the rejected segments have a duration smaller than 50ms). Listening to many of them shows that, at least for the shorter segments, the segmentation is not questionable. Other reasons may explain these rejections, in particular a partial or total devoicing of vowels, thus making formant detection more difficult (or even impossible) and potentially producing non-sense formant values. Similarly, when two formants of a vowel are close, especially in low frequency ranges, (which is the case for posterior closed vowels), the algorithm may detect only one formant instead of two, thus entailing a shift towards the higher order values. The /u/ is particularly prone to rejection, as all mentioned reasons may apply. More details and rejection rates for each vowel are described in Gendrot & Adda-Decker (2005).

2.3. Prosodic Categories selection

Four prosodic categories were evaluated in this study. From the lowest to the highest: syllables, words, accentual phrases (syntactic chunks), intonational phrases (pauses). We

describe the choices made to select these categories below, starting from word positions. Boundaries of each of these categories will be investigated, that is to say both initial and final phonemic positions.

Words were obtained from the manual transcription and segmented by the alignment system. Vowels described as word-initial were absolute word-initials: only taken from words starting with a vowel ('**a**rmée' [aʁme]: armée), not from initial syllables more simply ('p**a**rtir' [paʁtiʁ]: to leave). Indeed, as moving up the prosodic hierarchy, we found very little hyperarticulation for vowels in words' second phoneme position. However a more detailed analysis would be of interest in a future study as some effects were still observed between word second phonemes and word internal vowels. As for positions described as word-final, we considered both final and penultimate vowels ('part**i**e' [paʁ**ti**]: party; 'part**i**r' [paʁ**ti**ʁ]: to leave). Indeed both positions had an effect of the same magnitude along the prosodic hierarchy; this interesting point will be developed in the discussion.

Syllables were determined from phonemic segmentation. Syllabation rules inspired from Pallier (1994) and Adda-Decker et al. (2005) were used for the continuous string of phonemes, i.e. not taking word delimitations into account. For example the two words sequence 'bon ami' ('good friend') is segmented into 3 syllables : 'bo', 'na' and 'mi' unless they are separated by a pause. Pauses were considered as delimiters and syllables and, according to this principle, wouldn't contain pauses. If we were to follow the prosodic hierarchy in a strict manner, the lower level below word category is the syllable. However, it was difficult, if not impossible, to collect vowels that were syllable initial and still word internal. Vowels that are syllable initial but not word initial are very scarce as they can only be found in simple V-syllable kinds ('a**é**roport': airport, 2% in word internal positions in French). On the other hand, open syllables inside words are predominant (78 %) due to the open syllable status of French. We finally decided to take into account word syllables that were neither word final nor word initial, no matter which kind. Consequently, vowels considered here as word internal are predominantly (78 %) syllable final.

The third level analyzed in our prosodic hierarchy is what we identified as accentual phrases (see (1) for an example). They were obtained from a syntactic chunking based on automatically established grammatical categories, combined with several rules we describe here.

1. In the first place, each word is labelled with all available categories in the French dictionary of 'formes fléchies' ('Lefff': Clément et al. 2004), slightly modified for our needs. For example, we suppressed very rare word categories confounded with more frequent words (for example the adjective 'sûr' entailing false labellings).
2. Then the chunker from the 'Natural Language Toolkit' was used so as to generate two kinds of segments:
    a. Nouns, prepositions and verbs were gathered with their closest surroundings (clitics, determinants, prepositions, adjectives, etc.)
    b. All sequences of words not defined by the previous rule

In a final step, three merging rules were applied:

3. Merging of any segment ending on an auxiliary or a modal with the following segment.
4. Merging of any verbal segment with the following segment if the whole resulting chunk is less than seven syllables.
5. Merging of any other sequence of segments of less than seven syllables.

(1) combien de fois la justice française a-t-elle accepté de se remettre en question comme cela

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 Wioland (1985) - is a rough limitation to avoid overly long chunks. This constituent size limitation has also been mentioned by previous writers such as Grammont (1933). It may be supplanted in future studies by a duration threshold which would be more suited to be adapted to different speaking styles. 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. The accentual phrase, referred to as 'sense group' by Grammont ('groupe de sens') has a demarcative function as explained in the introduction. We could also have combined this morpho-syntactic 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 hyperarticulated in the first place. Hence, this procedure aims at evaluating accentual phrases on a syntactic (underlying) level rather than based upon their prosodic characteristics. We could also have used pitch rising but accentual phrases may also be realized with a lowering tone or a plateau as long as there is lengthening at their final boundary. That procedure would thus have excluded a non negligible part of the data.

The fourth and highest prosodic category we investigated is the intonational phrase. This was automatically detected with the help of pauses, the latter being taken into account when longer than 50ms only. Pauses were detected by the alignment system and this threshold was chosen so as to allow for possible detection errors. It is known in French that intonational phrases may be preceded/followed by pauses (Jun & Fougeron 2000). That's why we consider it as similar to intonational phrases. Intonational phrases may not be systematically preceded/followed by a pause, but when they are, they are surely considered as intonational phrases. However, a distinction has to be made here between rising and falling contours when a pause is following as it may be confounded with an utterance ending. As for the beginning of intonational phrases, we have no way of ensuring it cannot be an utterance (the highest prosodic level according to most authors).

We are aware that detecting automatically accentual and intonational phrases in such a way will lead some misidentification of categories. For example intonational phrases may be confounded with utterances as mentioned previously. Accentual phrases may also be taken for intonational phrases without pauses but may also be simply realized as word endings/beginnings. However the aim of this study was to determine four different theoretical categories, not based upon prosodic cues, and none of our examples can be in two categories at the same time. The large number of contexts is also believed to allow for a certain percentage of errors. Finally, an investigation of usual prosodic cues such as f0 and duration values on boundaries will help us check whether the detection of these categories has succeeded.

3. Results

3.1 Initial positions.

As was hypothesized, we can notice at a first glance on figure 2 that vowels occupy a much

larger vocalic space in the highest prosodic category (intonational phrase), and this gradually for every analyzed level. It seems that high vowels /i/ and /y/ don't undergo as much variation as the other vowels. The dispersion measurement detailed in the method section gives an indication of the statistical differences involved here.

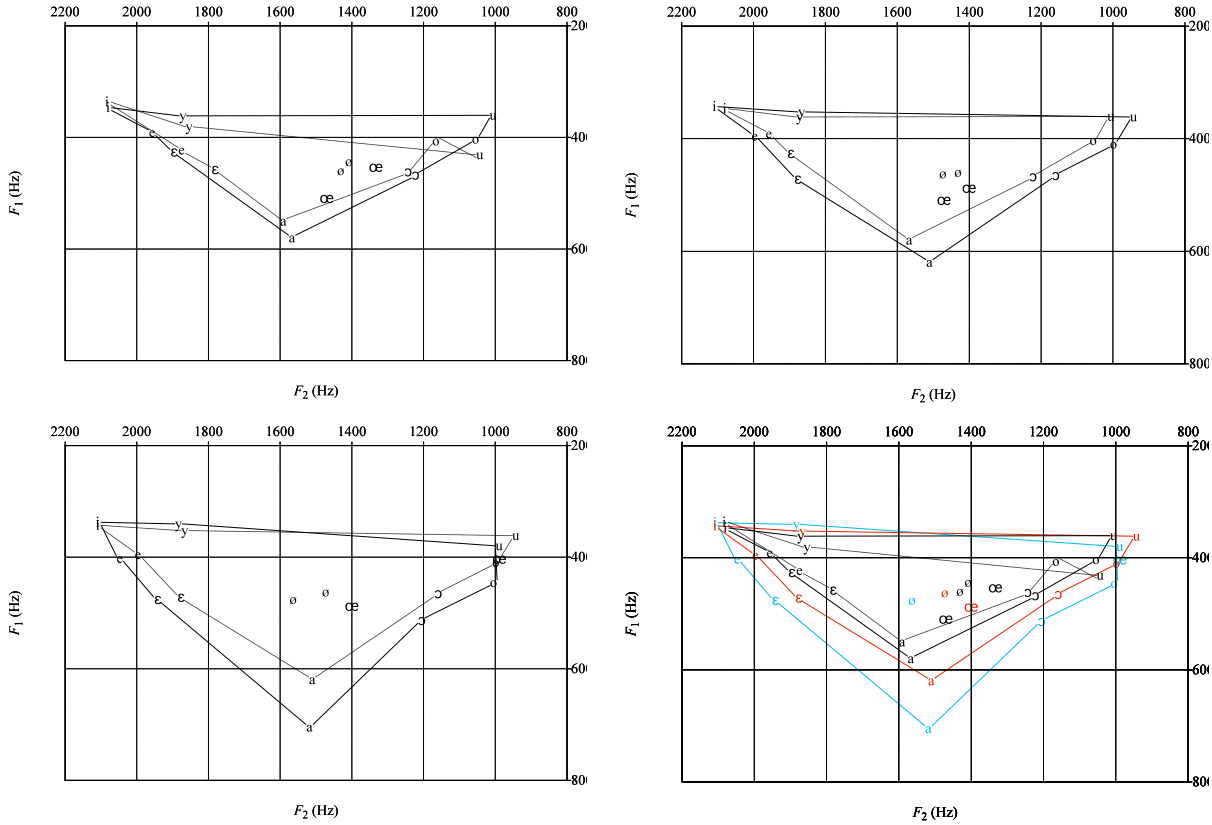


Figure 2.a.b.c.d:. from left to right and top to bottom. Comparison of all prosodic categories level by level: word internal, initial of word, accentual phrase and intonational phrase. The dashed triangle in the first three figures represents the lower level in each comparison pair. The bottom right figure summarizes all four positions.

Dispersion measurements results in figure 3 show that values generally increase with the level of the prosodic hierarchy, i.e. vowels move away from the acoustic centre of the acoustic space, allowing themselves to be more distinguishable from one another. Statistic tables won't be presented for sake of space but standard error bars shown on the figures give a hint about significant differences.

Only /e/ shows significant differences for all levels. Other vowels /ɛ/, /a/, /o/, /u/ show a significant dispersion according to the prosodic hierarchy for three levels out of four. For /ɔ/ and /i/, only two levels can be significantly distinguished, in both cases the syllable and word levels together versus accentual and intonational phrases. Only /y/ doesn't reveal a straightforward tendency according to all levels. We found in previous studies (Gendrot et al., 2008) that high front vowels /i/ and /y/ in French don't undergo as much variation in the F1/F2 plane as other vowels, but rather in the F3/F4 plane.

Third formant (F3) and fourth formant (F4) variations were also investigated so as to check whether these vowels undergo more variation in these dimensions. We observe that whereas F3 rises for all non rounded vowels, it lowers for rounded vowels /u/, /o/ and more

interestingly /y/. As for F4 values, they are lowering for /i/ when going up the prosodic hierarchy while they don't show any significant tendency for other vowels. These F3 and F4 movements allow /i/ and /y/ to be more focal by bringing together F3/F4 and F2/F3 respectively (Schwartz et al. 1997). In articulatory terms, /y/ is more rounded, while /i/ is more spread when going up the prosodic hierarchy, thus enhancing their articulatory characteristics.

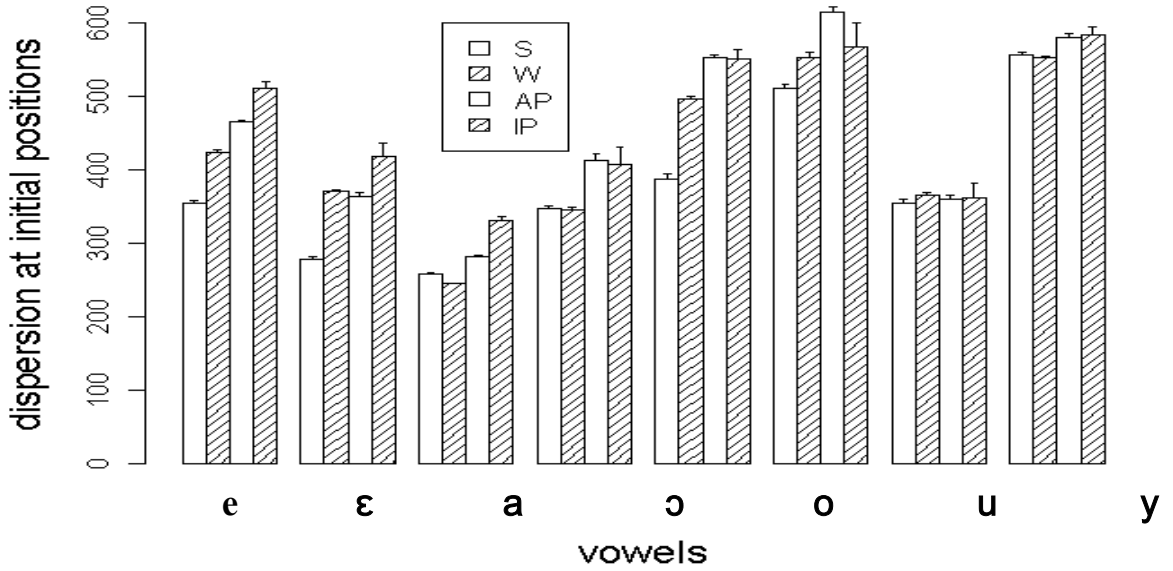


Figure 3: dispersion measurements at initial positions for all peripheral vowels according to the prosodic hierarchy (from S: syllable, to W: word, then AP: accentual phrase, and finally IP: intonational phrase.

A short glance at f0 and duration values in figure 4 reveals that both values are rising with the level of the prosodic hierarchy which confirms that phonemes are hyperarticulated not only in spectral but also in prosodic terms. Vowel lengthening favours hyperarticulation and a f0 raise generally increases all formant values. These prosodic measurements also confirm the fiability of our automatic detection of selected prosodic categories. The relation between f0, duration and formant measurements will be developed in the discussion.

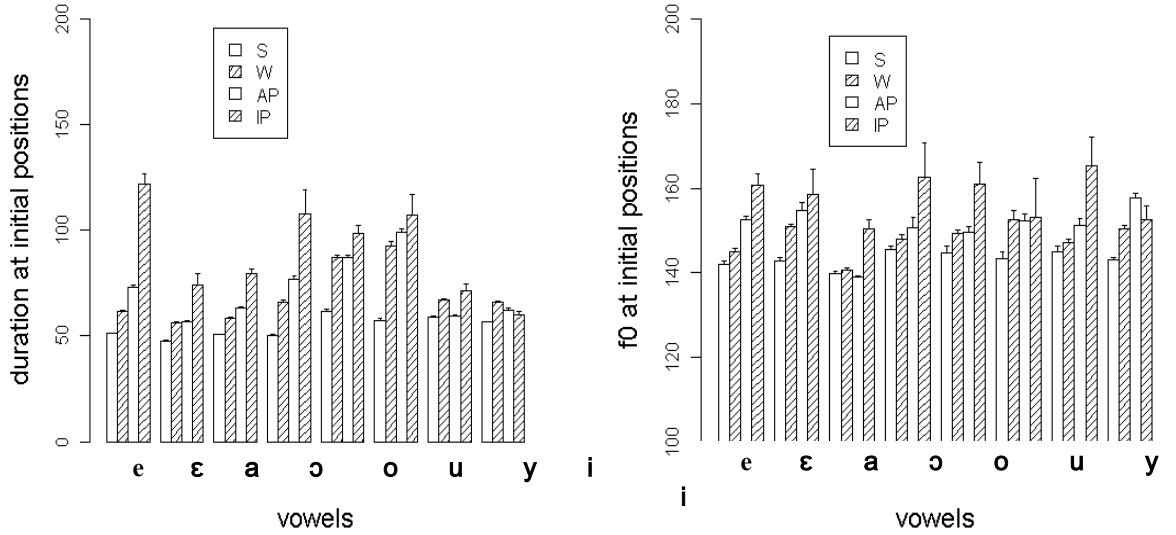
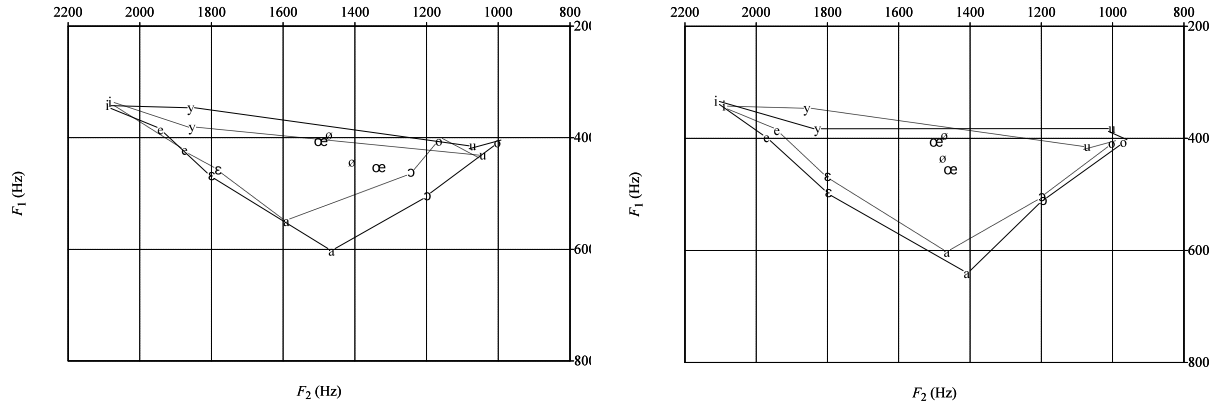


Figure 4a.b: duration (left) and f0 (right) values at initial positions according to the prosodic hierarchy (from S: syllable, to W: word, then AP: accentual phrase, and finally IP: intonational phrase).

3.2. Final positions

As observed for initial positions, we can see at first glance in figure 5 that vowels occupy a much larger vocalic space in the highest prosodic category (intonational phrase), and this gradually for every analyzed level. It seems that /i/ doesn't undergo as much variation compared to the other vowels, while /y/ is rather characterized by untypical variations.



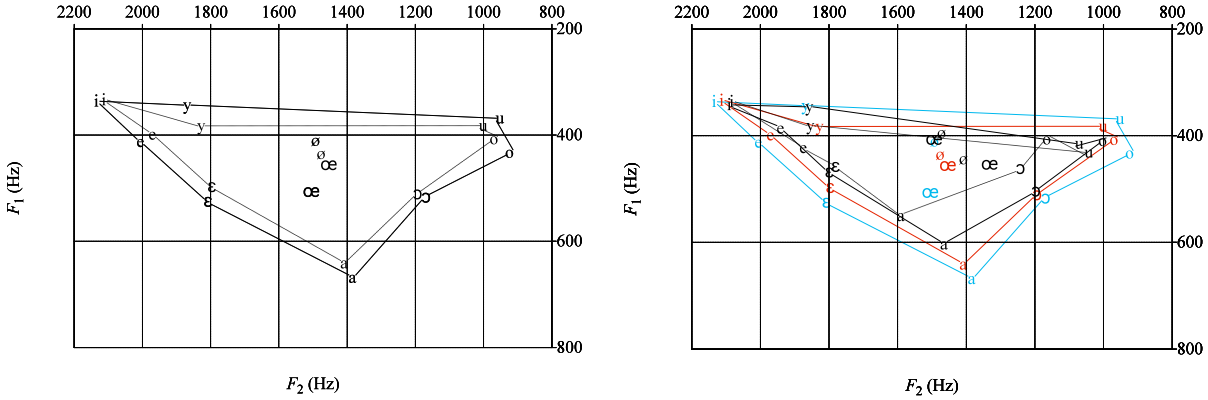


Figure 5.a.b.c.d.: from left to right and top to bottom. Comparison of all prosodic categories level by level: word internal, final of word, accentual phrase and intonational phrase. The dashed triangle in the first three figures represents the lower level in each case. The bottom right figure summarizes all four positions.

Once again, the dispersion measurements in figure 6 give an indication of the statistical differences involved here: the amplitude of variations is larger than for initial positions. Vowels /e/, /a/, /o/ and /i/ show significant differences for all levels, although with less amplitude for /i/. Other vowels /ɛ/, /ɔ/, and /u/ show a significant dispersion according to the prosodic hierarchy for three levels out of four. As for /y/ finally, variations don't quite go in the expected direction with the accentual phrase level having the smallest values. The highest level (intonational phrase) still has higher values than syllable and word levels, which follows the general tendency.

F3 and F4 variations were investigated as well: F3 lowers for rounded vowels /u/, /o/ and /y/ but for the latter, only two levels were distinguished (syllable and word level together versus accentual and intonational phrases). As for F4 values, they are lowering for /i/ when going up the prosodic hierarchy for final positions, in the same way it did for initial positions. As noticed for initial positions, these F3 and F4 movements allow /i/ and /y/ to be more focal, although less so for /y/.

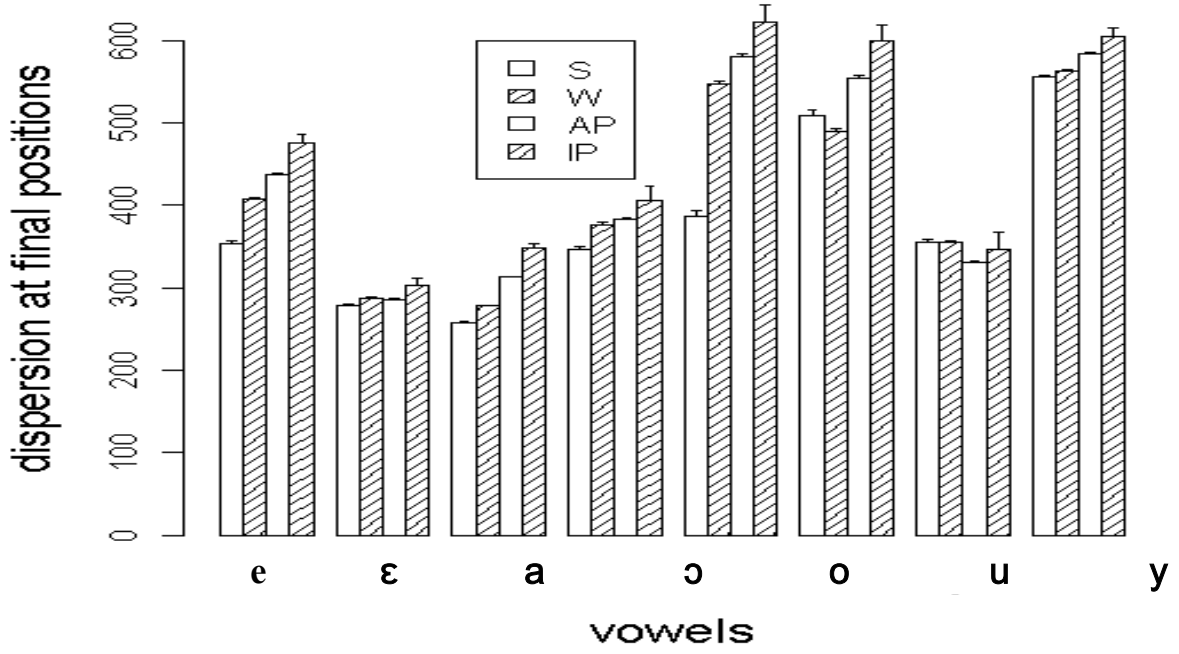


Figure 6: dispersion measurements at final positions for all peripheral vowels according to the prosodic hierarchy (from S: syllable, to W: word, then AP: accentual phrase, and finally IP: intonational phrase.

A short glance at f0 and duration values reveals that both values are rising with the levels of the prosodic hierarchy. As noticed for dispersion measurements, the amplitude of this variation is much larger than for initial positions. This corresponds to predictions of the French prosodic hierarchy and in turn, confirms that the four analyzed prosodic categories were correctly detected.

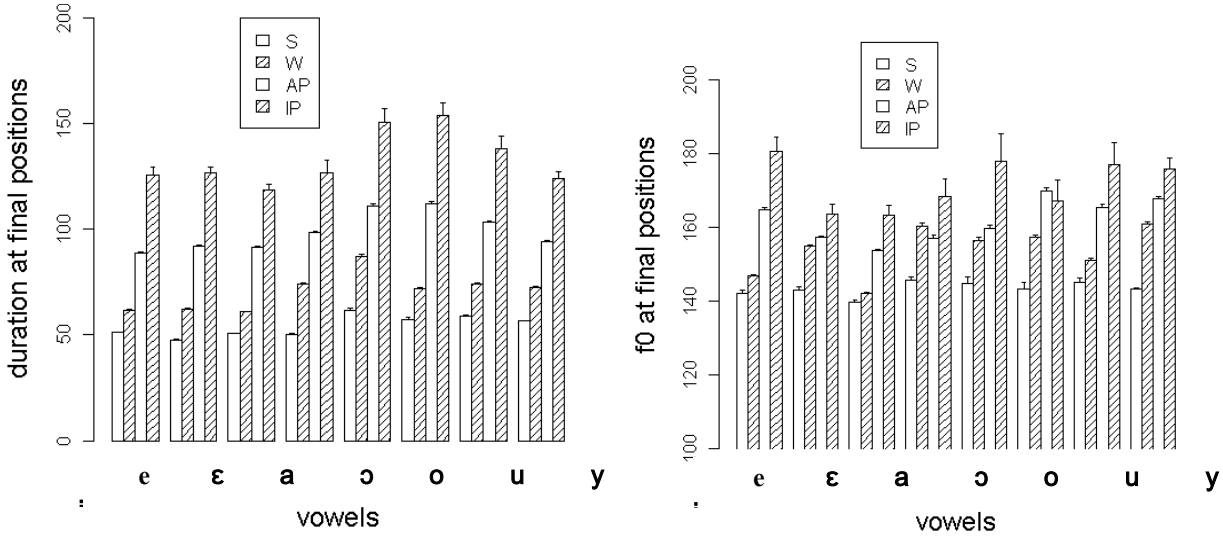


Figure 7a.b.: duration (left) and f0 (right) values at final positions according to the prosodic hierarchy (from S: syllable, to W: word, then AP: accentual phrase, and finally IP: intonational phrase).

4. Discussion

4.1. General results: significant differences and possible improvements

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 structure of French, the more hyperarticulated it is. However, as noticed by previous studies (Fougeron 2001; Tabain 2002; Tabain & Perrier 2005; Cho 2005), 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, /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 favouring their focal characteristics (Schwartz et al.,1997).

Other vowels such as /ɔ/, /o/, and /u/ revealed variations that could be unexpected in some contexts, sometimes larger or smaller than for other vowels. These results may be partly explained by the scarce number of these vowels at specific prosodic levels for phonotactic reasons. For example, we noticed that /ɔ/ has a non typical variation in (strictly) final positions with particularly high F2 values; /ɔ/ is typically found in closed syllables (‘or’, ‘corps’, ‘**cor**pus’, etc.) according to French Phonology. However these odd variations are partly covered as we considered both final and penultimate vowels. No F3 lowering was observed for /ɔ/, neither for initial nor final positions. This is rather surprising since /ɔ/ is usually mentioned as a rounded vowel. But French /ɔ/ has already been mentioned to be closer to /œ/ in some varieties of French (Boula de Mareuil et al. 2008). As for /u/ and /o/, they are the least represented peripheral vowels (around 2% each) which may account for their somewhat less regular variations along the prosodic hierarchy. Finally, we also have to mention that an f0 increase, as was noticed in figures 4b and 7b results in an increase of all F1 values. This fact may explain why we do not systematically get the same results as in figure 1, i.e. hyperarticulated close vowels having a lower and lower F1.

4.2. Information content and span of the hyperarticulation

No distinctions between grammatical categories (‘parts of speech’) were considered in this study. It has to be mentioned that distribution of grammatical categories is not even, depending on the analyzed prosodic positions. Indeed, beginning words of accentual of intonational phrases are usually prepositions, conjunctions or determinants, while for the two other prosodic categories (syllable and word), nouns and adjectives are the most represented. It could be expected that grammatical words are 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- & 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 accentual and intonational phrases might be more hyperarticulated towards the end and not at the beginning. However, in previous unpublished analyses, we noticed that vowels formants were not subject to much change when comparing grammatical with lexical words. We decided not to filter out these constituents with the hypothesis that in continuous speech, if the speaker does decide to

signal boundaries to the listener, then he will do it whatever the grammatical category.

Indeed, it was measured that both initial and positions are characterized by hyperarticulation 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 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 Byrd & Saltzman (2003), 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.

One interesting result was the fact that for final positions, strictly final but also penultimate vowels were considered for analysis, while for initial positions, only strictly initial vowels were taken into account. As suggested by Fougeron (2001) for French or Byrd et al. (2006) for American English, the span seems to more important on final than on initial positions. In fact some variations could be observed between syllable and word initial position but they were overall less important than strictly initial positions. The amplitude of variations for hyperarticulation was noticed to be larger in final positions. Duration and f0 variations as well are much larger for final than for initial positions. It is also possible that for grammatical words, the span of hyperarticulation might not be as spread as for lexical words and this should be soon investigated.

4.3. Hyperarticulation and 'prosodic' characteristics

So as to make sure to retain intonational phrases and filter out utterance endings, we measured f0 contours on analyzed vowels and kept only rising contours for intonational phrase category. The filtering of rising f0 contours in final positions had in fact few consequences. A similar hyperarticulation pattern could be found for lowering f0 contours. Phonemes in final positions of intonational phrases with a lowering f0 contour (supposedly utterance endings) were even slightly hypoarticulated compared to intonational phrases (with a rising f0 contour). This tends to show that - as suggested by Fougeron (2001) for French - there are few articulatory differences at boundaries between the intonational phrase level analyzed here and an utterance level.

We noticed that f0 and duration values increased in parallel to dispersion measurements, along the prosodic hierarchy. It seemed in the first place that these three parameters were linked since f0 and duration are also known to mark boundaries. Correlations were effected, but as noticed by Keating et al. (2004), they turned out to be weak (at initial positions: $r=0.21$ between dispersion and duration and $r=0.26$ between dispersion and f0; at final positions: $r=0.18$ between dispersion and duration and $r=0.27$ between dispersion and f0). If f0 and duration that are usually considered as boundary markers are not strongly correlated with our dispersion measurements, it suggests that there might be compensations between spectral and prosodic variations used to by speakers in order to signal boundaries. Therefore it also suggests that speaker variability should be investigated in a near future.

Some modifications are currently being added so as to improve the chunking results. First of all, we are working on some possible improvements to the syntagmatic chunking so as to avoid some irregular accentual phrases (a small proportion of these were noticed when checking manually the chunked portions of speech). Changing the number of 7 syllables (Wioland 1985) to an amount of time dependent on speech rate adopted by the speaker should be an interesting point in the future and finally, other kinds of chunking based on statistical training might be also used. Thanks to these, more significant differences may be found between all prosodic levels. We also started studying consonants in the same prosodic positions. Indeed, word initial vowels in French are scarce due to the CV predominance, and

significant results could be found for initial consonants.

5. Conclusion

In this study we showed that the level of prosodic constituent in French influences the acoustic realization of vowels at constituent boundaries. We observe a general prosodic hierarchy (from syllable to word, then accentual phrase and finally intonational phrase) based on formant measurements, showing that the higher a vowel is in the prosodic structure of French, the more hyperarticulated it is. However, as noticed by these previous studies, all levels cannot be statistically distinguished from one another in a systematic manner. Phonotactics and the lack of occurrences of a few phonemes in restricted contexts were possibly involved in some non significant or unexpected results. Possible improvements will be brought by other chunking methods.

These hyperarticulation results follow the increasing f0 and duration values measured on these phonemes although correlations were noticed to be weak. Both initial and final positions were subject to hyperarticulation for higher prosodic levels but hyperarticulation seems to have a longer 'span' for final positions. Finally, grammatical category of the word involved doesn't seem prevalent since 'determinants' and 'conjunctions' that are predominant at the beginning of accentual and intonational phrases are still hyperarticulated compared to 'nouns' or 'verbs' in the lower prosodic level.

References

Adda-Decker M., P. Boula de Mareüil, G. Adda & L. Lamel (2005). Investigating syllabic structures and their variation in spontaneous French. *Speech Communication 46*, pp. 119-139.

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

Boula de Mareuil, P., B. Vieru-Dimulescu, C. Woehrling & M. Adda-Decker. Accents étrangers et régionaux en français. *Traitement Automatique des Langues 49:3*, pp.135-163.

Bradlow, A.R., G.M. Torreta & D.B. Pisoni (1996). Intelligibility of normal speech: global and finegrained acoustic-phonetic characteristics. *Speech Communication 20*, pp. 255-272.

Bürki, A., C. Fougeron, C. Gendrot & U. Frauenfelder (to appear). Phonetic reduction versus phonological deletion: Methodological questions regarding the study of schwa/zero alternation in French. *Journal of Phonetics*.

Byrd, D. & E. Saltzman (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics 31:2*, pp. 149-180.

Byrd, D., J. Krivokapic & S. Lee (2006). How far, how long: On the temporal scope of phrase boundary effects. *Journal of the Acoustical Society of America 120*, pp. 1589-1599.

Cho T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in English. *Journal of Acoustical Society of America 11:6*, pp. 3867-3878.

Clément, L., B. Sagot & B. Lang (2004). Morphology based automatic acquisition of large-coverage lexica. In *proceedings of LREC'04*, Lisboa, Portugal, pp. 1841–1844.

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

Fougeron, C., C. Gendrot & A. Bürki (2007). On the acoustic characteristics of French schwa. In *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Saarbrücken, Germany. pp. 941-944.

Galliano, S., E. Geoffrois, D. Mostefa, K. Choukri, J.-F. Bonastre & G. Gravier (2005). ESTER Phase II Evaluation Campaign for the Rich Transcription of French Broadcast News. In *Proceedings of Interspeech*, Lisboa, Portugal, pp. 1149-1152.

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

Gendrot, C. & M. Adda-Decker (2007). Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages, In *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany. pp. 1417-1420.

Gendrot, C., M. Adda-Decker & J. Vaissière (2008). Les voyelles /i/ et /y/ du français : focalisation et variations formantiques. In *proceedings of 26èmes Journée d'Etude de la Parole*, Avignon, France, pp. 205-208

Gendrot, C. & M. Adda-Decker (to appear). Impact of duration and consonantal context on formant values of oral vowels: An automatic analysis of large broadcast news corpora in French. In M. Embarki & C. Dodane (Eds.), *Patterns of speech coarticulation: Between physics and metaphysics*. Cambridge: CUP.

Grammont M. (1933). *Traité de phonétique*, Paris: Librairie Delagrave.

Jun S.-A. & C. Fougeron (2000). A Phonological model of French intonation. In A. Botinis (ed.) *Intonation: Analysis, Modeling and Technology*. Dordrecht : Kluwer Academic Publishers. pp.209-242.

Keating P., T. Cho, C. Fougeron & C.-S. Hsu (2004). Domain-initial articulatory strengthening in four languages. In J.K. Local, R. Ogden and R.A.M. Temple (eds.) Papers in Laboratory Phonology VI: Phonetic interpretation *Cambridge: CUP, pp. 145-163.*

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

Lindblom B., 1990, Explaining phonetic variation : a sketch of the H & H theory, in Speech production and speech modelling, W. Hardcastle et A. Marchal, Dordrecht, Kluwer, pp. 403-440.

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

Pallier, C. (1994). Rôle de la syllabe dans la perception de la parole : études attentionnelles. *PhD thesis*, EHESS, Paris.

Schwartz J. L., L. J. Boë, N. Vallée & C. Abry (1997). The dispersion-focalization theory of vowel systems. *Journal of Phonetics 25*, pp. 255-286.

Stevens, K.N. (1997) Articulatory-acoustic-auditory relationships. In W.J. Hardcastle and J. Laver (eds.), *The Handbook of Phonetic Sciences* , Blackwell: Oxford, pp. 462-506.

Tabain M. (2003a). Effects of prosodic boundary on /aC/ sequences: acoustic results. *Journal of the Acoustical Society of America 113*, pp. 516-531.

Tabain M. (2003b). Effects of prosodic boundary on /aC/ sequences: articulatory results, *Journal of the Acoustical Society of America 113*, pp. 2834-2849.

Tabain M. & Perrier P. (2005). Articulation and acoustics of /i/ at prosodic boundaries in French, *Journal of Phonetics 33*, pp. 77-100.

Vaissière, J. (2007). Area Functions and Articulatory Modeling as a Tool for Investigating the Articulatory, Acoustic, and Perceptual Properties of Sounds Across Languages. In: Sole, M-J., Beddor, P.S., Ohala, M. (eds), *Experimental Approaches to Phonology*, Oxford University Press, pp. 55-72.

Wioland, F. (1985). Les structures syllabiques du français : Fréquence et distribution des phonèmes consonantiques, Contraintes idiomatiques dans les séquences consonantiques. *Slatkine-Champion*, Genève-Paris, 1985.

Wright, R. (2003). Factors of lexical competition in vowel articulation, in J. Local, R. Ogden & R. Temple (eds), Papers in Laboratory Phonology VI*, Cambridge: CUP, pp. 75-87.*