

**Cédric Gendrot**

Université Sorbonne Nouvelle – Paris 3 & Laboratoire de Phonétique et Phonologie (CNRS UMR 7018)

**Kim Gerdes**

Université Sorbonne Nouvelle – Paris 3 & Laboratoire de Phonétique et Phonologie (CNRS UMR 7018)

**Martine Adda-Decker**

Laboratoire de Phonétique et Phonologie (CNRS UMR 7018) & Université Sorbonne Nouvelle – Paris 3

---

# Détection automatique d'une hiérarchie prosodique dans un corpus de parole journalistique

## 1. INTRODUCTION

Cette étude vise à décrire la variabilité des phonèmes en français grâce à l'utilisation de corpus de grande taille transcrits orthographiquement, segmentés phonétiquement, puis analysés automatiquement. Grâce à ces corpus, 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 acoustique des voyelles était grandement influencée par leur durée : si l'on considère un espace acoustique formé par les deux premiers formants de chaque voyelle, plus les voyelles sont longues et plus l'espace vocalique occupé par l'ensemble de ces voyelles est important (Fig. 1-2). Ce résultat a également été observé pour d'autres langues telles que l'allemand, l'anglais, le mandarin, le portugais, l'espagnol, l'italien et l'arabe (Gendrot & Adda-Decker 2007).

Une 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 par l'auditeur. Ce résultat peut être interprété à l'inverse : à savoir que les voyelles les plus courtes ont des réalisations qui s'éloignent de leurs valeurs de référence et qui sont ainsi considérées comme réduites. Le lien entre la réalisation spectrale des voyelles (à travers leurs valeurs de formants) et leur durée a été validée depuis longtemps par B. Lindblom (1963) et beaucoup d'autres depuis. Il a également été montré que la réduction des voyelles courtes, telle que l'on semble l'observer sur les Figures 1 et 2, n'est pas une simple centralisation de l'espace acoustique, mais plutôt de la coarticulation : les voyelles courtes prennent les caractéristiques acoustiques des consonnes, ce qui les attire dans la majorité des contextes vers le centre de l'espace acoustique.

La relation entre la réalisation acoustique et l'articulation des voyelles peut être estimée grâce aux mesures de formants (Vaissière 2007 et références incluses). En effet, le premier formant (F1) est généralement associé à l'aperture de la voyelle, particulièrement pour les voyelles ouvertes et semi-ouvertes, alors que le deuxième formant (F2) est plus dépendant de l'antériorité/postériorité de la voyelle. Quant au troisième formant (F3), il est lié principalement à l'arrondissement des voyelles, notamment pour les voyelles /y/, /ø/ et /œ/. Les voyelles centrales /ø/ et /œ/ ne sont pas incluses dans cette étude (Bürki *et al.* 2010). Nous nous concentrerons ici sur les voyelles périphériques pour une approximation de l'espace vocalique par l'ensemble de ces voyelles. Les variations de /y/ (et de /i/) sont également détaillées dans d'autres études car leurs différentes réalisations peuvent dépendre de F3 plus que les autres voyelles (Gendrot, Adda-Decker & Vaissière 2008 et références incluses). Notre intérêt se porte ici essentiellement sur les variations de F1 (corrélé à l'aperture de la voyelle) et de F2 (relié à l'antériorité de la voyelle).

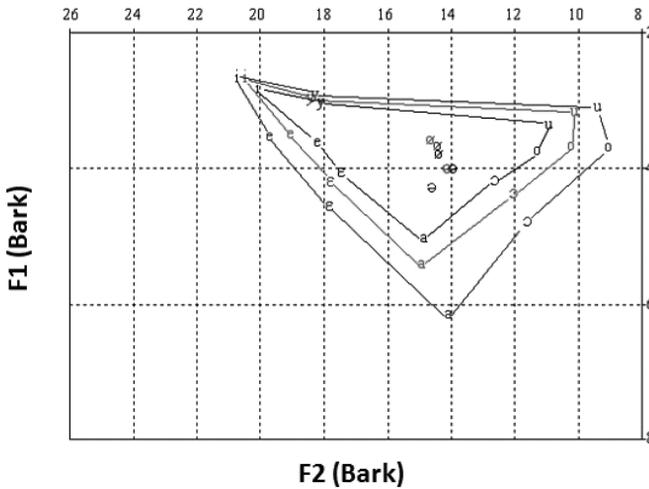
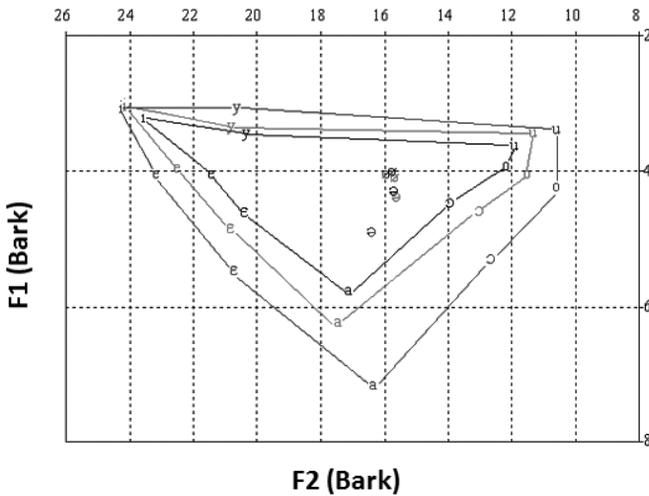


Figure 1 : Valeurs moyennes de F1/F2 des voyelles orales du français en fonction de leur durée, ici pour les hommes en Bark (Hertz normalisés) <sup>a</sup>

a. De l'intérieur vers l'extérieur : noir [30-50 ms], rouge [60-80], noir [90-110].



**Figure 2 : Valeurs moyennes de F1/F2 des voyelles orales du français en fonction de leur durée, ici pour les femmes en Bark <sup>a</sup>**

a. De l'intérieur vers l'extérieur : noir [30-50 ms], rouge [60-80], noir [90-110].

Les raisons des variations de durée des voyelles sont multiples et des facteurs, tels que le style et le débit du locuteur, ont une grande influence. Des facteurs plus linguistiques – tels que la prosodie, le contexte segmental, la position de la voyelle dans la syllabe, le mot, le syntagme ou l'énoncé – ont également une influence significative. Les quatre dernières unités mentionnées sont fréquemment traitées comme des constituants prosodiques puisque représentées par des variations caractéristiques de durée et de  $f_0$  dans un but démarcatif : ils signalent des occurrences de frontières (Vaissière & Michaud 2006). Ces constituants prosodiques sont également considérés comme étant organisés au sein d'une hiérarchie prosodique : chaque constituant est imbriqué dans un constituant de plus haut niveau (Nespor & 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 segments aux frontières de ces constituants sont longs (pour le français, Fougeron 2001 ; Tabain 2003). Ces positions ont plus rarement été étudiées d'un point de vue spectral (et/ou articulatoire). Pour le français, C. Fougeron (2001) à partir de l'électropalatographie (EPG) et M. Tabain (2003), M. Tabain et P. Perrier (2005) grâce à l'articulographie (EMMA : ElectroMagnetic Midsagittal Articulatograph) ont travaillé respectivement sur les positions initiales et finales. Leurs résultats montrent qu'un phonème en position initiale ou finale de constituant sera influencé par sa position dans la hiérarchie prosodique. Plus le niveau sera élevé, plus le phonème sera renforcé (voir aussi Cho 2005 pour l'anglais américain).

Nous tentons ici de répliquer ces études sur de la parole continue, à partir de mesures acoustiques seulement, mais pour toutes les voyelles périphériques du français. Quatre constituants prosodiques unanimement reconnus ont été choisis : la syllabe, le mot, le syntagme accentuel et le syntagme intonatif. Notre objectif est de détecter des catégories s'approchant de ces constituants et de les analyser à leurs frontières (en position initiale et finale) dans une procédure entièrement automatisée. Pour une de ces catégories, nous nous attacherons notamment à tester un outil de découpage syntaxique (section 2.4).

## 2. MÉTHODE

### 2.1. Précédents résultats et hypothèses

Nous souhaitons montrer que les voyelles positionnées à des frontières de constituants de plus haut niveau sont renforcées (*i.e.* mieux réalisées) comparées à leurs homologues de constituants de plus bas niveau. Ces résultats valideraient la détection automatique de nos catégories. Pour ce faire, nous évaluerons l'espace acoustique utilisé par l'ensemble des voyelles périphériques pour chacune des catégories choisies (en analysant les positions initiales et finales séparément). Selon les hypothèses avancées précédemment, nous nous attendons à un élargissement de l'espace vocalique à mesure que l'on remonte dans la hiérarchie de ces constituants, *i.e.* en partant de la syllabe, puis du mot, en passant par le groupe accentuel et pour finir le groupe intonatif. Afin de quantifier ces différences, nous calculerons la dispersion (par distance euclidienne) pour chaque niveau prosodique entre chaque voyelle et le centre acoustique (mesuré sur l'intégralité des données F1 à 450 Hz et F2 à 1550 Hz, calculé dans Gendrot & Adda-Decker (2007) pour le français, et inspiré de Bradlow, Torreta & Pisoni (1996) comme illustré dans la Figure 3). Si la voyelle s'éloigne du centre acoustique d'un niveau prosodique à l'autre de façon significative alors elle sera considérée comme renforcée. Nous gardons à l'esprit que cette mesure peut s'avérer imprécise puisque liée au degré de centralisation de la voyelle qui, comme précisé *supra*, n'est qu'une conséquence secondaire de la coarticulation. Cependant, lorsque toutes les voyelles s'éloignent les unes des autres afin de favoriser leur distinction, elles accroissent nécessairement l'espace vocalique (voir la théorie de la dispersion adaptative de B. Lindblom pour une interprétation de ce phénomène dans les préférences des systèmes vocaliques des langues du monde).

Pour conforter la validation de nos catégories vis-à-vis des catégories prosodiques recherchées, des mesures de durée et de f0 ont également été prises sur les voyelles.

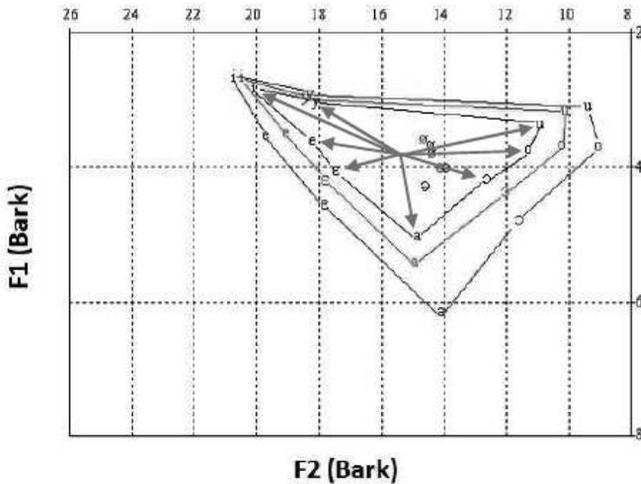


Figure 3 : Illustration de la mesure de dispersion vocale choisie ici

Dans les sections suivantes, nous résumerons les procédures ainsi que les méthodes utilisées dans les précédentes études sur lesquelles nous nous basons. Dans un premier temps, nous détaillerons les corpus et la segmentation automatique et, dans un deuxième temps, nous décrirons les mesures effectuées ainsi que les précautions prises pour éviter des erreurs de mesures automatiques.

## 2.2. Corpus et analyses

Les données de parole utilisées correspondent approximativement à une trentaine d'heures de parole radiophonique (environ 500 hommes et 300 femmes) 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 A. Bürki *et al.* (2008). Il s'agit de parole journalistique préparée : la parole est produite à un rythme soutenu mais doit rester facilement compréhensible pour un large public ; quelques hésitations, répétitions et faux-départs viennent ponctuer des structures syntaxiques qui restent souvent proches de l'écrit. Les phénomènes de réduction et de renforcement qui seront mentionnés dans cette étude y sont certainement moindres que dans un corpus de parole spontanée.

La transcription orthographique étant connue *a priori*, le système d'alignement est utilisé pour localiser les frontières de début et de fin, pour choisir parmi les alternatives potentielles de prononciation (en particulier, les liaisons et la présence de schwas) et pour identifier les silences, respirations et autres bruits. Des modèles de phones indépendants du contexte sont utilisés pour l'alignement. Alors que des modèles acoustiques dépendants du contexte (triphones) produisent des transcriptions de meilleure qualité (*i.e.* des taux d'erreur de détection de mots plus faibles), les modèles acoustiques indépendants du contexte

sont plus efficaces pour repérer les frontières de segments (Bürki *et al.* 2008). Pour des raisons techniques, la précision de la segmentation est limitée à 10 ms et la durée minimale d'un phonème est de 30 ms (pour information, la durée moyenne d'un phonème se situe à 70 ms environ). L'étiquetage ainsi produit n'est pas un étiquetage phonétique, mais plutôt un étiquetage phonémique correspondant à la transcription standard des mots. Les variations dans la prononciation des sons seront évaluées par des mesures acoustiques de durée, de  $f_0$  et des formants que nous présentons *infra*.

### 2.3. Extraction automatique des formants et mesures acoustiques

Les trois premiers formants (F1, F2 et F3) ont été mesurés automatiquement à l'aide de l'algorithme Burg implémenté dans PRAAT (Boersma & Weenink 2009). La détection des pics d'amplitude dans le spectre est limitée à une bande inférieure à 5 kHz pour les locuteurs masculins et inférieure à 5.5 kHz pour les locuteurs féminins. Les mesures ont été prises sur la partie médiane de la voyelle, respectivement à 1/3, 1/2 et 2/3 ; ces trois points ont ensuite été moyennés pour fournir une valeur unique par voyelle. L'interprétation des pics d'amplitude ainsi extraits en tant que formants peut poser problème dans un certain nombre de cas : bruits, nasalité, etc. Deux précautions méthodologiques ont été prises afin de prévenir les erreurs : (i) les voyelles nasales sont exclues de nos analyses ; (ii) les valeurs de pics d'amplitude sont filtrées dans le but de rejeter les items aberrants par rapport aux limites acceptables du conduit vocal. Des fourchettes de valeurs ont été établies afin de filtrer les valeurs fantaisistes (approximativement 4 % au final) dues à des erreurs de détection automatique. Plus de détails sur les mesures de formants, les précautions prises et leur interprétation sont fournis dans C. Gendrot et M. Adda-Decker (2005) et M. Adda-Decker, C. Gendrot et N. Nguyen (2008).

Les valeurs de formants ont été normalisées par locuteur selon la procédure dite de B. Lobanov (1971). Cette normalisation consiste à pondérer la valeur d'un formant pour une voyelle et un locuteur donné par la moyenne et l'écart-type des différentes réalisations de cette même voyelle et de ce même locuteur. Cette procédure, parmi les plus anciennes, est reconnue pour conserver les différences phonémiques tout en réduisant les différences inter-locuteurs, ce qui est l'objectif visé ici (Adank, Smits & van Hout 2004 et références incluses). Cette normalisation est considérée comme une précaution méthodologique plus qu'une nécessité, puisque les analyses présentées dans cet article ont été effectuées avec et sans normalisation pour des résultats très semblables ; l'utilité de la normalisation dans le cas de grands corpus est discutée dans C. Gendrot (2013). Les valeurs de durée ont également été normalisées, selon une procédure nouvelle prenant en compte le contexte segmental (Gendrot, Adda-Decker & Schmid 2012).

## 2.4. Catégories d'analyse

Quatre catégories ont été analysées dans cette étude, dans le but de s'approcher des quatre catégories prosodiques mentionnées dans l'introduction. De la plus petite à la plus grande en termes de taille (la Table 4 en section 4.4 récapitule leurs caractéristiques temporelles) : la syllabe, le mot, le groupe accentuel potentiel et le groupe intonatif potentiel. Dans notre approche, il ne peut y avoir aucun recouvrement entre les différentes catégories détectées : une voyelle analysée ne peut pas appartenir à deux catégories différentes, nous décrivons *infra* les choix effectués pour détecter chaque catégorie, en commençant par le mot. Les items correspondant à chaque catégorie seront analysés à leurs frontières, *i.e.* à leurs positions initiales et finales.

Le mot correspond ici à l'unité orthographique et les frontières de mots ont été obtenues sur la base de la transcription manuelle, puis sur la base de l'alignement effectué automatiquement. Les voyelles décrites comme initiales de mots sont considérées comme des initiales absolues, comme par exemple dans les mots *arme* [aʁm] et les positions décrites comme finales de mot sont en finale absolue, comme par exemple dans le mot *bras* [bʁa]. Nous aborderons dans la discussion la possibilité de prendre en compte la position en pénultième comme pour le mot *trame* [tʁam] et la deuxième position du mot comme pour le mot *larne* [laʁm]. Les voyelles initiales de mots sont fréquemment produites avec une liaison en français (« deux [dø] amis [ami] » → [døzami]). Dans le processus de segmentation automatique, les liaisons sont annotées si elles sont prises en compte dans le dictionnaire de prononciation. Si l'appartenance phonologique de la consonne de liaison est attribuée au deuxième mot en phonologie, elle est considérée dans le système automatique que nous avons utilisé comme appartenant au premier mot. Remarquons pour finir que la phonotactique du français implique que /ɔ/ ne peut pas se trouver en position finale et que, par conséquent, aucun résultat pour cette voyelle ne sera présenté en position finale.

Les syllabes sont également déterminées à partir de la segmentation, elles correspondent aux syllabes phonétiques. Des règles de syllabation inspirées de C. Pallier (1994) et M. Adda-Decker *et al.* (2005) ont été utilisées à partir du flux continu de phonèmes, *i.e.* sans prendre en compte les frontières de mots. Par exemple, la séquence de deux mots *bon ami* [bɔnami] est segmentée en trois syllabes : [bɔ], [na] et [mi] à moins que ces deux mots ne soient séparés par une pause. Les pauses sont considérées comme des délimiteurs et les syllabes, selon ce principe, ne peuvent pas contenir de pauses. Contrairement aux autres catégories, les syllabes ne seront pas analysées à leurs positions initiales et finales. En effet, les voyelles en position initiale de syllabe, mais non initiales de mot, sont extrêmement rares. Les seules syllabes de ce type sont les syllabes de type 'V' contenues par exemple dans le mot *aéroport* qui représentent 2 % des syllabes dans le corpus étudié ici. Nous avons donc décidé de prendre en compte les syllabes médianes au sein du mot (ni finales, ni initiales de mots), sans filtrer le type de syllabe. Par conséquent, les syllabes considérées ici sont de façon

prédominante (à 78 %) des syllabes ouvertes où la voyelle est en position finale de syllabe.

Le troisième niveau analysé ici est le *groupe accentuel potentiel* (GAP : voir l'exemple en Tableau 1) qui s'approche du groupe accentuel prosodique, mais sur une base syntaxique. Afin d'obtenir ce niveau, un *chunking* syntaxique a été effectué sur la base d'un étiquetage grammatical automatique combiné à quelques règles de regroupement mises en place grâce au *chunker* du *Natural Language Toolkit* ([http://nltk.org/index.php/Main\\_Page](http://nltk.org/index.php/Main_Page)) :

- Dans un premier temps, chaque mot-forme a été étiqueté de la catégorie donnée dans le dictionnaire français des formes fléchies (Lefff ; Clément, Sagot & Lang 2004), légèrement adapté à nos besoins. Par exemple, nous avons supprimé les catégories de mots rares qui se confondaient avec des catégories de mots plus fréquents (p. ex. pour l'adjectif *sûr* qui entraînait des étiquetages erronés).
- Ensuite, le *chunker* du NLTK (*Natural Language Toolkit*) a été utilisé pour générer deux types de segments :
  - les noms, les prépositions et les verbes ont été regroupés avec leur entourage le plus proche (clitiques, déterminants, prépositions, adjectifs, etc.),
  - toutes les séquences de mots, non concernés par la règle précédente, ont été regroupées.

Notons que le regroupement conjoint des verbes et noms avec leur entourage permet de se passer, le plus souvent, d'une désambiguïisation complète de chaque mot forme pluri-catégorielle : la suite de mots *les portions* se retrouvera dans un GAP, peu importe qu'il s'agisse de la suite <déterminant nom> ou de <clitique verbe>.

- Dans une étape finale, trois règles de regroupement ont été appliquées :
  - regroupement de tout segment terminant sur un auxiliaire ou modal avec le segment suivant,
  - regroupement de tout segment verbal avec le segment suivant si la combinaison fait moins de sept syllabes (Wioland 1985, Martin 2011),
  - regroupement de tout autre suite de segments qui fait moins de sept syllabes.

Un exemple du résultat de cet algorithme est illustré par la Figure 4.

Les segments découpés par cet algorithme peuvent avoir plus de sept syllabes, si les règles précédentes le permettent, par exemple avec *avec qui j'ai pu m'entretenir*, qui forme un groupe très naturel et difficile à découper. La règle des sept syllabes (Wioland 1985, Martin 2011) sera discutée ultérieurement dans cette étude.

Par ce *chunking*, nous tentons de nous approcher de la réalisation du groupe accentuel. Nous sommes conscients que tous ces groupes ne seront pas « accentués », *i.e.* qu'ils ne seront pas tous caractérisés par un allongement final et/ou

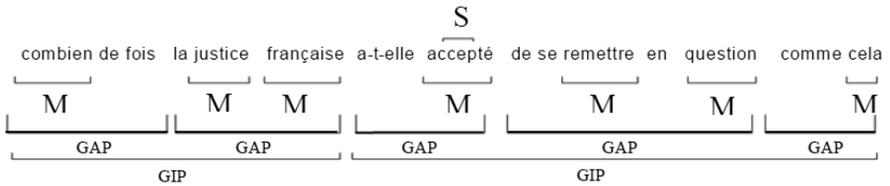


Figure 4 : Illustration du *chunking* en groupes accentuels potentiels <sup>a</sup>

a. S : syllabe – M : mot – GAP : groupe accentuel potentiel – GIP : groupe intonatif potentiel

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 groupes accentuels à un niveau syntaxique, plutôt qu'en considérant des groupes accentuels d'après leurs caractéristiques prosodiques. Nous étudions ainsi aussi la possibilité de calculer les groupes accentuels en se basant uniquement sur la suite des mots, très utile pour la synthèse de la parole ainsi que pour les modèles d'analyse automatique de la langue parlée.

La quatrième et dernière catégorie analysée est le *groupe intonatif potentiel* (GIP), détecté automatiquement sur la base des pauses (supérieures ou égales à 60 ms) indiquées par l'alignement. Ce seuil a été choisi pour éviter les nombreux petits segments étiquetés comme des pauses qui peuvent être insérés quand il y a une faible correspondance entre le signal et sa transcription. Nous nous sommes arrêtés sur ce critère de sélection puisqu'il a été montré que la présence de pauses est un facteur important pour signaler la réalisation d'un groupe intonatif (Jun & Fougeron 2000). Une détection de la forme du contour final de  $f_0$  (montant/descendant) a également été effectuée dans le but de ne prendre en compte que les syntagmes ayant un contour montant. Cela permet 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 groupes accentuels potentiels sera proposée ici : l'objectif de cette étude est d'obtenir quatre catégories, aussi proches que possible des catégories prosodiques mentionnées dans la littérature, sans se baser sur des caractéristiques prosodiques d'allongement.

Une analyse supplémentaire avec des valeurs de durée et de  $f_0$  permettra aussi de tester les quatre catégories ainsi sélectionnées, confirmant la fiabilité de ces catégorisations. La très large quantité de données utilisée pourrait également aider à compenser les éventuelles erreurs de catégorisation. Le Tableau 1 *infra* résume le nombre de voyelles analysées par catégorie en position initiale et finale.

**Tableau 1 : Distribution des voyelles analysées en fonction de leur catégorie et de leur position (initiale ou finale)**

Position initiale				
voyelle	syllabe	mot	GA potentiel	GI potentiel
e	3 629	5 948	1 041	679
ɛ	4 065	5 204	1 005	133
a	6 695	11 717	3 217	498
ɔ	3 773	3 134	451	48
o	748	2 522	714	164
u	716	306	52	47
y	1 893	3 021	1 055	81
i	8 525	3 924	2 643	299
<b>Total</b>	<b>30 044</b>	<b>35 776</b>	<b>10 178</b>	<b>1 949</b>
Position finale				
voyelle	syllabe	mot	GA potentiel	GI potentiel
e	3 629	19 841	11 129	330
ɛ	4 065	6 765	3 150	107
a	6 695	17 637	5 633	191
ɔ	3 773	0	0	0
o	748	2 474	1 995	78
u	716	2 648	853	40
y	1 893	5 743	1 259	31
i	8 525	10 111	5 866	167
<b>Total</b>	<b>30 044</b>	<b>65 220</b>	<b>29 886</b>	<b>1 264</b>

## 2.5. Analyses statistiques

L'utilisation de grands corpus de parole continue apporte un certain nombre d'avantages tel qu'un nombre très important d'occurrences produites de façon naturelle. Cependant, par rapport à un corpus de parole contrôlée, un certain nombre de paramètres, comme par exemple le contexte segmental et morpho-syntaxique, ne peuvent être contrôlés, et alors que certains contextes sont extrêmement fréquents, d'autres au contraire sont plus rares. Dans ce but, la durée des voyelles a été normalisée en tenant compte de leur contexte segmental (section 2.3) pour la durée tandis que les valeurs de  $f_0$  sont moins influencées par le contexte segmental. Pour les valeurs de formants cependant, il sera utile d'inclure des informations segmentales dans l'analyse statistique. Le contexte segmental, conformément à la méthode utilisée dans de précédentes études (Gendrot & Adda-Decker 2005, Bürki *et al.* 2010), est regroué en quatre lieux

d'articulation : labial, dental, palato-vélaire et uvulaire pour le contexte précédent et suivant. Ces regroupements ont été choisis puisqu'ils ont une influence similaire sur les transitions formantiques des voyelles adjacentes (Gendrot 2013). L'information morpho-syntaxique (ici sous la forme de classes grammaticales) sera gérée grâce à l'utilisation de modèles linéaires à effets mixtes avec, comme facteurs aléatoires, les différents mots analysés et les locuteurs du corpus (afin de tenir compte de la variation inter-locuteurs) (Baayen, Davidson & Bates 2008). Le nombre de syllabes dans le mot – calculé grâce à l'algorithme de syllabification mentionné dans le choix des catégories prosodiques – a également été intégré dans le modèle statistique comme un facteur aléatoire. Il est connu (Nooteboom 1997) que plus un mot contient de syllabes, plus les phonèmes de ce mot seront courts. La fréquence lexicale (transformée en unité logarithmique) et le débit (calculé comme le nombre de phonèmes par seconde incluant les pauses) ont été insérés comme des facteurs covariants puisqu'ils sont connus pour avoir une influence potentielle sur les réalisations des voyelles. Un modèle différent a été calculé pour chaque voyelle.

Les résultats qui nous intéressent ici sont, non pas l'influence des différents paramètres évoqués *supra* sur la réalisation des voyelles, mais l'influence des constituants prosodiques une fois que le poids de ces différents paramètres a été pris en compte. Seuls les résultats concernant les constituants prosodiques automatiquement extraits seront évoqués dans les sections suivantes.

### 3. RÉSULTATS

#### 3.1. Positions initiales

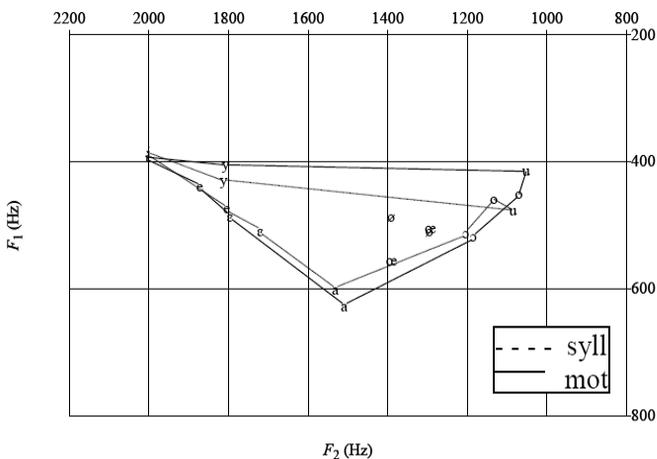


Figure 5 : Comparaison de deux catégories d'analyse en position initiale (de l'intérieur vers l'extérieur : syllabe vs. mot)

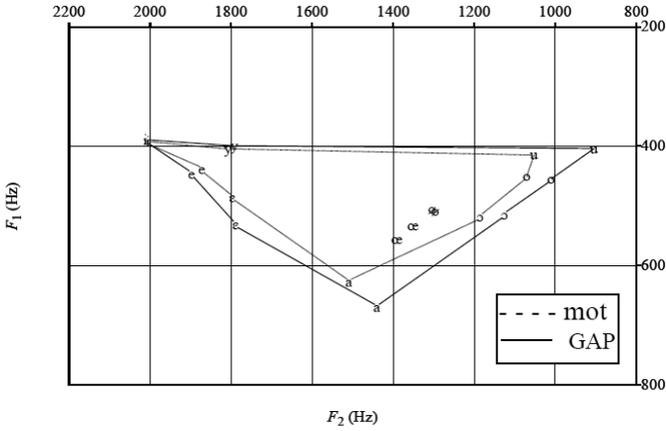


Figure 6 : Comparaison de deux catégories d'analyse en position initiale (de l'intérieur vers l'extérieur : mot vs. groupe accentuel potentiel)

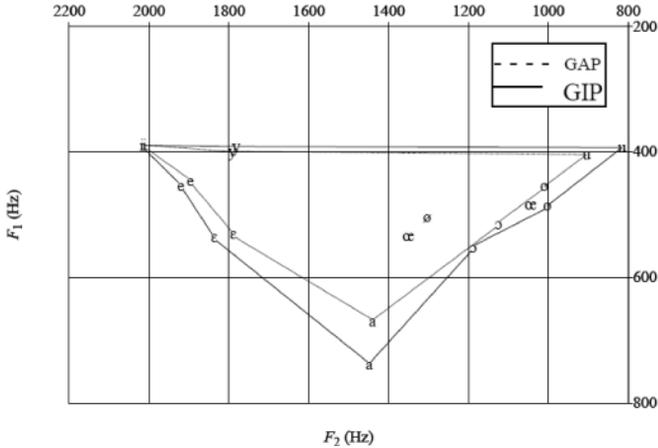


Figure 7 : Comparaison de deux catégories d'analyse en position initiale (de l'intérieur vers l'extérieur : groupe accentuel potentiel vs. groupe intonatif potentiel)

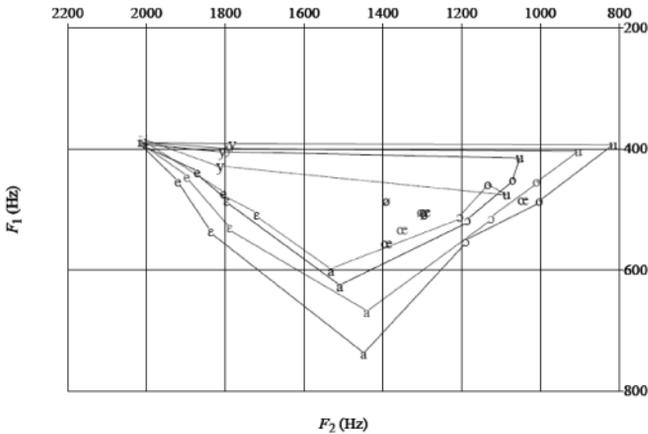


Figure 8 : Comparaison des quatre catégories d'analyse en position initiale (de l'intérieur vers l'extérieur : syllabe, mot, GAP et GIP)

Comme le montrent les Figures 5 à 8, les voyelles occupent un espace acoustique de plus en plus important en remontant les catégories d'analyse. Les mesures de dispersion présentées par la Figure 9 *infra* révèlent que les valeurs augmentent progressivement d'une catégorie à la suivante : les voyelles s'éloignent du centre de l'espace vocalique, étant ainsi plus distinctes les unes des autres perceptivement. Le parallèle que nous souhaitons faire avec la hiérarchie prosodique semble se confirmer ici. Seules les voyelles /i/ et /y/ varient moins.

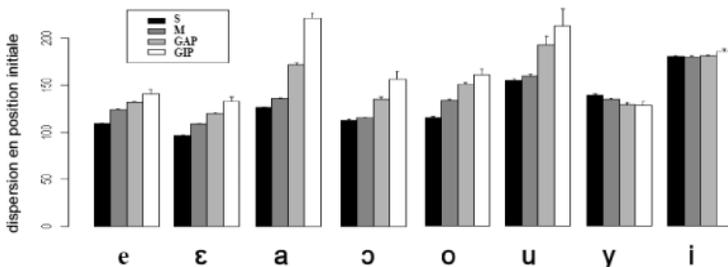


Figure 9 : Dispersion en position initiale en fonction des catégories d'analyse <sup>a</sup>

a. S : syllabe – M : mot – GAP : groupe accentuel potentiel – GIP : groupe intonatif potentiel

La Table statistique (Tableau 2 *infra*) montre que, si la catégorie d'analyse a une influence significative sur les valeurs de dispersion pour toutes les voyelles, seul /a/ affiche des différences significatives pour chaque niveau. Les voyelles /ε/, /e/, /o/ montrent une dispersion significative entre trois niveaux sur les quatre. Pour /o/ et /u/, seuls deux niveaux peuvent être distingués significativement,

dans les deux cas, les niveaux syllabe/mot *vs* groupe accentuel/intonatif potentiel. Pour la voyelle /i/, seul le dernier niveau (groupe intonatif potentiel) a une valeur de dispersion plus élevée que les autres. /y/ révèle une tendance inverse à celle observée pour les autres voyelles. Nous avons montré dans une étude précédente (Gendrot, Adda-Decker & Vaissière 2008) que /i/ et /y/ révèlent moins de variation dans le plan F1/F2 que les autres voyelles, car les variations pour ces voyelles en français sont ciblées sur F3 et F4.

**Tableau 2 : Résultats des tests statistiques pour les mesures de dispersion en fonction de la voyelle et de la catégorie d'analyse en position initiale <sup>a</sup>**

voyelle	significativité	test <i>post-hoc</i>
e	F(3,11293)=5.0517, p=0.001696	S * < M * < GAP < GIP
ɛ	F(3,10403)=13.980, p=4.3e-09	S * < M * < GAP < GIP
a	F(3,22123)=8.0357, p=2.387e-05	S * < M * < GAP * < GIP
ɔ	F(3,7481)=12.5, p=3.8e-08	S < M * < GAP * < GIP
o	F(3,4144)=41.182, p=0	S * < M * < GAP < GIP
u	F(3,1112)=4.8827, p=0.002235	S < M * < GAP < GIP
y	F(3,6046)=3, p=0.029	S = M > GAP = GIP
i	F(3,15387)= 9.8715, p=1.686e-06	S = M = GAP < GIP

a. Les '\*' indiquent une différence significative à  $p < 0.05$  (t value > 2)

Une analyse statistique identique à celle que nous venons d'effectuer, mais avec F3 et F4 comme valeurs dépendantes, a été réalisée pour /i/ et /y/ et montre que, pour /y/, à mesure que l'on monte dans la hiérarchie des constituants, F3 baisse significativement (mais pas F4). Pour /i/, F4 baisse significativement à mesure que l'on monte dans cette même hiérarchie (mais pas F3). Ces variations sur F3 et F4 permettent à /i/ et /y/ d'être plus focales en rapprochant les formants F3/F4 et F2/F3 respectivement (Schwartz *et al.* 1997 ; Gendrot, Adda-Decker & Vaissière 2008). En termes articulatoires, /i/ serait plus étiré à mesure qu'il monte dans la hiérarchie prosodique alors que /y/ serait plus arrondi, ce qui accroît leurs caractéristiques articulatoires.

Un rapide coup d'œil aux valeurs de durée et de f0 dans les Figures 10 et 11 permet d'observer que la durée des voyelles analysées, et dans une moindre mesure leur f0, augmentent avec le niveau potentiel de la hiérarchie prosodique, ce qui confirme que les phonèmes sont renforcés, non seulement au niveau spectral mais également au niveau prosodique. Remarquons que, pour les valeurs de durée, les variations pour /i/ et /y/ se distinguent à nouveau de celles des autres voyelles.

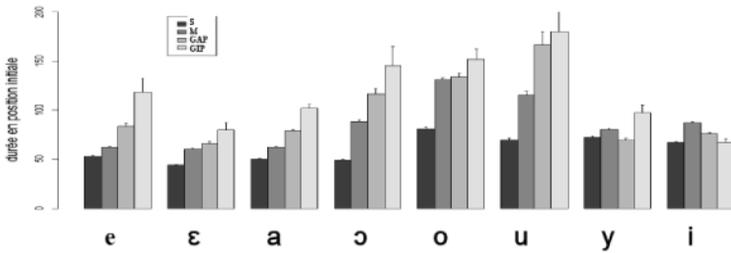


Figure 10 : Durée en position initiale en fonction des catégories d'analyse <sup>a</sup>

a. S : syllabe – M : mot – GAP : groupe accentuel potentiel – GIP : groupe intonatif potentiel

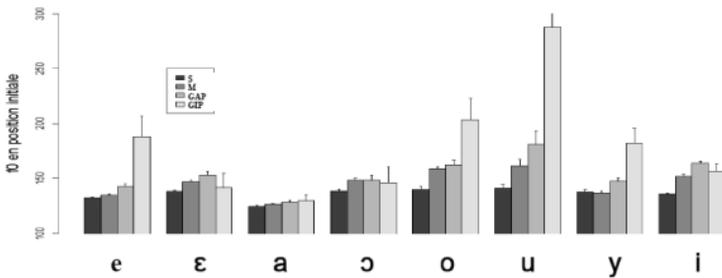


Figure 11 : f0 en position initiale en fonction des catégories d'analyse <sup>a</sup>

a. S : syllabe – M : mot – GAP : groupe accentuel potentiel – GIP : groupe intonatif potentiel

### 3.2. Positions finales

D'après les Figures 12 à 15, et comme observé dans la section 3.1 pour les positions initiales, les voyelles occupent un espace acoustique de plus en plus grand en remontant la hiérarchie prosodique pour les positions finales. Encore une fois, les variations observées pour /i/ sont plus faibles que celles observées pour les autres voyelles, tandis que /y/ est caractérisé par des variations erratiques. Les mesures proposées en Figure 16 indiquent la dispersion de ces voyelles en fonction de leur catégorie d'analyse. Les mêmes modèles statistiques que ceux présentés pour les positions initiales ont été effectués sur ces mesures de dispersion et montrent que les voyelles /e/, /o/ et /u/ révèlent des différences significatives pour tous les niveaux. /a/ et /y/ montrent, quant à elles, une variation de dispersion significative pour trois niveaux sur les quatre, et /i/ pour deux niveaux seulement. Nous pouvons noter que les valeurs F du test statistique sont plus importantes avec plus de différences significatives que pour les positions initiales, ce qui suggère un renforcement plus important en position finale. Les variations de F3 et F4 ont également été analysées pour les voyelles /i/ et /y/ comme dans la section précédente : comme pour les positions initiales,

pour /y/, à mesure que l'on monte dans la hiérarchie prosodique potentielle, F3 baisse significativement (mais pas F4). Pour /i/, F4 baisse significativement à mesure que l'on monte dans la hiérarchie prosodique (mais pas F3). Le Tableau 3 *infra* résume ces résultats.

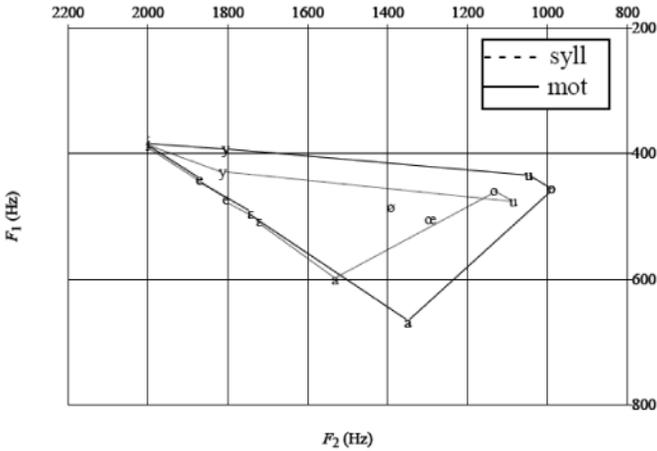


Figure 12 : Comparaison de deux catégories d'analyse en position finale (de l'intérieur vers l'extérieur : syllabe vs. mot)

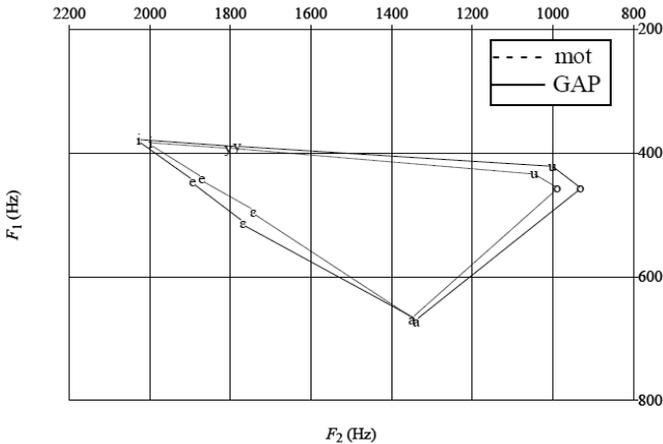


Figure 13 : Comparaison de deux catégories d'analyse en position finale (de l'intérieur vers l'extérieur : mot vs. groupe accentuel potentiel)

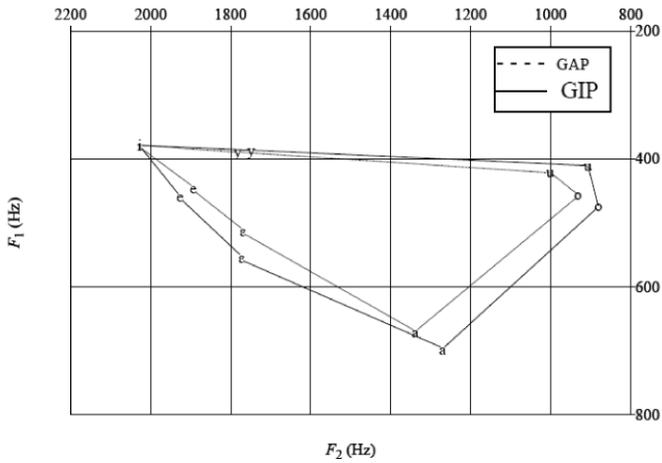


Figure 14 : Comparaison de deux catégories d'analyse en position finale (de l'intérieur vers l'extérieur : groupe accentuel potentiel vs. groupe intonatif potentiel)

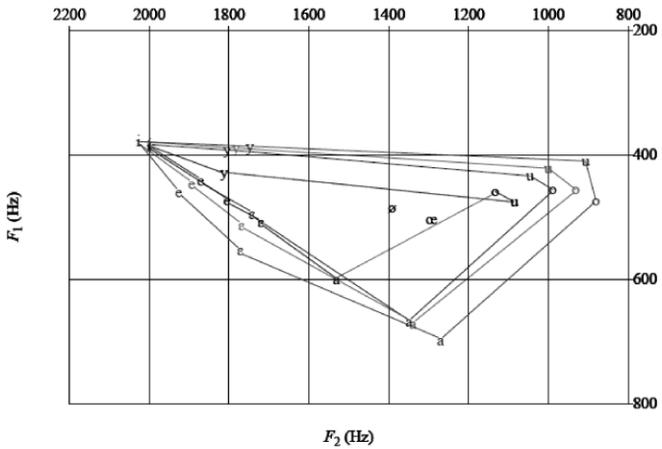


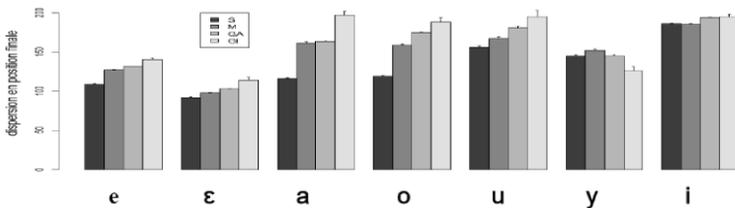
Figure 15 : Comparaison des quatre catégories d'analyse en position finale (de l'intérieur vers l'extérieur : syllabe, mot, GAP et GIP)

**Tableau 3 : Résultats des tests statistiques pour les mesures de dispersion en fonction de la voyelle et de la catégorie d'analyse en position finale <sup>a</sup>**

voyelle	significativité	test <i>post-hoc</i>
e	F(3,34934)=71.704, p=0	S * < M * < GAP * < GIP
ε	F(3,14085)=46.749, p=0	S * < M * < GAP * < GIP
a	F(3,30158)=61.903, p=0	S * < M < GAP * < GIP
o	F(3,5292)=185.13, p=0	S * < M * < GAP * < GIP
u	F(3,4253)=50.404, p=0	S * < M * < GAP < GIP
y	F(3,8925)=29.752, p=0	M * > S = GIP * > GAP
i	F(3,24665)=33.016, p=0	S = M * < GAP < GIP

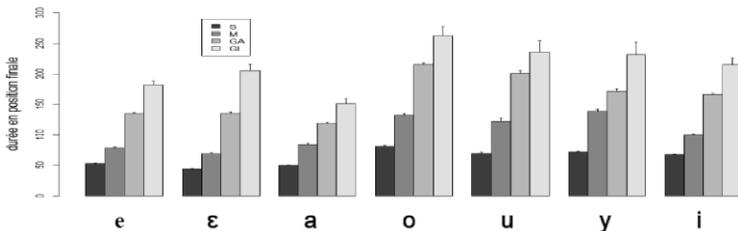
a. Les '\*' indiquent une différence significative à  $p < 0.05$  (t value > 2).

Les valeurs de durée et de  $f_0$  présentées dans les Figures 17 et 18 montrent que ces deux paramètres augmentent avec le niveau de hiérarchie prosodique potentielle. Comme noté pour les mesures de dispersion, l'amplitude de variation entre les niveaux est plus importante que celle observée pour les positions initiales.



**Figure 16 : Dispersion en position finale en fonction de la catégorie d'analyse <sup>a</sup>**

a. S : syllabe – M : mot – GAP : groupe accentuel potentiel – GIP : groupe intonatif potentiel



**Figure 17 : Durée en position finale en fonction de la catégorie d'analyse <sup>a</sup>**

a. S : syllabe – M : mot – GAP : groupe accentuel potentiel – GIP : groupe intonatif potentiel

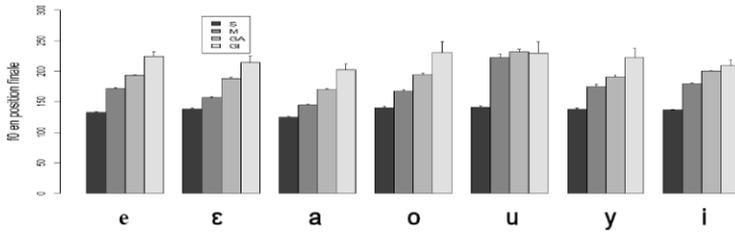


Figure 18 : f0 en position finale en fonction de la catégorie d'analyse <sup>a</sup>

a. S : syllabe – M : mot – GAP : groupe accentuel potentiel – GIP : groupe intonatif potentiel

## 4. DISCUSSION

### 4.1. Résumé des principaux résultats

En accord avec notre hypothèse, nous observons une hiérarchie prosodique (de la syllabe au mot, puis du groupe accentuel potentiel jusqu'au groupe intonatif potentiel) sur la base de mesures spectrales et de mesures prosodiques (de durée et de f0). Plus une voyelle est élevée dans la hiérarchie prosodique des constituants que nous avons déterminée, plus elle est renforcée : cela confirme que les catégories que nous avons détectées se positionnent dans une hiérarchie prosodique similaire à celle observée dans la littérature prosodique. Cependant, comme signalé par les études précédentes (Fougeron 2001, Tabain 2003b), 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 le plan F1/F2 que sur le plan F3/F4. D'autres voyelles telles que /ɔ/, /o/ et /u/ ont révélé des variations parfois non significatives, parfois dans des directions ne correspondant pas aux hypothèses. Ces résultats peuvent être en partie expliqués par une quantité plus restreinte de ces voyelles dans ces positions phonotactiques. Nous pouvons également mentionner qu'une augmentation de f0 (cf. Figures 11 et 18) aboutit acoustiquement à une augmentation des valeurs de F1, ce qui pourrait également expliquer pourquoi les voyelles fermées varient peu sur F1 en fonction des catégories d'analyse.

### 4.2. Détection automatique des catégories prosodiques

La première tentative d'explication de ces différences non significatives se reporte sur la détection automatique des catégories prosodiques potentielles. Nous avons pu remarquer, surtout pour les positions initiales, que les différences non significatives portaient sur les deux plus hautes catégories (groupes accentuels et intonatifs potentiels). Les tendances peuvent aller dans la direction attendue avec des valeurs pour le groupe intonatif potentiel plus élevées que pour le groupe accentuel potentiel, mais sans atteindre systématiquement le seuil de significativité (2 sur 8 pour les positions initiales ; 5 sur 7 pour les

positions finales). Se pose donc la question de savoir si les catégories déterminées automatiquement et intitulées « groupe accentuel potentiel » et « groupe intonatif potentiel » correspondent réellement à ce que l'on attend. Les mesures prosodiques effectuées semblent confirmer que ces catégories ne sont pas attribuées aléatoirement. Malgré tout, la faible quantité de données pour les groupes intonatifs potentiels (cf. Tableau 4) peut jouer sur les taux de significativité pour cette catégorie.

### 4.3. Nombre de syllabes par *chunk* pour les groupes accentuels

Ces résultats nous amènent à un nouveau questionnement à propos du choix des catégories prosodiques à détecter. Y a-t-il un nombre limité de catégories prosodiques ou pourrait-on détecter un continuum graduel de catégories ? La longueur (en phonèmes, mais aussi en millisecondes) de chaque catégorie augmente en montant la hiérarchie prosodique potentielle (syllabe, mot, groupe accentuel potentiel, groupe intonatif potentiel). Si la longueur de la catégorie était en fait un facteur favorisant le renforcement, nous pourrions observer un continuum de catégories prosodiques plutôt qu'un nombre fini. Pour ce faire, nous avons utilisé une nouvelle fois l'algorithme de *chunking* en modifiant le nombre maximum de syllabes. La Figure 19 nous montre la durée du dernier phonème en fonction de la catégorie prosodique potentielle (syllabe, mot, groupe accentuel potentiel, groupe intonatif potentiel), mais en faisant varier le nombre de syllabes maximale autorisé par l'algorithme de *chunking* pour le groupe accentuel (de 5 à 10 syllabes). On s'aperçoit que la variation à l'intérieur de la catégorie « groupe accentuel potentiel » est très faible alors que la durée du groupe augmente (cf. Tableau 4), ce qui conforte l'existence d'un groupe accentuel prosodique indépendamment de sa durée.

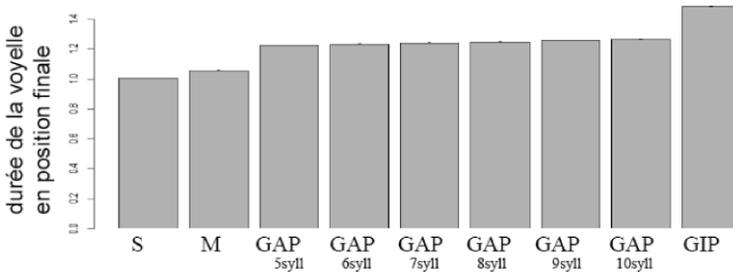


Figure 19 : Durée normalisée de la voyelle en position finale en fonction de la catégorie d'analyse <sup>a</sup>

a. S : syllabe – M : mot – GAP\_n\_syll : groupe accentuel potentiel de n syllabes – GIP : groupe intonatif potentiel

**Tableau 4 : Caractéristiques temporelles principales des constituants d'analyse présentés dans la présente étude et dans la Figure 19**

Constituant	Durée (ms)	Nombre de phonèmes
Syllabe	182	2.5
Mot	361	5
GAP 5 syllabes	887	12.2
GAP 6 syllabes	945	12.9
GAP 7 syllabes	1 015	13.9
GAP 8 syllabes	1 098	15
GAP 9 syllabes	1 187	16.2
GAP 10 syllabes	1 279	17.4
GIP	2 720	38.3

#### 4.4. Contenu informationnel et empan du renforcement

Dans cette étude, aucune sélection n'a été faite sur les classes grammaticales des mots portant les voyelles analysées, mise à part la prise en compte du mot en tant que variable aléatoire dans nos modèles statistiques. Or, la distribution des classes grammaticales n'est pas équilibrée, principalement à cause des débuts de groupes accentuels/intonatifs potentiels plus fréquemment composés de prépositions, conjonctions ou déterminants, alors que pour les deux autres catégories (syllabe et mot), les noms et les adjectifs sont les plus représentés. On pourrait s'attendre à ce que les mots grammaticaux soient moins bien prononcés (*i.e.* renforcés) puisqu'ils sont plus fréquents et qu'ils portent moins d'information cruciale dans la parole (une abondante littérature existe sur la relation entre contenu informatif des mots, ainsi que leur fréquence lexicale, et la réalisation des sons contenus dans ces mots ; voir la théorie de l'hypo- et hyper-articulation de B. Lindblom (1990) et R. Wright (2003) par exemple). On aurait donc pu s'attendre à des débuts de groupes accentuels et de groupes intonatifs avec un renforcement moindre que les débuts de mots par exemple. Les différences notées dans les variations sur les positions initiales et les positions finales pourraient corroborer ce fait. Cependant, il convient de se rappeler que les positions initiales de chaque catégorie prosodique sont concomitantes à la fin d'une catégorie de même niveau qui la précède. Comme proposé par D. Byrd, J. Krivokapic et S. Lee (2006), ces positions sont des moments de ralentissement articulatoire ('*pi-gesture*') qui favorisent le renforcement. Il n'est donc pas si surprenant d'observer les mêmes résultats et phénomènes de renforcement en position initiale et finale de tous nos niveaux prosodiques, et ce, quelle que soit la classe grammaticale impliquée. Si le locuteur choisit de signaler une frontière prosodique dans sa production, on peut émettre l'hypothèse qu'il le fera sans tenir compte de la classe grammaticale des mots autour de cette frontière.

Nous avons également testé l'analyse de voyelles en deuxième position de début ou de fin des constituants que nous avons analysés dans ce travail (respectivement en deuxième position ou en position pré-finale). Nous n'avons pas la place de détailler ces résultats ici, mais comme suggéré par C. Fougeron pour le français (2001) ou par D. Byrd, J. Krivokapic et S. Lee (2006) pour l'anglais américain, l'empan semble plus important pour les positions finales que pour les positions initiales. Le renforcement observé de la voyelle pré-finale est très proche de celui observé pour la voyelle finale alors que le renforcement de la voyelle en position initiale stricte est significativement plus important que le renforcement de la voyelle en deuxième position (post-initiale). Encore une fois, les différences notées entre les variations dans les positions initiales et les positions finales semblent corroborer ce fait.

#### **4.5. Nature du renforcement**

Le dernier point de cette discussion concerne l'impact de ces résultats sur les théories sur la nature du renforcement phonétique dans la parole dans les positions mises en valeur par les locuteurs : s'agit-il d'expansion de la sonorité (Beckman, Edwards & Fletcher 1992) ou d'hyper-articulation des voyelles (de Jong 1995) ? Ces deux hypothèses peuvent être confrontées en observant les valeurs de F1 pour les voyelles fermées. En effet, dans le cas de l'expansion de sonorité, un renforcement des voyelles basses aura pour effet une augmentation de la valeur de F1 alors pour l'hypothèse de l'hyper-articulation, un renforcement des voyelles basses aura pour effet un abaissement de la valeur de F1. En effet, cette dernière prédit que les traits des voyelles renforcés seront accrus, par exemple une voyelle basse sera encore plus basse, une voyelle arrondie sera encore plus arrondie, etc. Pour les autres voyelles, ces deux hypothèses auront des prédictions semblables avec une augmentation des valeurs de F1 pour les voyelles renforcées. Rappelons que pour les voyelles fermées, F1 n'est pas strictement lié à l'aperture car il s'agit d'une résonance de Helmholtz, qui s'abaisse en fonction d'une série de manœuvres articulatoires telles que l'abaissement du larynx, l'avancement de la langue, l'arrondissement des lèvres ou l'aperture. Les Figures 1 et 2 montrent clairement un cas d'hyper-articulation : à mesure que les voyelles basses gagnent en longueur, leur premier formant s'abaisse. Les Figures 5 à 8 pour les positions initiales et les Figures 12 à 15 pour les positions finales révèlent des variations de F1 dans le sens de l'hypothèse de l'hyper-articulation pour les voyelles fermées, bien que moins nettes que pour les Figures 1 et 2. Ces résultats sommaires doivent être confirmés par des analyses statistiques. Les résultats observés sur F3 et F4 pour /i/ et /y/ corroborent également cette hypothèse de l'hyper-articulation des voyelles renforcées puisqu'un /y/ renforcé voit son troisième formant s'abaisser, signalant un arrondissement accru de la voyelle, alors qu'un /i/ renforcé voit son troisième formant monter, signalant un étirement accru de la voyelle.

## 5. CONCLUSION

Dans cette étude qui a porté sur une trentaine d'heures de parole journalistique, nous avons montré qu'il était possible de détecter automatiquement des catégories s'approchant des catégories prosodiques universellement reconnues (syllabe, mot, groupe accentuel, groupe intonatif) grâce à leurs caractéristiques prosodiques et spectrales à leurs frontières. Au-delà des catégories prosodiques potentielles déterminées sur la base de l'alignement automatique (syllabe, mot et groupe intonatif), nous avons présenté un algorithme de *chunking* testé sur un nombre modulable de syllabes qui semble correspondre au groupe accentuel du fait de sa position dans la hiérarchie prosodique mesurée.

Cette étude permet également d'apporter des éléments de réponse aux études phonétiques et phonologiques dans le cadre de la prosodie articulatoire. (i) L'empan du renforcement des phonèmes semble pencher vers la fin des groupes prosodiques plutôt que vers les positions initiales de ces mêmes groupes. (ii) Le renforcement est caractérisé par une hyper-articulation (plutôt qu'une simple expansion de sa sonorité). Ce travail a été effectué sur un corpus de parole journalistique. Il est évident que pour un corpus de parole spontanée, avec une quantité plus importante de disfluences, de répétitions et d'interruptions, le *chunking* serait beaucoup plus délicat à réaliser ; d'autres types de *chunking* basés sur un entraînement statistique pourraient, par exemple, être utilisés. C'est dans cette voie que nous devons porter notre attention dans une future étude.

### Références bibliographiques

- ADANK P., SMITS R. & VAN HOUT R. (2004), "A comparison of vowel normalization procedures for language variation research", *The Journal of the Acoustical Society of America* 116 (5), 3099-3107.
- ADDA-DECKER M., GENDROT C. & NGUYEN N. (2008), « Contributions du traitement automatique de la parole à l'étude des voyelles orales du français », *Revue TAL* 49 (3), 13-46.
- ADDA-DECKER M. *et al.* (2005), "Investigating syllabic structures and their variation in spontaneous French", *Speech Communication* 46, 119-139.
- BAAYEN R. H., DAVIDSON D. J. & BATES D. M. (2008), "Mixed effects modeling with crossed random effects for subjects and items", *Journal of Memory and Language* 59, 390-412.
- BECKMAN M. E., EDWARDS J. & FLETCHER J. (1992), "Prosodic structure and tempo in a sonority model of articulatory dynamics", in G. J. Docherty & D. R. Ladd (eds), *Papers in Laboratory Phonology*, vol. II: *Gesture, Segment, Prosody*, Cambridge: Cambridge University Press, 68-89.
- BOERSMA P. & WEENINK D. (2009), *Praat: doing phonetics by computer* (v. 5.1.22). [<http://www.fon.hum.uva.nl/praat/>]
- BRADLOW A. R., TORRETA G. M. & PISONI D. B. (1996), "Intelligibility of normal speech: global and finegrained acoustic-phonetic characteristics", *Speech Communication* 20, 255-272.
- BÜRKI A. *et al.* (2010), "Phonetic reduction versus phonological deletion of French schwa: Some methodological issues", *Journal of Phonetics* 39 (3), 279-288.

- BÜRKI A. *et al.* (2008), « Alignement automatique et analyse phonétique : comparaison de différents systèmes pour l'analyse du schwa », *Revue TAL* 49 (3), 165-197.
- BYRD D., KRIVOKAPIC J. & LEE S. (2006), "How far, how long: On the temporal scope of phrase boundary effects", *The Journal of the Acoustical Society of America* 120, 1589-1599.
- CHO T. (2005), "Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in English", *The Journal of the Acoustical Society of America* 11 (6), 3867-3878.
- CLÉMENT L., SAGOT B. & LANG B. (2004), "Morphology based automatic acquisition of large-coverage lexica", *Proceedings of 4th International Conference on Language Resources and Evaluation – LREC 2004*, Lisbon (Portugal), 1841-1844.
- DE JONG K. J. (1995), "The supraglottal articulation of prominence in English", *The Journal of the Acoustical Society of America* 97, 491-504.
- FOUGERON C. (2001), "Articulatory properties of initial segments in several prosodic constituents in French", *Journal of Phonetics* 29 (2), 109-135.
- GALLIANO S. *et al.* (2005), "ESTER Phase II: Evaluation campaign for the rich transcription of French broadcast news", *Proceedings of the 9th European Conference on Speech Communication and Technology – INTERSPEECH/EUROSPREECH 2005*, Lisboa (Portugal), 1149-1152.
- GENDROT C. (2013), « Méthodes quantitatives pour le traitement de la variabilité dans les sons de la parole », in N. Nguyen (éd.), *Méthodes et outils pour l'analyse phonétique des grands corpus oraux*, Paris : Hermès/Lavoisier.
- 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 the 9th European Conference on Speech Communication and Technology – INTERSPEECH/EUROSPREECH 2005*, Lisboa (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", *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken (Germany), 1417-1420.
- GENDROT C., ADDA-DECKER M. & SCHMID C. (2012), « Comparaison de parole journalistique et de parole spontanée : analyses de séquences entre pauses », *Actes des xxviii<sup>es</sup> Journées d'Études sur la Parole – JEP 2012*, Grenoble (France), 649-656.
- GENDROT C., ADDA-DECKER M. & VAISSIÈRE J. (2008), « Les voyelles /i/ et /y/ du français : focalisation et variations formantiques », *Actes des xxvii<sup>es</sup> Journées d'Études sur la Parole – JEP'08*, Avignon (France), 205-208.
- JUN S.-A. & FOUGERON C. (2000), "A Phonological model of French intonation", in A. Botinis (ed.), *Intonation: Analysis, Modelling and Technology*, Dordrecht: Kluwer Academic Publishers, 209-242.
- LINDBLOM B. (1963), "Spectrographic study of vowel reduction", *The Journal of the Acoustical Society of America* 35 (5), 1773-1781.
- LINDBLOM B. (1990), "Explaining phonetic variation: A sketch of the H & H theory", in W. Hardcastle & A. Marchal (eds), *Speech Production and Speech Modelling*, Dordrecht: Kluwer, 403-440.
- LOBANOV B. M. (1971), "Classification of Russian vowels spoken by different speakers", *The Journal of the Acoustical Society of America* 49, 606-608.
- MARTIN P. (2011), « Ponctuation et structure prosodique », *Langue française* 172, 99-114.
- NESPOR M. & VOGEL I. (1986), *Prosodic Phonology*, Dordrecht: Foris.

- NOOTEBOOM S. G. (1997), "The prosody of speech: Melody and rhythm", in W. J. Hardcastle & J. Laver (eds), *The Handbook of Phonetic Sciences*, Oxford: Blackwell, 640-673.
- PALLIER C. (1994), *Rôle de la syllabe dans la perception de la parole : études attentionnelles*, Thèse de l'École des Hautes Études en Sciences Sociales.
- SCHWARTZ J. L. et al. (1997), "The dispersion-focalization theory of vowel systems", *Journal of Phonetics* 25, 255-286.
- TABAIN M. (2003a), "Effects of prosodic boundary on /aC/ sequences: Acoustic results", *The Journal of the Acoustical Society of America* 113, 516-531.
- TABAIN M. (2003b), "Effects of prosodic boundary on /aC/ sequences: Articulatory results", *The Journal of the Acoustical Society of America* 113, 2834-2849.
- TABAIN M. & PERRIER P. (2005), "Articulation and acoustics of /i/ at prosodic boundaries in French", *Journal of Phonetics* 33, 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 M.-J. Sole, P. S. Beddor & M. Ohala (eds), *Experimental Approaches to Phonology*, Oxford: Oxford University Press, 55-72.
- VAISSIÈRE J. & MICHAUD A. (2006), "Prosodic constituents in French: A data-driven approach", in I. Fónagy, Y. Kawaguchi & T. Moriguchi (eds), *Prosody and Syntax*, Amsterdam: John Benjamins, 47-64.
- WIOLAND F. (1985), *Les Structures syllabiques du français : fréquence et distribution des phonèmes consonantiques, contraintes idiomatiques dans les séquences consonantiques*, Genève/Paris : Slatkine/Champion.
- WRIGHT R. (2003), "Factors of lexical competition in vowel articulation", in J. Local, R. Ogden & R. Temple (eds), *Papers in Laboratory Phonology VI*, Cambridge: Cambridge University Press, 75-87.