

Hiérarchie prosodique et réalisation spectrale des voyelles

Cédric Gendrot et Kim Gerdes

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

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

ABSTRACT

The link between the duration of vowels and their spectral realization has been validated for a long time by Lindblom (1963), i.e. the longer the vowels the more strengthened they are. Similarly, the relation between prosodic constituents of different levels (the prosodic hierarchy) and the duration of phonemes close to their boundaries has been demonstrated (in French, Fougeron 2001; Tabain 2003, 2005). In this study we aim at reproducing these results on non-controlled continuous speech. Procedures that try and detect automatically known prosodic categories are detailed. We show that the higher a vowel is in the prosodic structure of French, the more strengthened it is.

Keywords: prosody, formants, vowels, strengthening.

1. INTRODUCTION

Cette étude s'inscrit dans le cadre d'un projet de plus grande envergure s'attachant à décrire la variabilité des phonèmes en français. Grâce à l'utilisation de corpus de très grande taille segmentés automatiquement, nous avons pu étudier un nombre important de contextes afin de mieux quantifier leurs influences et leurs interactions. Nous avons observé lors d'études précédentes (Gendrot & Adda-Decker 2005) que la réalisation spectrale des voyelles était grandement influencée par leur durée : les voyelles les plus longues sont considérablement renforcées par rapport aux voyelles les plus courtes. Si l'on considère un espace acoustique formé par les 2 premiers formants de chaque voyelle, plus les voyelles sont longues et plus l'espace vocalique occupé par l'ensemble de ces voyelles est important (figure 1).

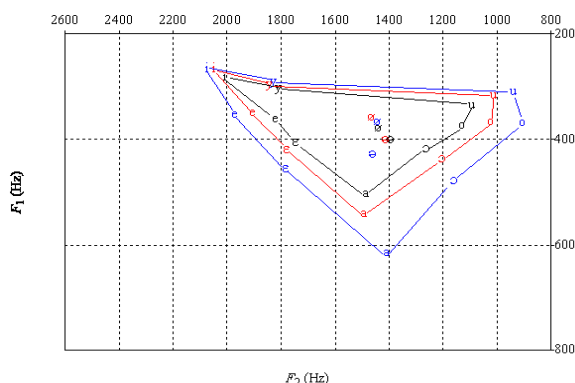


Figure 1: Valeurs moyennes de F1/F2 des voyelles orales du français en fonction de leur durée. Par ordre ascendant et de l'intérieur vers l'extérieur : noir [30-50ms], rouge [60-80], noir [90-110]).

Cette plus grande occupation de l'espace acoustique permet notamment aux voyelles de mieux se distinguer les unes des autres afin d'être plus facilement identifiables.

Les raisons de ces variations de durée sont multiples et des facteurs tels que le style et le débit du locuteur ont une grande influence. Des facteurs plus linguistiques tels que le contexte segmental, la position de la voyelle dans la syllabe, le mot, le syntagme ou l'énoncé ont également une forte influence. Les quatre unités précédemment mentionnées sont traitées comme des constituants prosodiques puisque représentés par des variations caractéristiques de durée et de f_0 dans un but démarcatif (ils signalent des occurrences de frontières). Ces constituants prosodiques sont généralement considérés comme étant organisés au sein d'une hiérarchie prosodique : chaque constituant est imbriqué dans un constituant de plus haut niveau (Nespor et Vogel, 1986). Dans ce courant de recherche, il a été montré que plus le niveau du constituant prosodique est élevé dans la hiérarchie, plus les phonèmes aux frontières de ces constituants sont longs (pour le français, Fougeron 2001; Tabain 2003, 2005). Ces positions ont plus rarement été étudiées d'un point de vue spectral (et/ou articulatoire). Pour le français, Fougeron (2001) à partir de l'électropalatographie sur les consonnes (EPG) et Tabain (2003, 2005) grâce à l'articulographie sur les voyelles (EMMA) ont travaillé respectivement sur les positions initiales et finales. Leurs résultats montrent qu'un phonème en position strictement initiale ou finale de constituant sera influencé par sa position dans la hiérarchie prosodique. Plus le niveau sera élevé, plus le phonème sera renforcé.

Nous tentons ici de répliquer ces études sur de la parole continue à partir de mesures acoustiques. Les productions des locuteurs ne sont ainsi plus contrôlées ; si les données sont considérablement plus naturelles que celles de corpus lus, il est impossible de contrôler la catégorie prosodique produite par les locuteurs. Il est donc nécessaire de les détecter de façon indirecte. Notre objectif est de montrer que le niveau des constituants dans la hiérarchie prosodique a un effet sur la réalisation spectrale des voyelles. Quatre constituants prosodiques unanimement reconnus ont été choisis : la syllabe, le mot, le syntagme accentuel et le syntagme intonatif. Nous expliquerons quels choix ont été faits pour approcher ces constituants de façon automatique, et ceux-ci seront analysés à leurs frontières (en position initiale et finale).

2. MÉTHODE

2.1. Corpus et analyses

Les données de parole utilisées correspondent approximativement à 30 heures de parole radiophonique (environ 500 hommes et 300 femmes extraites du corpus ESTER (Galliano et al., 2005) et à partir de la segmentation effectuée par l'IRISA. Les procédures de transcription et de segmentation sont détaillées dans Bürki et al. (2008)

Des mesures des trois premiers formants (F1, F2 et F3) ont été effectuées automatiquement à l'aide de l'algorithme Burg implémenté dans Praat (Boersma et Weenink 2009). Des fourchettes de valeurs ont été établies afin de filtrer les valeurs fantaisistes (approx. 4% au final) dues à des erreurs de détection automatique. Plus de détails sur les mesures de formants et les précautions prises sont fournis dans Gendrot et Adda-Decker (2005). Comme mentionné dans l'introduction, les voyelles renforcées forment un espace acoustique beaucoup plus large, elles sont donc plus éloignées du centre acoustique. Nous émettons l'hypothèse que les voyelles aux frontières des constituants prosodiques de haut niveau (syntagme accentuel, puis syntagme intonatif) seront renforcées comparées aux voyelles en frontière de constituants prosodiques de plus bas niveau (le mot, puis la syllabe). Afin de quantifier cette variation, nous avons effectué par simple distance euclidienne une mesure d'éloignement du centre acoustique déterminé sur l'ensemble de nos données (F1 : 450Hz, F2 : 1540Hz) (Gendrot & Adda-Decker 2007; Bradlow 1996).

2.2. Sélection des catégories prosodiques

Quatre catégories ont été analysées dans cette étude, ces catégories tentent d'approcher les catégories prosodiques les plus acceptées : de la plus basse à la plus élevée, la syllabe, le mot, le syntagme accentuel et le syntagme intonatif. Nous décrivons ci-dessous les choix effectués pour détecter automatiquement chaque catégorie, en commençant par le mot. Les frontières de chaque catégorie seront analysées, c'est-à-dire aux positions initiales et finales. Les voyelles décrites comme initiales sont strictement initiales ("armée"), et non pas plus simplement dans la syllabe initiale. En effet, des résultats préliminaires ont montré que l'effet dû à la hiérarchie prosodique était essentiellement concentré sur la voyelle strictement initiale seulement. A l'inverse, pour les positions finales, les voyelles à la fois finales et en pénultième position ont été prises en compte car un effet semblable a été remarqué pour ces 2 positions.

Les frontières de mots ont été obtenues sur la base de la transcription manuelle, puis de l'alignement effectué automatiquement.

Les syllabes ont été déterminées à partir de la segmentation phonémique. Des règles de syllabation inspirées de Pallier (1994) ont été utilisées depuis le flux continu de phonèmes, i.e. les frontières de mots n'interviennent pas dans la syllabation. Par exemple, la séquence de mots "bon ami" est segmentée en 3 syllabes "bo", "na" et "mi" sauf s'ils sont séparés par une pause. Par manque d'occurrences, nous n'avons pu déterminer des positions initiales et finales pour le niveau syllabique ; par exemple les voyelles initiales de syllabe mais internes de mots sont très rares ("aéroport") et toutes les voyelles en syllabe interne de mot ont donc été analysées pour la catégorie "syllabe".

Le troisième niveau analysé ici est le syntagme accentuel (voir exemple en (1)). Un chunking syntaxique a été effectué sur la base d'un étiquetage grammatical automatique (Leff : Clément et al. 2004) combiné à quelques règles de

regroupement mises en place grâce au chunker du Natural Language Toolkit (http://nltk.org/index.php/Main_Page) :

- o Les segments nominaux, prépositionnels et verbaux sont regroupés avec leur entourage le plus proche (clitiques, déterminants, prépositions, adjectifs, etc.)

- o Puis, 3 règles de combinaisons sont appliquées :

- 1- Combinaison de tout segment terminant sur un auxiliaire ou modal avec le segment suivant.
- 2- Combinaison de tout segment verbal avec le segment suivant si la combinaison fait moins de 7 syllabes.
- 3- Combinaison de tout autre suite de segments qui fait moins de 7 syllabes.

Les segments découpés par cet algorithme peuvent avoir plus de 7 syllabes, si les règles précédentes le permettent, par exemple « avec qui j'ai pu m'entretenir », qui forme un groupe très naturel et difficilement à découper. La règle des 7 syllabes (Vaissière, 1971) pourra ultérieurement être supplantée par un calcul plus respectueux du débit de parole.

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

Par ce "chunking", nous tentons de nous approcher de la réalisation du syntagme accentuel. Nous sommes conscients que tous ces syntagmes ne seront pas "accentués", c'est-à-dire qu'ils ne seront pas tous caractérisés par un allongement final et/ou un contour mélodique montant. Cependant, utiliser des informations prosodiques pour s'en assurer aurait introduit un caractère circulaire dans notre étude puisque les voyelles les plus longues sont elles-mêmes renforcées. Notre méthode vise donc à évaluer la réalisation spectrale de syntagmes accentuels à un niveau syntaxique (sous-jacent), plutôt qu'en considérant des syntagmes accentuels d'après leurs caractéristiques prosodiques.

La quatrième catégorie analysée vise à s'approcher du syntagme intonatif et est détectée automatiquement sur la base des pauses (supérieures à 50ms) indiquées par l'alignement. En effet, il a été montré que la présence de pauses est un facteur important pour signaler la réalisation d'un syntagme intonatif (Jun & Fougeron 2000). Une détection de la forme du contour final de f0 (montant/descendant) a été effectuée dans le but de ne prendre en compte que les syntagmes ayant un contour montant. Cette méthode permet ainsi de les distinguer d'une position finale d'énoncé ayant un contour descendant. Aucune précaution de cet ordre n'a pu être effectuée pour les positions initiales. Nous sommes encore une fois conscients que la détection automatique de cette catégorie peut engendrer un certain nombre de détections erronées. Une explication semblable à celle évoquée pour les syntagmes accentuels sera proposée ici : l'objectif de cette étude est d'obtenir quatre catégories prosodiques, aussi proches que possible de celles mentionnées dans la littérature, sans se baser sur des caractéristiques prosodiques d'allongement. Une analyse supplémentaire (non détaillée ici) a montré des valeurs de durée et de f0 augmentant progressivement pour nos 4 catégories de la hiérarchie prosodique, confirmant ainsi la fiabilité de ces catégorisations.

3. RÉSULTATS

3.1. Positions initiales

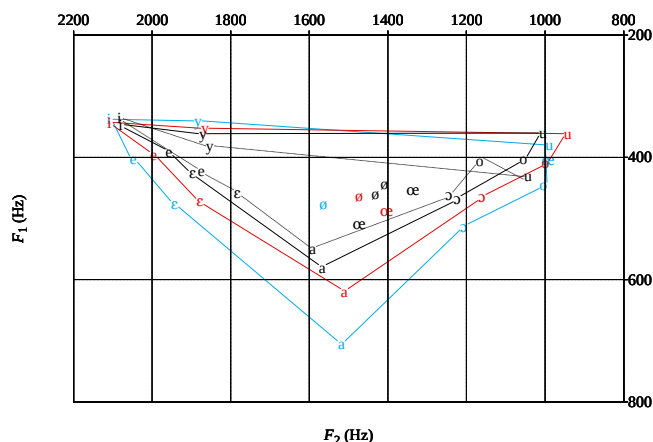


Figure 2 : Comparaison des 4 catégories prosodiques en position initiale. De l'intérieur vers l'extérieur : syllabe, mot, syntagme accentuel et syntagme intonatif.

Comme le montre la figure 2, les voyelles occupent un espace acoustique de plus en plus important en remontant la hiérarchie prosodique. Les mesures de dispersion présentées par la figure 3 révèlent que les valeurs augmentent globalement avec le niveau de la hiérarchie prosodique, c'est-à-dire que les voyelles s'éloignent du centre de l'espace vocalique, étant ainsi plus distinctes les unes des autres perceptivement.

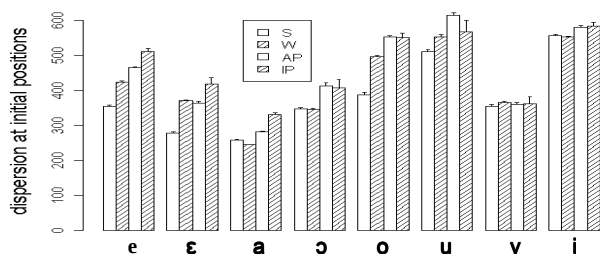


Figure 3 : Dispersion en position initiale en fonction de la hiérarchie prosodique (S: syllabe, W: mot, AP: syntagme accentuel, et IP: syntagme intonatif).

Des ANOVAS à 1 facteur (catégorie) ont été effectuées pour déterminer les différences significatives pour chaque voyelle. Seul /e/ affiche des différences significatives pour chaque niveau. Les voyelles /ε/, /a/, /o/ et /u/ montrent une dispersion significative entre trois niveaux. Pour /ɔ/ et /i/, seuls deux niveaux peuvent être distingués significativement, dans les deux cas, les niveaux syllabe/mot vs. syntagme accentuel/intonatif. Seul /y/ ne révèle aucune tendance en fonction de la hiérarchie prosodique.

3.2. Positions finales

Comme observé en 3.1., les voyelles occupent un espace acoustique de plus en plus important en remontant la hiérarchie prosodique pour les positions finales.

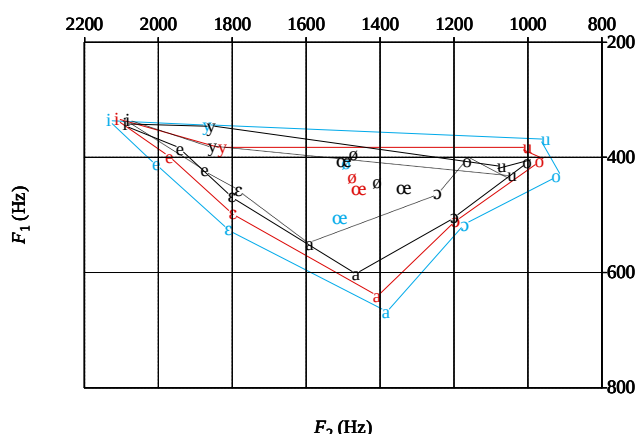


Figure 4 : Comparaison des 4 catégories prosodiques en position finale. De l'intérieur vers l'extérieur : syllabe, mot, syntagme accentuel et syntagme intonatif.

Les mesures de dispersion présentées par la figure 5 révèlent également des valeurs qui augmentent globalement avec le niveau de la hiérarchie prosodique. Les voyelles /e/, /a/, /o/ and /i/ montrent des différences significatives pour tous les niveaux, malgré une amplitude moins importante pour /i/. Les voyelles /ε/, /ɔ/ et /u/ révèlent une dispersion significativement plus importante pour trois niveaux. Pour /y/ encore, les variations sont soit inattendues, soit non significatives.

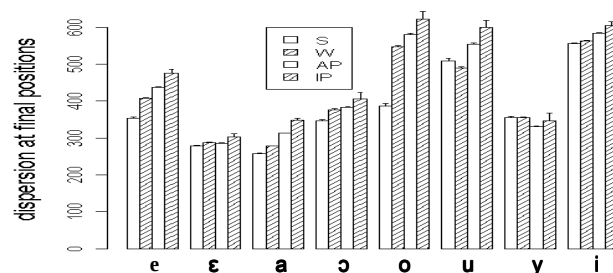


Figure 5 : Dispersion en position finale en fonction de la hiérarchie prosodique (S: syllabe, W: mot, AP: syntagme accentuel, et IP: syntagme intonatif).

4. DISCUSSION ET CONCLUSION

Comme nous en avons fait l'hypothèse, nous observons une hiérarchie prosodique sur la base de mesures spectrales à partir des catégories prosodiques que nous avons déterminées. A la différence des travaux effectués sur de la parole lue, les catégories sont déterminées à postériori sur le continuum de parole, et cette procédure nécessitera à l'avenir une réflexion méthodologique : combien de catégories significativement différentes pourraient ainsi se distinguer statistiquement ? Comme signalé par les études précédentes dans ce domaine (Fougeron 2001; Tabain 2003, pour le français), tous les niveaux ne peuvent pas être distingués de façon systématique. Par exemple nous avons noté que /i/ et surtout /y/ étaient moins variables sur l'axe F1/F2. Gendrot et al. (2008) avaient précédemment observé ce même résultat en fonction des variations de durée, en indiquant que les variations spectrales étaient plus importantes sur un plan F3/F4. Une analyse des valeurs de F3 et F4 a pu montrer que /y/ est caractérisé par un

rapprochement de F2 et F3 à mesure qu'il s'élève dans la hiérarchie prosodique. /i/ est quant à lui caractérisé par un rapprochement de F3 et F4. Ces rapprochements sont ainsi plus pertinents que les variations chaotiques observées sur F1 et F2 et servent à accroître les caractéristiques focales de ces voyelles (Schwartz et al., 1997).

D'autres voyelles telles que /ɔ/, /o/ et /u/ ont révélé des variations inattendues, parfois non significatives, parfois dans des directions ne correspondant pas aux hypothèses. Ces résultats peuvent être en partie expliqués par une quantité assez restreinte de ces voyelles dans ces positions phonotactiques. Par exemple, nous avons remarqué que /ɔ/ a des variations atypiques dans les positions finales strictes avec des valeurs de F2 particulièrement élevées ; or /ɔ/ apparaît en syllabe fermée ('or', 'corps', 'corpus', etc.) selon les règles phonologiques du français. Finalement, nous pouvons également mentionner qu'une augmentation de f0 comme celle signalée en 2.2 aboutit à une augmentation des valeurs de F1, ce qui pourrait également expliquer pourquoi les voyelles fermées varient peu sur F1 en fonction des positions prosodiques, contrairement à ce que nous avons observé sur la figure 1, i.e. les voyelles fermées renforcées ont traditionnellement des valeurs de F1 de plus en plus basses.

Il convient de se rappeler que les positions initiales chaque de catégorie prosodique correspondent également à la fin d'une catégorie de même niveau. Comme proposé par Byrd &

Saltzman (2006), ces frontières sont des moments de ralentissement articulatoire ('pi-gesture') qui favorisent le renforcement. Il est donc normal d'observer les mêmes résultats et phénomènes de renforcement en position initiale et finale de tous nos niveaux prosodiques, et ce quelque soit la catégorie grammaticale impliquée (en effet, les positions initiales de syntagmes accentuels et intonatifs sont majoritairement occupées par des mots grammaticaux, généralement considérés comme hypoarticulés). Un autre résultat intéressant est le fait que pour les positions finales, les voyelles finales mais également pénultièmes ont été considérées pour l'analyse. Par contre, pour les positions finales, seules les voyelles strictement initiales ont été prises en compte. Comme suggéré par Byrd et al. (2006) pour l'anglais américain, l'empan semble plus important sur les positions finales que sur les positions initiales.

Pour finir, nous avons remarqué que les valeurs de durée et de f0 augmentaient parallèlement aux valeurs de dispersion, en remontant la hiérarchie prosodique. Il semblait dans un premier temps que ces paramètres étaient liés puisque la f0 et la durée sont connus pour marquer la présence de frontières. Des mesures de corrélations ont été effectuées mais se sont révélées faibles, ce qui pourrait suggérer que des phénomènes de compensation pourraient exister, au sein de stratégies entre locuteurs, entre des variations spectrales et des variations prosodiques pour marquer la présence de frontières.

BIBLIOGRAPHIE

- [1] Boersma, P. & D. Weenink (2009). Praat: doing phonetics by computer (Version 5.1.22) [Computer program]. Retrieved September 15, 2009.
- [2] Bradlow, A.R., G.M. Torretta & D.B. Pisoni (1996). Intelligibility of normal speech: global and finegrained acoustic-phonetic characteristics. *Speech Communication* 20, pp. 255-272.
- [3] Byrd, D., J. Krivokapic & S. Lee (2006). How far, how long: On the temporal scope of phrase boundary effects. *J. Acoust. Soc. Am.* 120, pp. 1589-1599.
- [4] Clément, L., B. Sagot & B. Lang (2004). Morphology based automatic acquisition of large-coverage lexica. In *proc. of LREC'04*, Lisboa, Portugal, pp. 1841-1844.
- [5] Fougeron C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics* 29:2, pp. 109-135.
- [6] 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 *Proc. of Interspeech*, Lisboa, Portugal, pp. 1149-1152.
- [7] 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 *Proc. of Interspeech* – Lisboa, Portugal, pp. 2453-2456.
- [8] Gendrot, C., M. Adda-Decker & J. Vaissière (2008). Les voyelles /i/ et /y/ du français : focalisation et variations formantiques. In *proc. of 26èmes Journée d'Etude de la Parole*, Avignon, France, pp.205-208.
- [9] Lindblom B. (1963). Spectrographic study of vowel reduction. *J. Acoust. Soc. Am.* 35, pp.1773-1781.
- [10] Nespore M. & I. Vogel (1986). *Prosodic phonology*, Dordrecht : Foris.
- [11] Pallier, C. (1994). Rôle de la syllabe dans la perception de la parole : études attentionnelles. *PhD thesis*, EHESS, Paris.
- [12] 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.
- [13] 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.
- [14] Tabain M. (2003). Effects of prosodic boundary on /aC/ sequences: acoustic results. *J. Acoust. Soc. Am.* 113, pp. 516-531.
- [15] Tabain M. & Perrier (2005). Articulation and acoustics of /i/ in pre-boundary position in French. *Journal of Phonetics*, 33, pp. 77-100.
- [16] Vaissière, J. (1971). Contribution à la synthèse par règles du français. Thèse de doctorat, Université de Grenoble.