

Université Sorbonne Nouvelle — Paris 3
Institut de Linguistique et Phonétique Générales et Appliquées (ILPGA)
Laboratoire de Phonétique et Phonologie, UMR 7018

Modélisation de la durée et de la modulation cepstrale dans deux types de parole : Facteurs prosodiques et syntaxiques

MEMOIRE DE MASTER 2/2 RECHERCHE MENTION SCIENCES DU LANGAGE
Parcours : Phonétique et Phonologie

Présenté par
Carole MILLOT
21800813

Devant le jury composé de
Martine Adda-Decker (Dir.Recherche CNRS)
Nicolas Audibert (MCF)
Cédric Gendrot (Professeur)

Sous la direction de M.Cédric GENDROT

Déclaration sur l'honneur

Je, soussignée, Carole MILLOT, déclare avoir rédigé ce travail sans aides extérieures ni sources autres que celles qui sont citées. Toutes les utilisations de textes préexistants, publiés ou non, y compris en version électronique, sont signalées comme telles. Ce travail n'a été soumis à aucun autre jury d'examen sous une forme identique ou similaire, que ce soit en France ou à l'étranger, à l'université ou dans une autre institution, par moi-même ou par autrui.

Le 02/06/2023

Signature :



Résumé

Ce mémoire s'inscrit dans une perspective acoustique de la phonétique, et se propose d'étudier des problématiques de prosodie articulatoire sous un angle acoustique. Plus particulièrement, nous étudions la durée et la modulation cepstrale aux frontières prosodiques de la séquence et du mot et comparons nos résultats avec la littérature existante. De nombreux travaux de recherche se sont penchés sur la production et la perception de frontières prosodiques en français comme Keating et al. (2004) ou Tabain (2003) : les mots en position finale de constituant prosodique ont la particularité d'être allongés et mieux articulés que les autres – ils sont hyperarticulés –, alors que les mots en position initiale de constituant prosodique sont principalement mieux articulés, sans allongement – ils sont renforcés.

Cependant, ces résultats n'ont été jusqu'ici obtenus que sur des corpus contrôlés en laboratoire dont les énoncés avaient pour but d'éliciter au maximum les résultats attendus. Il nous paraît donc important de vérifier ces résultats sur un corpus de parole spontanée afin de voir s'ils sont attestés dans la parole non contrôlée. En plus de ce travail, nous étudions l'impact des parties du discours et des dépendances syntaxiques sur les mesures car elles sont encore peu observées dans les recherches sur l'hyperarticulation finale et le renforcement initial. Nos principales hypothèses sont que les phénomènes d'hyperarticulation finale et de renforcement initial sont retrouvés dans les deux corpus, et dans les mots grammaticaux comme lexicaux ; ces phénomènes pouvant varier selon la dépendance syntaxique observée et sa place canonique dans l'énoncé.

Pour ce faire, nous étudions des corpus au niveau du diphone ; nous analysons la durée du diphone mais aussi une mesure encore peu usitée : la modulation cepstrale. Celle-ci se base sur les MFCC, des coefficients capables de rendre compte des phénomènes articulatoires produits par le locuteur à partir du signal acoustique. La modulation cepstrale nous permet d'obtenir le taux de changement acoustique entre deux segments : si ce changement est élevé, alors nous faisons l'hypothèse qu'au moins l'un des deux segments a été très bien articulé. Une hyperarticulation finale se traduirait alors par une durée et une modulation cepstrale élevées, alors qu'un renforcement initial ne se traduirait que par une modulation cepstrale élevée. Nous utilisons deux corpus afin de tester ces mesures : le corpus de parole spontanée NCCFr ainsi qu'un corpus de parole journalistique, Ester, que nous employons comme témoin dans cette étude – la parole journalistique étant plus lente et mieux articulée, nous devrions nous approcher des résultats trouvés dans des corpus contrôlés.

Les deux constituants prosodiques étudiés sont la séquence, située entre deux pauses – ses frontières sont la position postpausale (après une pause) et prépausale (avant une pause) –, et le mot : ses frontières sont la syllabe initiale et la syllabe finale.

Nos résultats montrent que l'hyperarticulation finale et le renforcement initial sont bien retrouvés dans les deux corpus : la mesure de modulation cepstrale a donc été efficace pour notre étude. La Figure 1 montre que la différence de durée des diphones en début et milieu de séquence n'est différente que de 1ms et ne les distingue donc pas perceptivement l'un de l'autre contrairement aux diphones en fin de séquence, mais la modulation cepstrale réussit à les départager significativement.

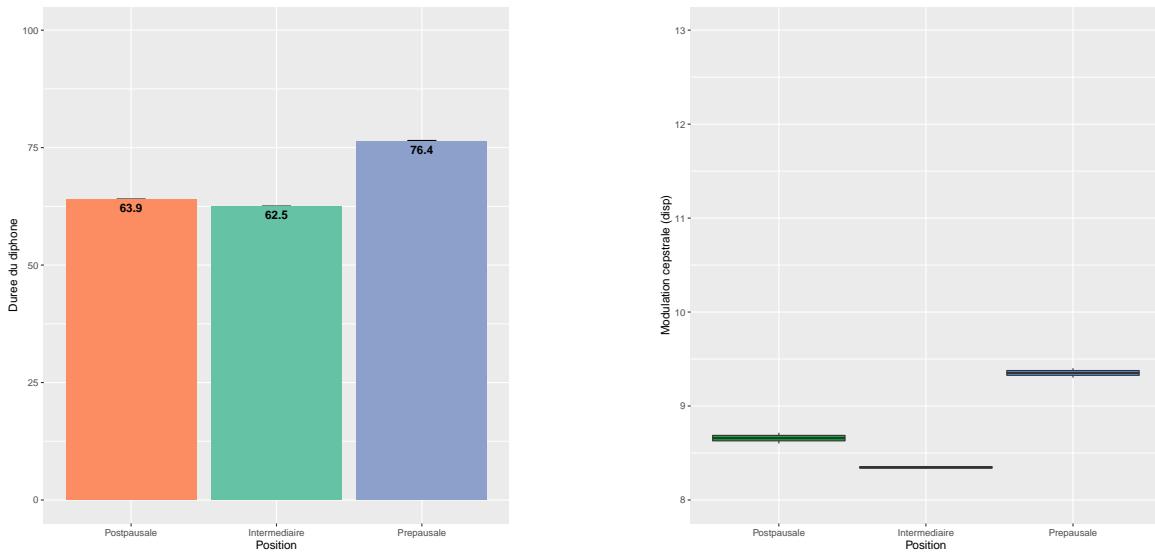


FIGURE 1 – Durée et dispersion de la modulation cepstrale du diphone selon sa position dans une séquence du corpus NCCFr : après une pause, dans la séquence, et avant une pause.

Les mots grammaticaux comme lexicaux subissent également ces phénomènes, bien que plus marqués pour les mots lexicaux. Notre étude des dépendances syntaxiques a révélé qu'il y a des différences entre les dépendances *sujet* et *complément d'objet* : cette dernière possédait des valeurs de durée et de modulation cepstrale plus importantes, que nous interprétons comme dues à une plus forte informativité du complément d'objet (plus un élément est informatif, plus il est produit clairement) car il arrive plus tard que le sujet dans l'énoncé. Enfin, nous avons observé des différences de renforcement et d'hyperarticulation entre les rimes des syllabes ouvertes et fermées : la syllabe ouverte est renforcée/hyperarticulée uniquement sur la voyelle, alors que la syllabe fermée les répartit sur la rime entière ; certaines consonnes des syllabes fermées comme les occlusives sourdes ont même une durée et une modulation cepstrale plus longue que la voyelle qui les précède.

L'hyperarticulation finale et le renforcement initial sont donc retrouvés dans la parole spontanée,

et des facteurs phonologiques et syntaxiques peuvent les faire varier. La modulation cepstrale est adaptée à cette étude, et permet de mesurer des différences d'articulation de segments à partir d'un simple signal acoustique de parole, ce qui la rend utile dans de nombreux contextes d'analyse.

Remerciements

De nombreuses personnes ont contribué à la création et au bon déroulé de ce mémoire, et je souhaite les en remercier ici.

Tout d'abord, merci à M.Cédric Gendrot pour avoir accepté de m'encadrer à nouveau pour ce mémoire, vos explications concernant les données et les tests statistiques ont toujours été très claires et vous avez su m'orienter vers les bonnes références. Merci également pour votre aide sur certains scripts et détails techniques de Praat.

Je tiens aussi à remercier M.Kim Gerdes pour son aide précieuse concernant les annotations en parties du discours et en dépendances syntaxiques, et M.Leonardo Lancia pour le script calculant la modulation cepstrale et ses explications à son sujet : mon mémoire sous sa forme actuelle n'aurait pas existé sans eux. Merci également à Marine Courtin pour ses explications sur les dépendances syntaxiques et les parties du discours.

Enfin, merci à Nicolas Audibert et Martine Adda-Decker d'avoir accepté le rôle de jury pour mon oral de mémoire, et à Jane Stuart-Smith et Claire Pillot-Loiseau d'y assister.

Table des matières

1	Introduction	1
2	État de l'art	3
2.1	Prosodie	3
2.1.1	Accents primaire et secondaire	3
2.1.2	Constituants prosodiques	4
2.1.3	Prosodie et communication	5
2.1.4	Articulation et frontières prosodiques	5
2.1.5	Syntaxe	9
2.2	Coarticulation	10
2.2.1	Définition	10
2.2.2	Coarticulation et communication	11
2.2.3	Hyperarticulation et réduction	12
2.3	Phonologie articulatoire	13
2.3.1	Théorie motrice de la parole	13
2.4	Modulation cepstrale	15
3	Objectifs et hypothèses	21
4	Méthode	23
4.1	Corpus et mesures utilisés	23
4.1.1	Corpus	23
4.1.2	Modulation cepstrale	24
4.2	Obtention des données	24
4.2.1	Durée et frontières prosodiques	26
4.2.2	Parties du discours et dépendances syntaxiques	28
4.2.3	Calcul de la modulation cepstrale	29
5	Résultats	35
5.1	Durée et modulation cepstrale aux frontières prosodiques	35
5.1.1	Analyse d'un sous-ensemble de diphones	43

5.1.2	Impact des hésitations	46
5.2	Durée et modulation cepstrale selon les parties du discours	47
5.2.1	Analyse d'un sous-ensemble de diphones	52
5.2.2	Étude de quelques parties du discours	55
5.3	Impact du type de syllabe	57
5.3.1	Impact de la consonne	60
5.3.2	Impact du nombre de syllabes	61
5.4	Impact des dépendances syntaxiques	62
5.4.1	Sujets et compléments d'objet	62
5.4.2	Dépendance syntaxique des déterminants	71
6	Discussion	74
6.1	Mesures de durée et modulation cepstrale	74
6.2	Résultats selon les données syntaxiques	77
6.2.1	Parties du discours	77
6.2.2	Dépendances syntaxiques	78
6.3	Phonèmes et types de syllabe	79
6.4	Autres facteurs de variation	81
6.4.1	Fréquence fondamentale	81
6.4.2	Mots monosyllabiques ou entre deux pauses	82
6.4.3	Modulation cepstrale et locuteurs	83
6.5	Critiques	84
7	Conclusion	89
8	Bibliographie	91
9	Annexe	96

1 Introduction

Les analyses phonétiques sur des corpus de parole spontanée ont longtemps été peu nombreuses, de par plusieurs raisons :

- la qualité des microphones hors chambre sourde ou pièce spécialisée n'était pas assez bonne pour obtenir des résultats exploitables ;
- le corpus de parole spontanée ne possède pas d'énoncé "toutes choses égales par ailleurs" où l'on pourrait étudier l'environnement d'un segment dans plusieurs conditions totalement contrôlées : cela n'est possible qu'avec des corpus élaborés à l'avance. La seule possibilité est de compenser ce manque de contrôle par une quantité de données importantes, mais les systèmes de stockage n'étaient alors pas assez grands.

Cependant, la qualité des microphones s'est depuis grandement améliorée – à tel point qu'un microphone de téléphone portable permet des enregistrements de qualité suffisante pour une étude sur la parole (tant que celle-ci n'est pas être fine et ne s'intéresse pas aux formants ou à la fréquence fondamentale par exemple) – et les systèmes de stockage supportent à présent une quantité de données très importante.

De ce fait, utiliser des corpus de parole spontanée est maintenant possible, et nombre d'études s'attèlent à analyser leurs données de façon quantitative. Par exemple, Gendrot and Audibert (2019) comparent la distinction entre /e/ et /ɛ/ dans un corpus de parole spontanée et un corpus journalistique, non contrôlé mais dont les locuteurs emploient un français plus normé. Ainsi, ils peuvent observer si le processus de fusion entre ces deux phonèmes a déjà lieu dans le corpus spontané (dans lequel il est plus probable de l'observer, car les changements phonétiques apparaissent plus volontiers en parole spontanée), et s'il a commencé à apparaître en parole journalistique également.

Les corpus de parole spontanée permettent également de vérifier si des observations effectuées dans des corpus contrôlés sont vérifiées dans la parole spontanée, et donc à quel point elles sont représentatives de celle-ci.

C'est ce que nous nous proposons de faire dans ce mémoire consacré à l'étude de l'hyperarticulation finale et du renforcement initial en français : après avoir étudié les travaux déjà effectués dans ce domaine, nous présenterons nos hypothèses, ainsi que les mesures et les corpus (spontané et journalistique) que nous avons choisis pour vérifier la présence de ces phénomènes prosodiques

en parole spontanée. Nous utilisons notamment la mesure de modulation cepstrale afin d'observer la présence du renforcement initial. Enfin, après la description des résultats, nous discuterons de ceux-ci et de leurs implications, ainsi que le travail encore possible sur ce sujet avant de conclure sur nos hypothèses.

Tous les scripts nommés sont accessibles à l'adresse https://github.com/C-Millot/memoire_m2.

2 État de l'art

La variation est inhérente à la parole et est présente dans tous les aspects de la vie d'un locuteur ; différentes théories phonologiques la modélisent, comme la phonologie articulatoire et la phonologie acoustique. Les sources de cette variation sont multiples, et peuvent provenir de contraintes linguistiques, pathologiques ou encore culturelles. Des facteurs de variation inter et intra locuteurs tels que des particularités physionomiques, l'âge ou certaines activités sont également vecteurs de variations. Par exemple, les plis vocaux qui permettent la production de phonation au sein de phonèmes peuvent être affectés par des nodules s'ils sont trop sollicités, se ferment moins bien avec l'âge et possèdent souvent une petite ouverture chez les femmes.

La prosodie de la langue est un autre facteur de variation, modifiant les segments de façon prédéfinie selon un ensemble de règles propres à une langue. La syntaxe peut également jouer un rôle dans l'élicitation de ces variations.

2.1 Prosodie

2.1.1 Accents primaire et secondaire

La prosodie est un ensemble de règles s'appliquant à des segments dans un énoncé, et est composée de plusieurs facteurs de performance suprasegmentaux comme l'accentuation, l'intonation ou le rythme. Les corrélats acoustiques de la prosodie en français incluent des modulations de la fréquence fondamentale, du type de phonation, de l'intensité et la durée ([Vaissière and Michaud, 2006](#)). Ces corrélats sont encore étudiés actuellement, mais ils ont été analysés dès les années soixante : Delattre (1963) analysait les contours prosodiques du français et d'autres langues dans des énoncés exprimant la continuation (signalant que l'énoncé n'est pas fini) ou la finalité (la fin de l'énoncé) ([Figure 2](#)). Les études postérieures ont servi à hiérarchiser et étendre les connaissances sur ces sujets.

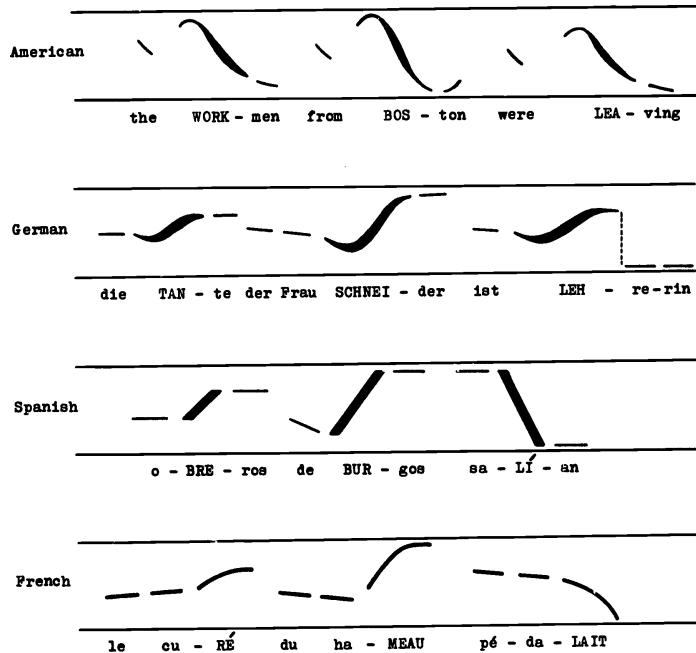


FIGURE 2 – Comparaison des principaux contours intonatifs américain, allemand, espagnol et français pour les expressions de la continuation et de la finalité, d'après Delattre (1963).

La prosodie de la langue a un impact important sur la variation des phonèmes produits. Welby (2003) référence les différentes caractéristiques prosodiques du français dans l'état de l'art de sa thèse centrée sur les montées précoces (*early rises*) de fréquence fondamentale dans les énoncés en français. L'accent primaire du français provoque une durée plus longue de la dernière syllabe d'un syntagme prosodique ainsi qu'une intensité plus importante. De plus, si la syllabe n'est pas syllabe finale de l'énoncé, on observe une montée de fréquence fondamentale f_0 . Le français possède aussi un accent secondaire, au début du syntagme : Welby 2003 et références incluses ne font pas état d'une durée ou une intensité systématiquement plus importantes, mais d'une petite montée de f_0 .

Tous ces paramètres s'appliquent à divers endroits des productions orales, et sont hiérarchisés dans différents niveaux prosodiques.

2.1.2 Constituants prosodiques

Il existe différents niveaux prosodiques, comme mis en évidence par Nespor and Vogel (1983) (liste non exhaustive) :

- l'énoncé, ou séquence ;
- le syntagme intonatif ;
- le syntagme phonologique, ou syntagme accentuel ;
- le mot phonologique ;

- le pied ;
- la syllabe ;
- la rime.

Dans chaque constituant, un segment peut être en position initiale, finale, ou au milieu de celui-ci.

Cependant, toutes les langues ne distinguent pas ces différents niveaux, et ils ne revêtent pas tous la même importance : dans les études que nous évoquerons ci-après, les auteurs rapportent régulièrement que ces niveaux, élaborés au fil de théories prosodiques, ne sont pas toujours reconnus par les locuteurs dans Keating et al. (2004), les auteurs rapportent que les sujets ne différencient pas l'énoncé du syntagme intonatif par exemple, alors que des niveaux plus espacés comme l'énoncé et le mot sont correctement différenciés. Dans Tabain (2003), les syntagmes intonatifs et accentuels sont traités comme un seul constituant prosodique.

Ainsi, il est important de sélectionner adéquatement les niveaux prosodiques utilisés dans notre étude : nous comparerons le niveau de la séquence et celui du mot, qui sont régulièrement décrits comme correctement distingués.

2.1.3 Prosodie et communication

La prosodie a une dimension portée vers l'interlocuteur : par exemple, les montées aux deux extrémités des syntagmes peuvent avoir pour but de délimiter les frontières syntaxiques afin de les rendre plus compréhensibles pour les auditeurs, et pour éviter que la parole ne soit qu'un long enchaînement de phonèmes non accentués sans rythme (Astésano et al., 2003). Elles peuvent également servir à la désambiguïsation de certains syntagmes (Fónagy, 1980) ; par exemple, dans « les ballons et les trophées noirs », c'est la montée secondaire qui aide à comprendre si uniquement les trophées sont noirs (il y a alors montée sur « trophées »), ou si les ballons le sont aussi (il y a alors montée uniquement sur « ballons »).

2.1.4 Articulation et frontières prosodiques

La durée, l'intensité et la fréquence fondamentale ne sont pas les seuls paramètres distinguant les frontières de constituants prosodiques en français : plusieurs études montrent que l'articulation des segments est également un indice important.

Keating et al. (2004) étudient le contact palato-lingual et sa durée de constriction selon la position dans la frontière de différents niveaux prosodiques ; l'étude se place dans le cadre des monosyllabes de forme CVC (la consonne étant /n/ ou /t/ et la voyelle /a/), et analyse la production des consonnes selon si elles sont à la fin du constituant prosodique, ou au début. Les niveaux prosodiques étudiés sont l'énoncé (*utterance*), le syntagme intonatif (*intonational phrase*), le syntagme plus court (*smaller phrase*) – syntagme accentuel –, le mot et la syllabe. Cette étude est réalisée à l'aide d'un électropalatographe afin de mesurer le taux de contact entre la langue et le palais, sur quatre langues : l'anglais, le français, le coréen et le taiwanais. Les phrases produites sont construites afin de mettre en exergue les frontières des différents niveaux prosodiques, par exemple :

« Paul aime Tata. Nadia les protège en secret. » (énoncé) (1)

« La pauvre Tata, Nadia et Paul n'arriveront que demain. » (syntagme intonatif) (2)

Plusieurs résultats sont constatés par l'article : tous les locuteurs différencient plusieurs niveaux de constituants prosodiques, mais ce n'est pas pour autant que tous sont également reconnus : l'énoncé et le mot sont systématiquement différenciés, mais il est plus difficile de différencier entre énoncé et syntagme intonatif par exemple. La position « initiale de mot » n'est pas toujours distinguée des positions « initiale de syllabe » et « initiale de syntagme court ».

La comparaison des langues est intéressante puisqu'elle met en valeur une qualité du français par rapport au coréen : alors qu'en finale de frontière la consonne est hyperarticulée pour les deux langues, en début de frontière elle n'est que renforcée pour le français : la durée, deuxième élément de l'hyperarticulation, n'est un élément déterminant que pour le coréen en début de frontière.

En ce qui concerne l'articulation des voyelles selon les frontières prosodiques du français, Tabain (2003) étudie la production articulatoire des /aC/ (avec la voyelle en position finale de constituant) selon les frontières suivantes : l'énoncé, le syntagme intonatif, le syntagme accentuel et le mot. L'articulatographe électromagnétique EMA est utilisé ; il s'agit d'une instrumentation utilisant un champ électromagnétique et des capteurs placés à différents endroits de la bouche (langue, lèvres, mâchoire) afin d'en modéliser les mouvements en trois dimensions sur ordinateur (Rebernig et al., 2021).

Dans cette expérience, les capteurs étaient positionnés sur la mâchoire, l'apex de la langue, le corps de la langue et la lèvre supérieure. Les phrases utilisées sont proches de celles de Keating

et al. (2004) :

« Paul aime Tata. Baba les protège en secret. » (3)

pour le niveau de l'énoncé, avec la consonne suivante /b/. La voyelle finale de constituant prosodique est toujours /a/, mais la consonne qui la suit peut varier parmi /b d g f s ſ/. Ainsi, les productions des locuteurs sont très contrôlées, et un locuteur natif du français connaissant le système d'annotation ToBI vérifie que les phrases ont été produites avec l'intonation attendue.

Afin d'analyser les données, Tabain récolte la durée des voyelles, des consonnes, et leur pic de Déplacement – pour les voyelles, il s'agit du minimum dans l'axe des ordonnées du capteur du dos de la langue et de la mâchoire. Pour les consonnes, il s'agit de la valeur la plus haute entre la fermeture et l'aperture d'un des capteurs suivants : l'apex de la langue dans le cas des consonnes dentales /d ſ/, le corps de la langue dans le cas de la consonne vélaire /g/ et la mâchoire dans le cas des consonnes labiales /b f/. De plus, sont calculés le chemin entre chaque capture du corps de la langue (en utilisant la somme des distances euclidiennes entre chaque échantillon), et l'ampleur et la durée et le pic de vélocité de l'ouverture et de fermeture de la bouche.

Les résultats montrent qu'encore une fois, les syntagmes intonatifs et accentuels ne sont pas systématiquement différenciés et sont traités comme une seule classe ; l'énoncé et le mot sont, eux, systématiquement différenciés. Un autre résultat intéressant est que la durée des phonèmes est plus grande plus la frontière prosodique est d'un niveau élevé, mais « la voyelle prend un pourcentage plus important de la syllabe plus la frontière prosodique est importante. Cela concorde avec un effet plus important de la durée de la voyelle que celle de la consonne de la frontière prosodique. » (Tabain, 2003, p.7 [traduit de l'anglais]). L'ampleur du mouvement du corps de la langue est aussi plus important à mesure que le niveau prosodique est élevé. Malheureusement, l'article ne distingue pas la fin d'une frontière prosodique et le début d'une autre, et ne regarde que la transition entre les deux.

L'article Tabain and Perrier (2005) étend cette expérience à la voyelle /i/, en ajoutant le niveau prosodique de la syllabe aux autres précédemment étudiés. Les résultats obtenus avec la voyelle /a/ ne sont pas les mêmes : bien que l'effet de la durée est toujours présent pour différencier les frontières prosodiques, l'ampleur des mouvements supralaryngés n'est pas significativement différente selon le niveau prosodique. Tabain et Perrier émettent l'hypothèse que, plutôt qu'hyperarticuler le /i/ comme le /a/, les locuteurs préfèrent appuyer sur le F3 élevé de cette voyelle afin de renfor-

cer ses caractéristiques acoustiques, et ce « que cela entraîne un avancement, une hauteur ou un renforcement du corps de la langue plus important » (Tabain and Perrier, 2005, p.24 [traduit de l'anglais]) – il n'y a pas d'articulation particulière qui entraîne ce renforcement, ce qui expliquerait que rien ne soit trouvé dans les données.

En alliant prosodie et coarticulation, Gendrot et al. (2016) étudient la détection de constituants prosodiques de par leur hypo ou hyperarticulation : les voyelles en frontière de constituant prosodique seront en effet théoriquement mieux articulées que les autres. Pour cela, ils utilisent un corpus journalistique duquel ils extraient les voyelles des syllabes et leurs trois premiers formants ; le but est d'analyser la dispersion des voyelles dans l'espace vocalique – si celle-ci s'éloigne du centre acoustique, alors elle sera considérée comme renforcée. La Figure 3 montre les résultats obtenus ; les résultats pour les voyelles finales sont similaires. De plus, ils montrent que les constituants prosodiques sont indépendants de leur durée : la durée du dernier phonème, selon le niveau prosodique en faisant varier son nombre de syllabes, ne change pas.

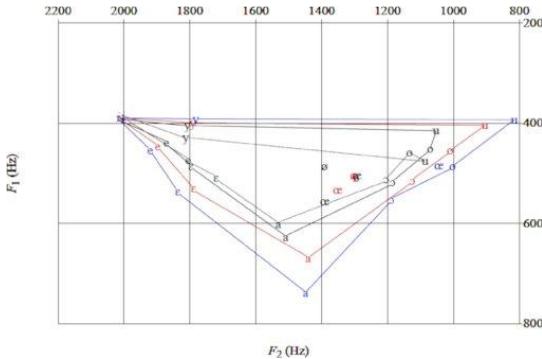


FIGURE 3 – Répartition des voyelles en position initiale dans un espace vocalique en deux dimensions selon le constituant syntaxique auquel elles appartiennent, d'après Gendrot et al. (2016).

Toutes ces études utilisent des corpus issus de parole de laboratoire : les énoncés, interlocuteurs et lieu d'enregistrement sont contrôlés pour obtenir les mesures les plus fines possibles, mais cela leur coûte la spontanéité qui est une des caractéristiques de la parole telle qu'elle est normalement produite. Dans notre étude, il sera donc fructueux d'utiliser un corpus de parole spontanée afin de vérifier si les résultats de durée et d'articulation aux frontières prosodiques peuvent être retrouvés.

Gahl et al. (2012) est une étude qui utilise un corpus de parole spontanée déjà existant : les auteurs effectuent des analyses acoustiques sur ses données, en étudiant la durée des mots selon leur fréquence, leur partie du discours, et leur probabilité de bigramme : les bigrammes modélisent la probabilité qu'un mot apparaisse avant ou après un autre mot. Leurs principaux résultats, sur la

Figure 4, montrent que la probabilité de bigramme prédit la durée d'un mot de façon semblable à la fréquence de celui-ci.

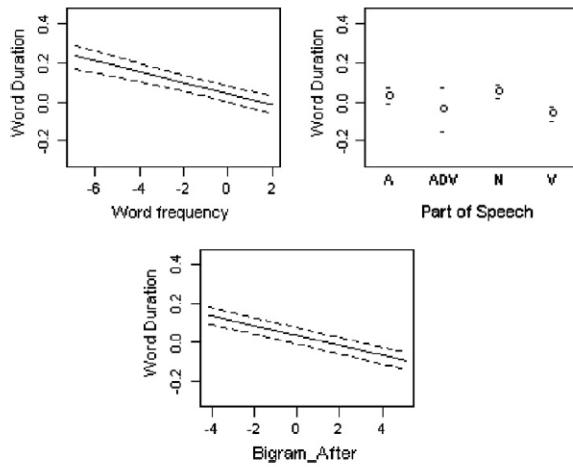


FIGURE 4 – Corrélations entre la durée d'un mot et sa fréquence, sa partie du discours et sa probabilité de bigramme, d'après Gahl et al. (2012).

2.1.5 Syntaxe

La syntaxe est un autre facteur de variation au sein de la parole : d'une manière similaire à la prosodie, la syntaxe hiérarchise les éléments de l'énoncé avec des relations de dépendance – le complément d'objet et le sujet sont subordonnés à la racine de l'énoncé, le plus souvent un verbe, les modificateurs d'un constituant sont subordonnés à celui-ci, *etc...* Le lien entre prosodie et syntaxe est discuté, néanmoins il est constaté que « la connaissance de la syntaxe d'un énoncé offre une base suffisante pour l'émulation d'un contour de fréquence fondamentale acceptable » (Vaissière and Michaud, 2006, p.2 [traduit de l'anglais]).

Il existe différentes théories syntaxiques considérant les propriétés syntaxiques comme plus ou moins liées au sens de l'énoncé : la théorie Sens-Texte considère une proximité entre relations syntaxiques et relations sémantiques, tentant de produire l'ensemble des énoncés pouvant exprimer un même sens Marengo, 2014 – elle postule qu'il existe des relations de dépendance entre les représentations syntaxique et sémantique de l'énoncé.

À l'inverse, Gnamian (2018) poursuit une vision chomskyanne de la syntaxe, qui stipule qu'elle est autonome par rapport à la sémantique – le sens ne serait donc plus premier, mais plutôt une interprétation des formes syntaxiques. Dans le cadre d'une étude conciliant parole (la prosodie est notamment basée sur le sens de l'énoncé) et syntaxe, nous nous plaçons dans le cadre d'une syntaxe

régie par le sens et non autonome.

Toutefois, le rapport entre liens de dépendance et coarticulation est encore peu étudié. Certaines études laissent pourtant à penser que l'articulation a un rôle à jouer dans la perception de mots de classe ouverte (mots lexicaux) contre ceux de classe fermée (mots grammaticaux) ((Goodman et al., 1990)). En effet, selon l'hypothèse que les mots lexicaux sont plus informatifs que les mots grammaticaux, alors ils ont plus de chances d'être mieux articulés car les mots les plus informatifs devraient être produits plus clairement afin d'être correctement compris. Étant donné que peu d'études ont analysé les liens entre syntaxe de dépendance et coarticulation, nous nous proposons d'observer ces liens dans notre mémoire.

Au sujet des parties du discours (catégories grammaticales), Gahl et al. (2012), dont l'étude portait sur la réduction des mots fréquents, avaient opéré une séparation entre les verbes, les noms, les adjectifs et les adverbes de leur corpus ; celle-ci n'avait pas apporté de différences significatives dans la durée des différentes catégories grammaticales évoquées. Cependant, nous souhaitons noter que toutes ces catégories sont des mots lexicaux, et qu'aucun test n'a été effectué avec des mots grammaticaux comme des pronoms ou des déterminants.

2.2 Coarticulation

Nous avons vu que l'articulation d'un segment avait un impact sur la perception et la production des frontières prosodiques. Ainsi, nous étudions à présent le concept de coarticulation.

2.2.1 Définition

Parmi les phénomènes provoquant de la variation dans la parole, la coarticulation est l'un des plus importants. Elle est définie comme un processus qui transmet les propriétés d'un segment, causant un chevauchement dans le temps de celles-ci avec les indices des segments adjacents (Zellou, 2022). Elle est due à l'inertie des articulateurs, et peut être calculée comme la différence entre deux segments, en termes de formants par exemple.

Malgré les causes physiologiques de la coarticulation, celle-ci n'est pas uniquement régie par nos contraintes physiques, et la phonologie de la langue joue un rôle important : le français par exemple, possède le trait de nasalité pour les consonnes et les voyelles alors que l'anglais n'a que des consonnes nasales ; ainsi, alors que les locuteurs du français contrôlent davantage leur velum afin que le flux d'air dans la cavité nasale cesse lorsque des voyelles orales doivent être produites en vicinité de

consonnes nasales, les locuteurs de l'anglais n'ont pas à y prêter une aussi grande attention car la nasalité vocalique n'est pas un trait contrastif dans cette langue ; la voyelle précédant une consonne nasale est donc souvent nasalisée [Cohn \(1993\)](#) ; [Chen \(1997\)](#).

Il faut néanmoins noter que, comme montré dans [Zellou \(2017\)](#) sur l'anglais, il existe des variations individuelles dans la coarticulation : certains locuteurs coarticulent plus que d'autres, et cela influe sur leur perception de la coarticulation. Notamment, un sujet qui hypoarticule attribue la cause de la nasalité qu'il perçoit à la voyelle nasalisée et non pas la consonne nasale qui a déclenché celle-ci, alors qu'un sujet hyperarticulant attribue la cause à la consonne, en percevant moins la nasalisation de la voyelle.

2.2.2 Coarticulation et communication

La coarticulation n'est pas seulement un phénomène phonétique dû à la manière dont nous produisons des sons, mais fait partie intégrante de la communication : [Scarborough \(2003\)](#) montre que le degré de coarticulation d'un mot produit a un lien avec sa compréhension par l'interlocuteur. Elle fait produire à des participants des mots plus ou moins susceptibles d'être sujets à coarticulation (nasale ou harmonie vocalique). Ensuite, sont déterminés parmi ces mots ceux qui ont une fréquence lexicale plus ou moins élevée et leur nombre de voisins – des mots différent de seulement un phonème par addition, délétion ou substitution.

Les résultats montrent que « les mots plus à même d'être lexicalement confondus, qui ont une fréquence faible par rapport à leurs voisins lexicaux, sont prononcés avec un plus grand degré de coarticulation que les mots moins confondus [...] les auditeurs ont été capables d'identifier des mots coarticulés plus vite que ceux dans lesquels la coarticulation avait été éliminée » ([Scarborough, 2003](#), p.10 [traduit de l'anglais]) : la coarticulation aide à clarifier la parole pour l'interlocuteur, car en anticipant les phonèmes suivants, l'oreille perçoit d'autant plus d'indices lui permettant de comprendre ce qui est produit.

La coarticulation a aussi une importance lors de la récupération des informations lexicales et phonologiques d'un mot dans la mémoire de l'auditeur ([Fink and Goldrick, 2015](#)) : les modèles de la mémoire dits lexicalistes prennent en compte le fait que des locuteurs perçoivent au cours de leur vie de nombreux exemples de mots fréquents réduits, et de nombreux exemples de mots non fréquents hyperarticulés : la mémoire enregistrerait ces détails avec les mots, ce qui influencerait la façon dont ils seront produits par le locuteur. Il faut cependant noter que ces modèles de mémoire

n'expliquent pas tous les phénomènes liés au lexique que nous percevons, notamment les variations liées à un contexte linguistique ou paralinguistique particulier.

2.2.3 Hyperarticulation et réduction

La notion de coarticulation est donc liée à celle de réduction et d'hyperarticulation : les mots les moins fréquents du lexique et ayant de nombreux voisins phonologiques sont produits avec des voyelles hyperarticulées, dans un espace vocalique plus grand que les mots plus courants, comme le montrent Gahl et al. (2012). Les transitions entre phonèmes sont alors plus saillantes et on y remarque moins de coarticulation. Les mots fréquents, au contraire, observent un phénomène de réduction : les cibles ne sont pas atteintes articulatoirement et le mot est produit plus rapidement, avec davantage de coarticulation. Ce continuum entre hyper et hypoarticulation a été constaté dès 1990 par Lindblom (1990) dans son *Hypo & Hyper articulation model* : la parole doit constamment réaliser un compromis, en visant la minimisation de l'effort tout en restant intelligible.

De nombreuses études induisent une hyperarticulation contrôlée ; par exemple Munson and Solomon (2016) utilisent des *bite blocks*, des blocs que les participants doivent mordre et qui les empêche de fermer totalement la mâchoire (ici, 1cm sépare les molaires supérieures et inférieures) – ceux-ci sont donc contraints d'hyperarticuler afin d'atteindre les cibles articulatoires des phonèmes. De plus, les résultats montrent qu'avec le *bite block* les participants ne réduisent pas les mots même quand ceux-ci sont plus courants, alors qu'ils les réduisent dans des conditions normales de production : les auteurs suggèrent que la présence de blocs perturbe tellement l'articulation que celle-ci ne peut plus être la même que celle qu'ils utilisent habituellement.

La réduction peut être due à de nombreux paramètres ; nous avons cité plus haut le nombre de voisins phonologiques et la fréquence des mots, mais d'autres facteurs peuvent entrer en compte. Par exemple, Audibert et al. (2015) montrent que la réduction de la voyelle est liée à la durée de celle-ci et au style de parole utilisé : ils utilisent un corpus journalistique (Ester), un corpus de lecture (BREF) et un corpus de parole spontanée (NCCFr) afin de comparer ces styles de parole, et des mesures sur l'espace de vocalique des voyelles analysées (/i,e,a,o,u/) comme l'étendue des deux premiers formants des voyelles, leur dispersion dans l'espace, leur distance par rapport au centre de l'espace et si elles conservent un contraste phonologique entre elles ou non. Les résultats montrent que les voyelles les plus courtes, et celles produites dans le corpus de parole spontanée, sont plus sujettes à réduction que les autres. Les deux autres corpus, dits de *clear speech*, possèdent

moins de voyelles réduites. Les corrélats acoustiques de la réduction sont une plus grande dispersion intra-classe vocalique et une plus grande centralisation des voyelles, ainsi qu'une perte de contraste phonologique entre certains exemplaires de voyelles.

Wu and Adda-Decker (2021) observent que certaines consonnes sont particulièrement à risque d'être réduites : les liquides, les semi-consonnes et /v/. Cela est expliqué comme étant dû à leurs durées intrinsèques plus courtes que les autres consonnes, et leur ressemblance acoustique plus prononcée à des voyelles. Concernant les voyelles, les voyelles orales arrondies sont celles résistant le moins à la réduction alors que les nasales lui résistent le mieux – cela est expliqué par leur durée intrinsèque plus longue et le trait [+nasal] qui leur conférerait un renforcement intrinsèque.

2.3 Phonologie articulatoire

La coarticulation fait fondamentalement partie des objets d'étude de la phonologie articulatoire, qui vise à rendre compte des phénomènes de production de la parole. Cette modélisation de la parole peut donc nous en apprendre davantage sur elle.

2.3.1 Théorie motrice de la parole

La phonologie articulatoire, décrite dans Browman and Goldstein (1992), introduit le geste comme unité constituant la parole : ce sont les événements articulatoires se déroulant lors de la production de parole. Ces gestes peuvent donner lieu à des corrélats acoustiques, ou des traits phonologiques, mais sont bien distincts de ceux-ci ; causés directement par les articulateurs de la parole, ils sont dynamiques et s'influencent les uns les autres – souvent, ils fonctionnent en groupe coordonné. Par exemple, l'ouverture des lèvres est influencée par la lèvre supérieure, la lèvre inférieure et la mâchoire. Ces gestes sont modélisés sur une partition comme sur la Figure 5 qui représente les gestes impliqués dans la production du mot anglais <palm>.

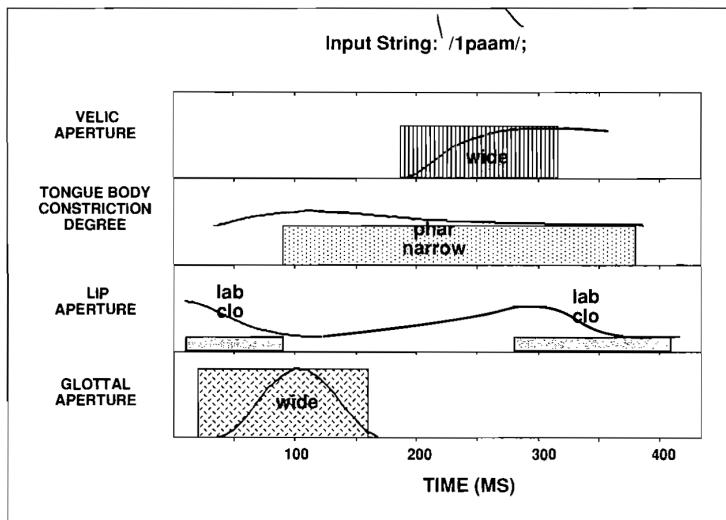


FIGURE 5 – Partition gestuelle pour le mot <palm> /paam/, d'après Browman and Goldstein (1992).

La caractéristique gestuelle de la parole est intéressante par bien des aspects. Elle prévoit les sons qui ressembleront à des sons de la parole pour des langues, même si elles ne les comportent pas : « même si une langue n'utilise pas de constrictions orale critiques [fricatives], le degré de ces constrictions existe entre une constriction complète et une constriction étroite et devrait donc être englobé par n'importe quelle langue qui utilise ces deux constrictions. » (Best, 1995, p.21 [traduit de l'anglais]).

Du point de vue de la perception, ce modèle prévoit que, plutôt que le signal acoustique de la parole, nous percevons les gestes sous-jacents – c'est-à-dire la source du signal. Le Réalisme direct fait même l'hypothèse que ce que nous percevons est directement l'object perceptuel et non sa représentation acoustique ; cela infère un module cérébral capable d'appréhender la nature directe de ce qui est transmis. Ce module est aussi proposé par la théorie motrice de la parole, l'un des modèles les plus proéminents de la phonologie articulatoire, même si celle-là ne part pas du principe que nous percevons directement l'objet perceptuel. Cependant, Galantucci et al. (2006) ont montré que les principales revendications de cette théorie n'étaient pas confirmées à ce jour : que les auditeurs perçoivent les gestes articulatoires n'a pas été totalement prouvé, et que la parole est traitée de façon spéciale par le cerveau non plus.

Un apport à la phonologie articulatoire intéressant pour notre étude est le concept de *π-gesture* (Byrd and Saltzman, 2003) : son but est d'inclure les suprasegments prosodiques en tant que gestes à part entière dans les partitions gestuelles afin que la phonologie articulatoire puisse précisément

rendre compte des phénomènes prosodiques. En effet, la prosodie détermine en partie l'articulation du mot et sa coordination avec les mots voisins, et ce modèle ne saurait expliquer toutes les variations dans la parole sans en tenir compte.

Pour ce faire, au-dessus des gestes articulatoires est érigé le π -gesture (de *prosodic gesture*) : celui-ci, situé aux frontières prosodiques, ralentit le temps des gestes produits au même moment que lui lorsqu'il apparaît (voir Figure 6). Comme les gestes sont ralenti, ils se chevauchent aussi moins et il y a donc moins de coarticulation : on retrouve les paramètres de la durée plus importante et de la meilleure articulation déjà décrits en prosodie. Si la frontière prosodique est importante dans la langue, celle-ci provoquera un π -gesture d'autant plus important qui provoque davantage l'hyperarticulation des gestes impliqués.

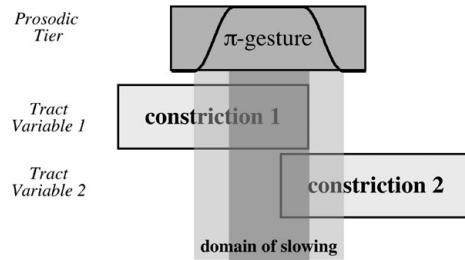


FIGURE 6 – Partition gestuelle pour deux gestes pendant une frontière prosodique instantiée via un π -gesture, d'après Byrd and Saltzman (2003).

Il faut ajouter que le π -gesture n'est pas symétrique dans son influence : plus on s'éloigne de la bordure de l'énoncé, plus le π -gesture est actif et ralentit les gestes impliqués.

Les apports de la phonologie articulatoire sont importants afin de pouvoir expliquer les résultats acoustiques que nous allons obtenir dans notre analyse. Mais quel est le lien entre les représentations articulatoires et acoustiques de la parole ?

2.4 Modulation cepstrale

Les travaux de Louis Goldstein, ancrés dans la phonologie articulatoire, sont importants dans l'étude des relations entre représentations articulatoires et acoustiques : les interactions entre production et perception de la parole (montrées dans des expériences de production en écoutant un retour de soi altéré par exemple) requièrent une collaboration entre les informations perçues auditivement – acoustiques – et la planification articulatoire de la production de parole. Nous verrons que la compréhension de ces interactions passe par l'utilisation des MFCC donnant des indices sur

l'articulation alors effectuée par le locuteur. Cette mesure est plus efficace que la durée, qui n'est qu'une conséquence des gestes articulatoires alors que les MFCC capturent directement l'information de la forme du conduit vocal (Slis et al., 2021).

Afin d'établir un rapprochement entre phonologies articulatoire et acoustique, Goldstein (2019) met en lumière les interactions entre représentations articulatoires et acoustiques du signal de parole par le locuteur, et propose d'expliquer cette interaction sensori-motrice par la perception des modulations temporelles, c'est-à-dire des changements dans les représentations phonétiques au cours du temps.

Pour ce faire, il étudie les données d'un corpus répertoriant des enregistrements aux rayons X de sujets, notamment au niveau de la langue – lèvre haute (UL), lèvre basse (LL), incisive basse (LI), et quatre marqueurs de la langue (de l'apex au dos de la langue : T1, T2, T3, T4). L'étude de Goldstein se base sur le fait que le signal de parole change de façon structurée temporellement, et non pas de façon continue. Il obtient les fonctions de modulation articulatoires et acoustiques des énoncés du corpus, c'est-à-dire des fonctions mesurant le changement acoustique et le changement articulatoire, et conjecture qu'elles devraient toutes former des pulsations, exhibant des zones de changement rapide et d'autres de stagnation : ces pulsations seront liées à la structure syllabique de l'énoncé — dans une syllabe CV, une pulsation est attendue dans la zone de changement entre la fin de la consonne et le début de la voyelle.

Le calcul de la fonction de modulation articulatoire est fait en obtenant la somme de la vélocité des marqueurs linguaux multipliés par leur nombre de dimensions (ici deux) : plus la vélocité est importante dans l'espace buccal, plus le changement articulatoire sera grand. La fonction de modulation acoustique se base sur les MFCC (Mel Frequency Cepstral Coefficients), et est généralement calculée en suivant les étapes de la Figure 7 : elle est appelée *modulation cepstrale*. Voici une brève description des étapes de calcul d'un MFCC d'après Muda et al. (2010), visibles sur la Figure 7 :

- 1) L'étape de *pre-emphasis* augmente l'énergie dans les hautes fréquences du signal afin d'émuler l'audition humaine.
- 2) Le *framing* segmente le signal en petites portions de 20 à 40ms généralement.
- 3) On applique à ces segments une fenêtre de Hamming (voir Figure 8), qui a la particularité de prendre en compte le segment précédent lors de l'extraction des paramètres du signal.
- 4) La Transformée de Fourier Rapide permet de décomposer le signal en somme de sons purs, permettant ainsi au signal de passer d'un domaine temporel à un domaine fréquentiel.

- 5) Cette étape permet au signal d'avoir une échelle en Mel afin d'être plus proche de la perception humaine du signal.
- 6) La transformée en cosinus discrète permet d'achever la création du MFCC en repassant dans un domaine temporel ; le MFCC est alors un vecteur contenant les caractéristiques acoustiques du signal.
- 7) L'étape *Delta Energy and Delta Spectrum* fournit un moyen d'encoder le changement des caractéristiques cepstrales au cours du temps, ce qui est utile dans le cadre de l'article de Goldstein : les features obtenues représentent les changements entre les fenêtres, on trouve généralement treize caractéristiques delta liées à la vitesse, et trente-neuf liées à l'accélération.

Les MFCC sont une mesure très usitée dans les domaines de la technologie vocale, dans les réseaux de neurones adaptés à la parole par exemple. C'est une mesure entièrement acoustique qui ne se base que sur le signal de parole.

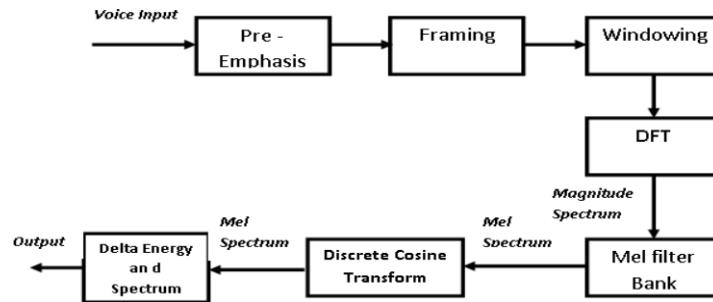


FIGURE 7 – Diagramme des étapes de calcul d'un MFCC, d'après Muda et al. (2010).

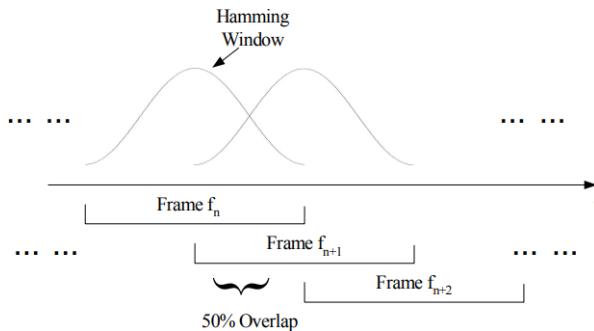


FIGURE 8 – Utilisation du fenêtrage Hamming, d'après Han et al. (2006).

Une fois les deux fonctions de modulation calculées, Goldstein calcule le taux de corrélation entre elles et montre qu'elles sont effectivement corrélées. Sachant que les dimensions de changement ne sont pas tout à fait les mêmes pour les deux fonctions, la corrélation trouvée est donc robuste. La forme de la syllabe a également une incidence sur la structure des pulsations observées dans le

signal, montrant que les fonctions de changement ont bien un rapport avec la structure syllabique. En plus de ces résultats, l'article montre la robustesse des MFCC dans le contexte de l'analyse acoustique de signaux de parole.

Dans la lignée de cet article, Lancia et al. (2020) utilise également les MFCC en s'appuyant sur les travaux de Goldstein : le but de l'article est d'étudier la parole perturbée et de comprendre les changements rythmiques effectués par les participants lors de la perturbation, en se basant sur la théorie suivante : les frontières prosodiques influencent les gestes qui s'allongent alors pour atteindre une frontière prosodique. Pour cela, ils utilisent un protocole avec un *feedback* auditif retardé : les sujets portent un casque qui joue ce qu'ils sont en-train de produire, mais en ajoutant aléatoirement du délai compris entre 0 et 180ms ; ainsi, les participants s'entendent de plus en plus tardivement, leur rythme de parole sera perturbé et ils commettent plus d'erreurs. Les prédictions sont que la complexité temporelle des énoncés augmentera avec le délai, et que les MFCC seront capables d'encoder cette complexité à l'aide de deux estimations : la richesse des dynamiques sous-jacentes au signal en MFCC, et la richesse des déformations temporelles du signal (constatée à partir de répétitions du signal).

Les résultats montrent que plus le délai est important, plus les locuteurs allongent la durée de leur énoncé, notamment lors de voyelles accentuées. Les deux estimations sont fonctionnelles, et permettent de détecter de petites différences dans la complexité sous-jacente du signal perturbé. Les auteurs concluent positivement par rapport aux MFCC : « Cette approche ne dépend pas d'une segmentation du signal et est basée sur des traits acoustiques de bas niveau, desquels la robustesse garantit l'application à un vaste éventail de populations (par exemple des nourrissons ou des locuteurs avec une pathologie) et de conditions d'énonciation. » (Lancia et al., 2020, p.2 [traduit de l'anglais]).

En renfermant les données articulatoires du conduit vocal, les MFCC sont également utiles pour étudier des maladies touchant la parole, comme la dysarthrie : une maladie affectant la respiration, le larynx et l'articulation. Elle peut être causée par la maladie de Parkinson ou la sclérose latérale amyotrophique par exemple. Slis et al. (2021) cherchent à utiliser les MFCC afin de décrire les problèmes articulatoires rencontrés par les patients atteints de sclérose latérale amyotrophique ou la maladie de Parkinson : ces deux maladies entraînent des effets sur la parole qui sont perceptuellement distincts, mais ont pourtant du mal à être classifiés en utilisant des mesures acoustiques

traditionnelles comme les espaces vocaliques ou la f_0 . Les résultats de cette étude montrent que les MFCC apportent effectivement des informations sur les changements sous-jacents dans le conduit vocal, étant capable de repérer des patients atteints de dysarthrie tout en les différenciant de ceux atteints de dysarthrie plus modeste et de sujets sains : les MFCC sont donc une mesure réussissant à traduire le signal acoustique en un vecteur de changements articulatoires.

Une autre étude explore les possibilités des MFCC pour les maladies liées à la parole : Lévêque et al. (2022) se proposent de différencier plusieurs types de dysarthrie, causés par différentes dégénérescences dans le conduit vocal – si les MFCC capturent les changements articulatoires du conduit vocal, ils devraient pouvoir différencier ces dysarthries. D'après les résultats, les MFCC réussissent à différencier les patients selon la cause de leur dysarthrie en conservant des informations sur la variabilité articulatoire dans les motifs acoustiques stables.

Actuellement, de nombreux articles réfléchissent à des améliorations quant à l'implémentation des MFCC, ou même d'algorithmes plus performants encore. Par exemple, Han et al. (2006) se proposent d'optimiser l'algorithme d'obtention des MFCC via la réduction de la taille du fenêtrage de 160 échantillons pendant 20ms, à 80 échantillons. Cela aboutit à une résolution fréquentielle moins bonne, mais une meilleure estimation du spectre global. L'avantage de cette modification est qu'elle nécessite moins de puissance calcul car le nombre de multiplications effectuées dans l'algorithme a été significativement réduit. De plus, la perte en précision n'impacte négativement la précision de l'analyse que de 1,5%.

Les performances des MFCC peuvent aussi être liées aux filtres passe-bande utilisés sur le signal, l'échelle utilisée à l'étape 5 des étapes de calcul d'un MFCC : il peut être important de modifier ces paramètres selon la qualité de l'enregistrement de départ (Zheng et al., 2001).

Enfin, Li et al. (2000) établissent un algorithme surpassant les MFCC en utilisant une échelle en Bark, une fonction de transfert qui modélise le gain de pression selon l'oreille externe, l'oreille moyenne et l'oreille interne et une fonction nonlinéaire afin de simuler la nonlinéarité du taux de décharge des nerfs auditifs ; la marge d'erreur de l'analyse est ainsi réduite entre 11% et 23%.

Les MFCC encodant les propriétés acoustiques du signal en un vecteur, il est ainsi possible d'encoder les caractéristiques acoustiques d'un locuteur. Tiwari (2010) utilise un algorithme afin de quantifier les centroïdes de ceux-ci ; en calculant la distance euclidienne entre ces centroïdes et ceux d'un autre signal, on peut évaluer la probabilité que le locuteur des deux extraits soit le même.

Résumé

- La coarticulation joue un rôle prédominant dans la parole : elle allège l'effort moteur pour les mots produits fréquemment (hypoarticulation) et s'assure de la clarté des mots peu fréquents (hyperarticulation).
- La prosodie permet la délimitation de constituants prosodiques. Tous ne sont pas systématiquement retrouvés dans la parole, et les plus couramment différenciés sont l'énoncé, le mot et la syllabe. Dans ces constituants, la production du phonème initial est renforcée (durée plus importante) alors que le phonème final est hyperarticulé (durée plus importante et meilleure articulation).
- La syntaxe a également un impact sur la parole car elle crée des liens de dépendance entre les mots.
- Des résultats montrent que ce sont majoritairement des considérations acoustiques et auditives qui sont prises en compte par les locuteurs lors de leurs productions vocales.
- Des vecteurs englobant les caractéristiques acoustiques du signal sont maintenant utilisés dans des analyses phonétiques de la parole, dont les résultats sont très robustes.

3 Objectifs et hypothèses

Nous avons vu que de nombreux facteurs entrent en compte dans la parole et peuvent favoriser sa variation : la coarticulation, la structure prosodique et la syntaxe d'un énoncé sont trois exemples importants. Les expériences que nous avons présentées montrent cette variation de façons très diverses, cependant nous n'avons vu que très peu d'études adoptant un corpus de parole spontanée : la plupart des observations sont donc ancrées dans une théorie linguistique du français, plutôt que l'usage qui en est fait.

En nous inspirant des travaux vus dans l'[État de l'art](#), nous nous proposons d'étudier, sous un angle acoustique, la durée et la modulation cepstrale à certaines frontières prosodiques du français dans deux corpus distincts :

- un corpus journalistique, dont la parole particulièrement normée du milieu du journalisme est attendue comme possédant les caractéristiques décrites par notre État de l'art ;
- un corpus de parole spontanée enregistré entre amis. Ce type de corpus est très peu utilisé pour acquérir des données linguistiques bien qu'il soit le plus représentatif des langues telles qu'elles sont usitées. Il sera donc très instructif de l'analyser afin de voir à quel point les modélisations de l'État de l'art peuvent s'y appliquer.

Nous nous proposons en premier lieu d'étudier le renforcement et l'hyperarticulation dans ces corpus, en s'appuyant sur la démarche de Keating et al. (2004) avec une approche acoustique basée sur la durée de diphones dans divers constituants prosodiques et leur modulation cepstrale. Ensuite, nous étudierons l'impact de la catégorie grammaticale sur ces mesures, ainsi que la forme de la syllabe dont est issue le diphone. Voici nos hypothèses :

- (i) Les mesures de durée et de modulation cepstrale permettent de différencier les éléments initiaux et finaux des constituants prosodiques du français. Cette distinction sera d'autant plus visible pour le corpus journalistique qui emploie un français plus normé et mieux articulé.
- (ii) Ces différences sont éprouvées peu importe la catégorie grammaticale du diphone. Nous prévoyons toutefois que les mesures aient des performances moins soutenues sur les catégories des mots grammaticaux (pronoms, déterminants, *etc...*).
- (iii) La forme de la syllabe (CV, VC...) a un impact sur les mesures car le phonème touché par les contraintes prosodiques de la frontière de constituant n'est pas le même.
- (iv) Les dépendances syntaxiques ont également un rôle à jouer dans les résultats : des mots

porteurs d'une nouvelle information seront hyperarticulés comparés à ceux portant une information déjà connue (complément d'objet et sujet respectivement par exemple, dans le cas d'une phrase conservant la structure informationnelle).

4 Méthode

4.1 Corpus et mesures utilisés

4.1.1 Corpus

Afin d'évaluer nos hypothèses, nous utilisons deux corpus : un corpus de parole spontanée, et un corpus de parole journalistique : cela nous permet de voir si les variations habituellement retrouvées en parole contrôlée sont également présentes dans la parole spontanée.

Le corpus de parole spontanée utilisé est le corpus NCCFr (Nijmegen Corpus of Casual French, Torreira et al. (2010)). Il s'agit d'un recueil de trente-six heures d'enregistrements de parole française spontanée produite par quarante-six locuteurs natifs du français lors de conversations entre amis. Du fait du caractère conversationnel de l'enregistrement, des micros sont posés devant chaque locuteur, et certains fichiers se répondent donc, donnant lieu à des moments de silence où seule une voix lointaine peut être entendue – celle-ci sera intelligible dans son fichier correspondant. Il est transcrit orthographiquement et segmenté phonétiquement, ce qui en fait un corpus adéquat car nous pourrons donner la transcription en entrée afin d'obtenir les données syntaxiques, et l'utiliser dans des analyses phonétiques fines.

Le corpus de parole journalistique est le corpus Ester (Galliano et al., 2006), composé originellement de 100 heures d'émissions de radio française enregistrées et transcris manuellement avec des annotations comme le locuteur par exemple, présents au nombre de 2172 dans le corpus transcrit. Des outils tels que l'alignement automatique et un dictionnaire d'équivalences ont permis d'améliorer les transcriptions : l'alignement automatique a aidé à vérifier la justesse des transcriptions.

Des campagnes d'évaluation autour de Ester se sont tenues, ayant pour but de fixer des méthodes d'évaluation d'annotations de grands corpus et de ressources issues de diffusions radio en les testant sur Ester afin de voir leurs performances (Galliano et al., 2009). Des tâches telles que la localisation de locuteurs, la détection d'entités nommées ou le suivi d'événements sonores (détection de la musique en premier plan ou en arrière-plan, suivi de la voix avec de la musique en fond) ont été effectuées sur le corpus avec différents résultats : par exemple, la détection de parole a obtenu de bons résultats, alors que la détection de musique s'est avérée plus compliquée – le corpus comportant peu de segments uniquement musicaux et les évaluateurs n'ayant souvent pas accès à un système spécialisé dans la détection de musique. Ces campagnes d'évaluation ont aidé

les laboratoires participants à comprendre les enjeux derrière la transcription et le traitement des données issues de diffusions des actualités.

Les corpus utilisés ont été obtenus avec chaque enregistrement en .wav ainsi qu'un fichier .TextGrid. Les TextGrids contiennent des champs comme phonème, mot orthographique, locuteur, sexe, mot SAMPA et syllabation. Des extraits des corpus sont visibles Figure 9.

4.1.2 Modulation cepstrale

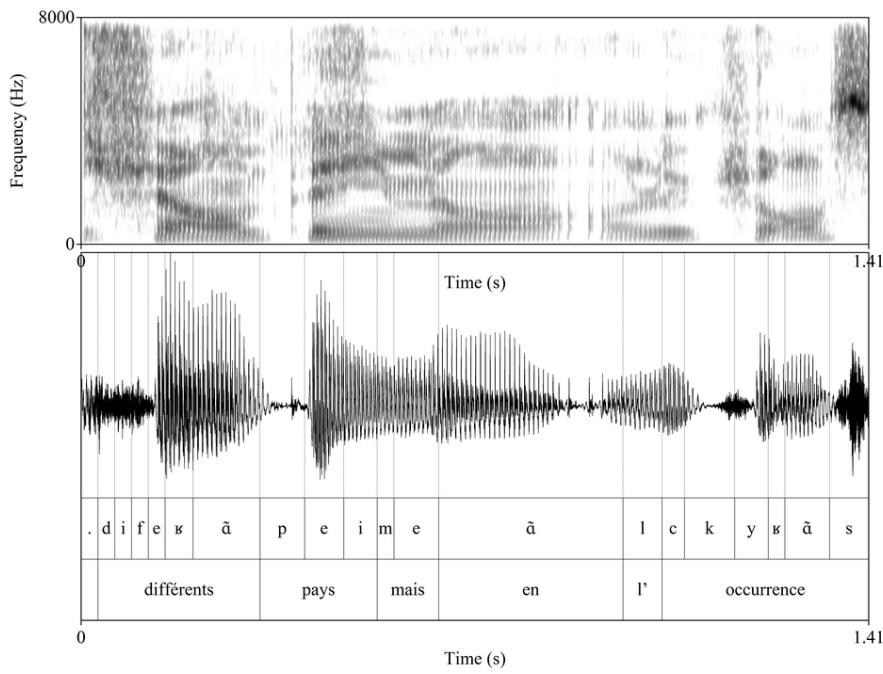
La première mesure utilisée est la modulation cepstrale qui permet de mesurer, d'une fenêtre temporelle à la suivante, la différence spectrale à partir de MFCC (Mel Frequency Cepstral Coefficients), des coefficients permettant de caractériser un signal donné (voir [État de l'art](#)). Elle est applicable à tout type de signal de parole, voyelle comme consonne, et est donc plus robuste que des mesures basées sur les formants ou les durées.

Elle est ici utilisée sur des diphones (comportant la deuxième partie d'un phonème et la première partie du phonème suivant) et non des phonèmes car il nous a paru plus pertinent de traiter les phénomènes d'hypo et hyperarticulation en analysant la combinaison de deux phonèmes plutôt qu'un seul, la parole étant un continuum. Cette mesure permet d'obtenir une courbe de variation spectrale pour laquelle nous émettons l'hypothèse qu'elle est – pour un même diphone – plus importante si le diphone produit est mieux articulé/hyperarticulé. Cette mesure est normalisée d'après un ensemble de corpus pré-existants. La durée des diphones sera également mesurée, ainsi que la f_0 de chaque phone quand celui-ci est voisé, et les bigrammes pour chaque mot.

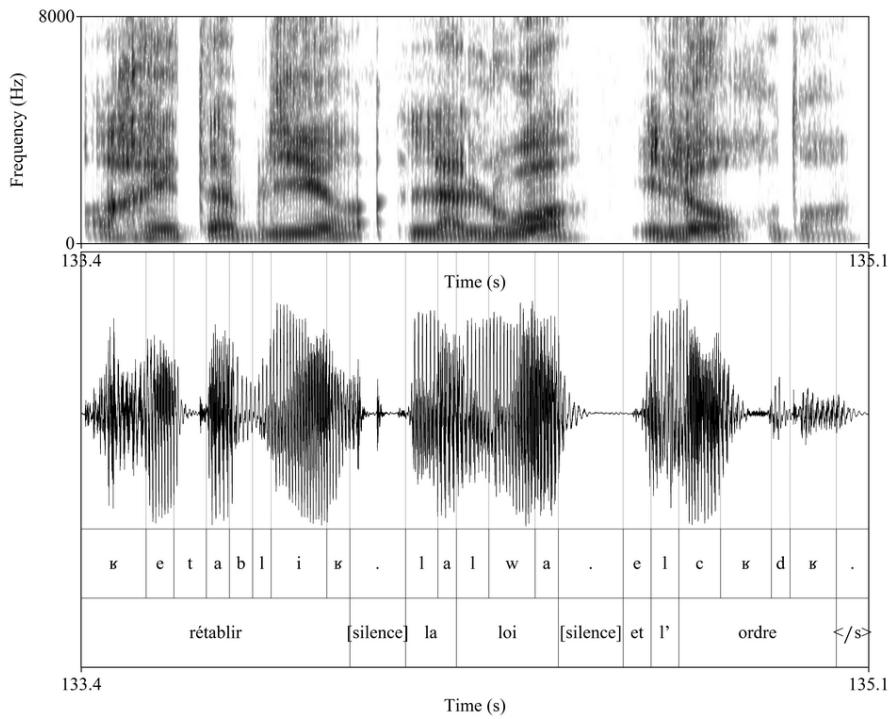
L'utilisation de la modulation cepstrale sur un corpus de parole spontanée nous permet de mettre à l'épreuve les résultats précédemment trouvés à l'aide de corpus de parole contrôlée (parole journalistique, ou parole de laboratoire). Le corpus journalistique nous permettra de vérifier la validité de nos mesures sur un type de parole déjà éprouvé par des études phonétiques. La combinaison durée/modulation cepstrale va permettre de mieux comprendre les paramètres de la variation de la coarticulation, en fonction du niveau prosodique des segments et de leurs liens de dépendance.

4.2 Obtention des données

Pour obtenir nos résultats, nous avons besoin des éléments suivants : la durée de chaque diphone, leur transcription ainsi que celle du mot auquel ils appartiennent, les mesures de modulation



(a) Extrait du corpus NCCFr.



(b) Extrait du corpus Ester.

FIGURE 9 – Exemples issus des corpus utilisés.

cepstrale, la catégorie grammaticale des mots, leurs dépendances syntaxiques, la f_0 des phones, et le locuteur.

4.2.1 Durée et frontières prosodiques

La durée des diphones aux frontières prosodiques est obtenue en faisant la moyenne de la durée des deux phones : si le premier phone dure 90ms et le deuxième 30ms, la durée sera égale à $(90 + 30)/2 = 60\text{ms}$; en effet un diphone est défini comme la deuxième moitié du premier phone et la première moitié du deuxième phone. Par exemple, le diphone /si/ est composé de la deuxième moitié de /s/ et la première moitié de /i/.

Les deux niveaux prosodiques étudiés sont le mot et la séquence, définie comme une suite de mots entre deux pauses. Pour identifier les séquences, de nombreuses annotations peuvent indiquer des pauses dans les corpus : on trouve dans le champ « mot » les codes [silence], {breath}, et les débuts et fins de phrase indiqués par <s> et </s> respectivement. Dans le champ des mots en SAMPA et des phonèmes, des combinaisons avec . et H sont également employées. Il nous suffit de chercher les diphones suivis par ces annotations afin de trouver les diphones prépausals par exemple – les diphones postpausals sont repérés en cherchant les diphones précédés par ces annotations. Par conséquent, les diphones précédent et suivant chaque diphone font partie des données que nous extraierons des TextGrids.

L'impact de la durée des pauses et l'annotation de celles-ci feront l'objet d'un examen dans les parties [Résultats](#) et [Discussion](#).

L'identification des débuts et fins de mots se base sur l'annotation des TextGrids, mais du fait que nous travaillons avec des diphones, il nous faut clarifier les phénomènes concernant la transition entre deux mots, comme les liaisons. Dans nos données, les diphones se trouvant entre deux mots comme /rd/ dans <partir de> sont considérés, en tant que syllabe, comme appartenant à « partir ». Mais indépendamment, chaque phone est catégorisé comme faisant partie de « partir » et « de » respectivement. Plus exactement : la position de la syllabe dans le mot pour les phones 1 (/r/) et 2 (/d/) est 2_2, correspondant à la dernière syllabe de « partir », mais la position du phone 1 dans le mot est 6_6 (sixième phone de « partir ») et celle du phone 2 est 1_2 (premier phone de « de »). Un exemple de diphones dans un même mot ou un mot différent est montré Figure 10. Afin de pouvoir filtrer les diphones appartenant à un même mot et ceux appartenant à deux mots

différents, nous créons dans le tableau de données final sur R, une colonne distinguant les deux cas de figure selon que le mot suivant celui auquel appartient le diphone actuel soit le même ou non. L'impact de l'appartenances des diphones à un même mot, ou à des mots différents, sera examiné dans *Résultats et Discussion*.

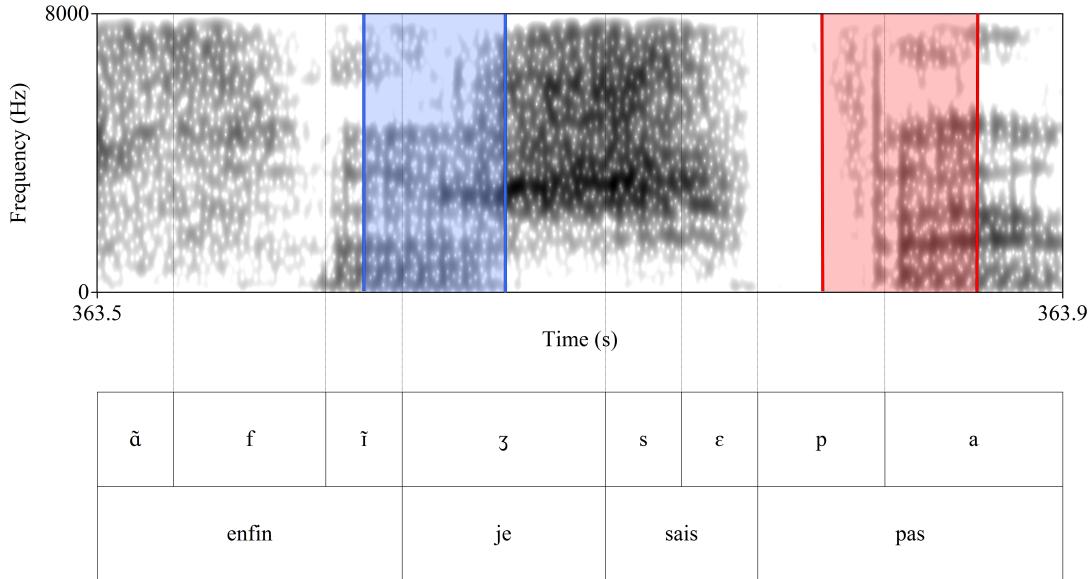


FIGURE 10 – Exemple de diphones appartenant à un mot différent (« **enfin j'sais** ») et à un même mot (« **pas** »), tiré du fichier son 05_11_07_nb1_1_16.wav : « *enfin j'sais pas* ».

En ce qui concerne le phénomène de liaison, prenons l'exemple de « nous envoie » du fichier 03_12_07_nb1_1_16. La liaison a bien lieu : on retrouve les diphones /nu/, /uz/, /zã/, /ãv/, /vw/ et /wa/. Les deux diphones concernés par la liaison, /uz/ et /zã/, sont considérés comme appartenant au mot « nous ». De fait, dans notre nouvelle colonne, /zã/ est catégorisé comme appartenant à deux mots différents tandis que /uz/ appartient au même mot « nous ». Cette classification est justifiée ; lexicalement parlant, c'est le mot « nous » qui cause l'apparition du /z/, et à ce titre la liaison lui appartient. Cependant, sous un autre angle, syllabiquement parlant c'est /zã/ – la syllabe CV – qui porte la liaison et donc a fortiori le mot « envoie » : il y a resyllabation de la consonne finale sur l'initiale vocalique suivante (Encrev , 1988).

4.2.2 Parties du discours et dépendances syntaxiques

L'annotation en syntaxe de dépendance et en parties du discours est fournie par M.Kim Gerdès, et est basée sur un apprentissage supervisé à partir de corpus de données.

Afin d'obtenir ces annotations, il nous faut extraire le texte contenu dans les TextGrids. Nous utilisons un script Praat `transformation_textgrid_en_phrases.praat` qui récupère les mots de la TextGrid et les retranscrit de sorte à obtenir un fichier .txt avec chaque séquence de l'enregistrement ligne par ligne : quand le script repère un mot il l'ajoute avec une espace après, et quand il repère une fin de séquence il passe une ligne – les fins de séquence sont signalées dans la transcription par `<\s>`.

Avec ces fichiers .txt, l'obtention des parties du discours et des dépendances syntaxiques est ainsi possible. Celles-ci sont recueillies sous la forme de fichiers .conllu, organisés de la façon suivante à chaque ligne : mot, lemme, partie du discours, flexions du mot, et dépendance syntaxique – un système de numérotation permet de savoir quelle dépendance est rattachée à quel élément –, il y a également des lignes indiquant la phrase entière, le locuteur et l'indice temporel. Un exemple est donné dans le Tableau 1. Dans ce tableau, on voit que le déterminant « la » est syntaxiquement dépendant de « pièce » car son numéro de dépendance est 4, ce qui correspond au numéro de « pièce ». Ce dernier est lui-même une dépendance de « de ».

Num	Mot	Lemme	PoS	Flexions	Dépendance	Num dépendance
1	partir	partir	VERB	VerbForm=Inf	root	—
2	de	de	ADP	—	comp :obl	—
3	la	le	DET	Definite=Def Gender=Fem Number=Sing PronType=Art	det	4
4	pièce	pièce	NOUN	Gender=Fem Number=Sing	comp :obj	2

TABLE 1 – Annotations syntaxiques pour l'énoncé « Partir de la pièce ».

Nous ajoutons ces informations dans les TextGrids avec le script Praat `conversion_annotation_dependances_en_textrgids.praat`; les parties du discours sont situées après trois tabulations tandis que les dépendances sont après deux tabulations à partir de la fin de chaque ligne. Nous utilisons ces informations afin de mémoriser les parties du discours et les dépendances dans le script, et de les placer dans de nouveaux champs de la TextGrid délimités comme les mots ; si le mot n'est pas vide, ni un début de séquence ou un souffle, on ajoute au champ la partie du discours et la dépendance syntaxique.

4.2.3 Calcul de la modulation cepstrale

Le calcul de la modulation cepstrale commence avec un script Python fourni par Leonardo Lancia, suivant une méthodologie similaire à celles décrites dans l'[État de l'art](#) : le fichier audio est ouvert avec une fréquence maximale de 10 000Hz. Ensuite, on prend des échantillons par pas de 0,005ms avec un fenêtrage de 0,05ms ; les paramètres de la Transformée de Fourier Rapide sont 512 points par échantillon de signal. Après plusieurs transformations du signal, on obtient alors un cepstre – au lieu d'avoir un spectre, la fréquence selon l'amplitude, on a le temps en fonction de l'amplitude. On applique un filtre passe-bas, et on acquiert les MFCC, desquels on peut obtenir le taux de changement cepstral dans le temps sous forme de tableau.

Dans le script `mesures_coarticulation_etape1_reference.praat`, on ouvre ce tableau et crée un signal à partir de celui-ci. On récupère ensuite les intervalles des différents diphones afin de savoir quel taux de changement cepstral appartient à quel diphone – on prélève six mesures à partir de ce taux : le maximum du taux de changement cepstral, son minimum, la dispersion de modulation cepstrale par diphone (*maximum–minimum*) – illustrée sur la Figure 11 –, la moyenne, l'écart-type (racine carrée de la variance, il mesure la distance moyenne entre chaque valeur et la moyenne) et la durée.

On utilise ensuite les scripts `mesures_coarticulation_etape2_ratio_norm.praat` et `mesures_coarticulation_etape3_ratio_nn_norm.praat` qui vont respectivement calculer la modulation cepstrale normalisée, et la modulation cepstrale non normalisée ; la différence entre les deux est que la version normalisée voit ses mesures divisées par les mesures de l'ensemble du fichier de taux de changement cepstral original. Ces scripts stockent les informations obtenues dans un fichier texte, avec le diphone concerné, le nom du fichier son et le mot auquel le diphone appartient.

Sur la Figure 12, on peut voir à quoi ressemblent les deux versions de la modulation cepstrale obtenues une fois les valeurs converties en signal sur Praat à l'aide des paramètres `Read table → Down to Matrix → Transpose → To Sound`. On voit que la version non normalisée conserve toutes variations fines de changement cepstral, lesquelles sont propres aux échantillons de signal individuels. La version normalisée, elle, a subi des modifications entre chaque transition entre phone causées par la normalisation ; cette modulation cepstrale a un aspect crénelé qui, pour chaque créneau, correspond à la valeur de la modulation sur cet échantillon divisée par les valeurs de l'ensemble des modulations cepstrales. Nous avons comparé ces deux versions, mais les mesures sont très similaires. De ce fait, la normalisation demandant plus de ressources et risquant possiblement de "tasser" les

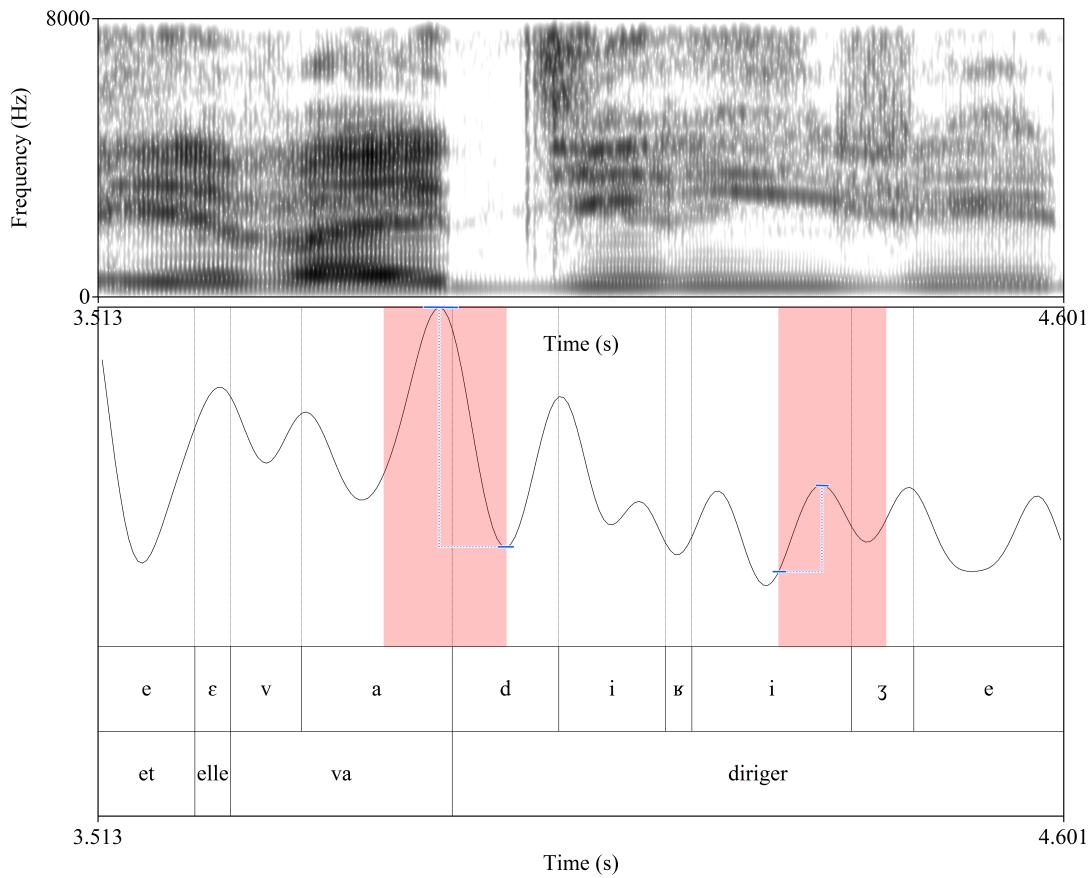


FIGURE 11 – Illustration de la dispersion de la modulation cepstrale non normalisée par diphone (différence entre le maximum et le minimum), à partir du fichier 23_11_07_nb1_2_16.

résultats du fait du nombre d’occurrences important dû à la grande taille des corpus, nous utilisons la mesure non normalisée par la suite. L’écart-type de la modulation cepstrale n’apportait pas de résultat intéressant et n’a pas été gardé non plus.

Une modulation cepstrale haute équivaut à un taux de changement cepstral important, brutal entre deux segments. La différence entre les deux segments est donc élevée. À l’inverse, une modulation cepstrale basse indique un taux de changement cepstral bas, ce qui veut dire qu’il y a peu de changement entre les deux segments. Nous faisons l’hypothèse qu’une modulation cepstrale haute – une différence importante entre les deux segments du diphone – indique qu’au moins un des deux segments est hyperarticulé, d’après les articles sur la modulation cepstrale de notre État de l’art. Afin d’avoir un aperçu du fonctionnement de la modulation cepstrale, nous avons prélevé des extraits de spectrogrammes selon divers paramètres : la même durée ($\sim 100\text{ms}$) et des modulations cepstrales différentes (3,49 et 10,49) Figure 13, et la même modulation cepstrale ($\sim 9,10$) et des

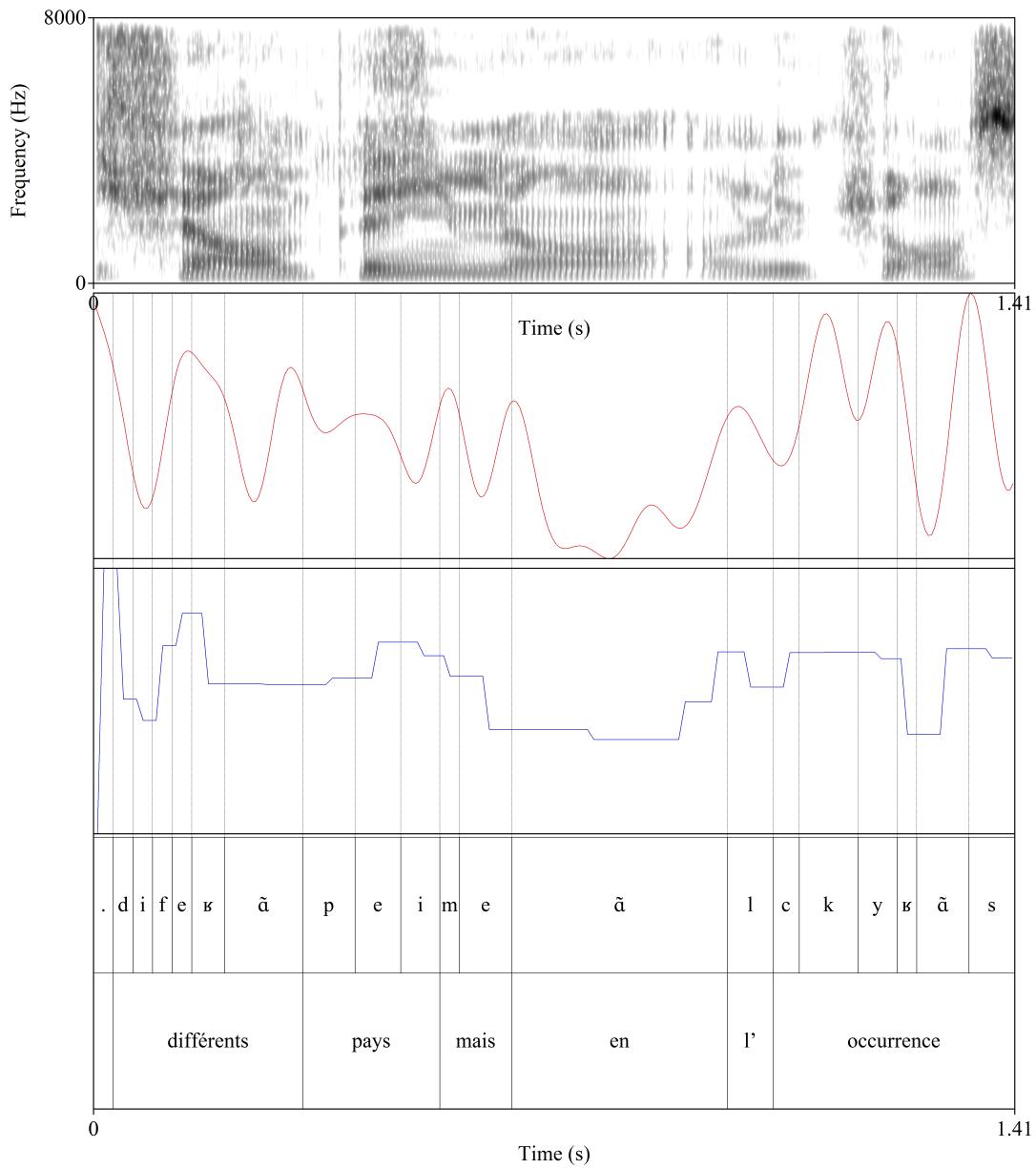


FIGURE 12 – Modulations cepstrales non normalisée (en haut, rouge) et normalisée (en bas, bleu) pour la première seconde et demie du fichier son 05_11_07_nb1_1_16.wav : « différents pays mais en l'occurrence ».

durées différentes (90ms et 1sec) Figure 14.

Sur la Figure 13, on voit sur l'image (a) que le diphone est plus articulé que sur l'image (b) : /u/, qui devrait surtout être composé de basses fréquences, possède aussi des fréquences plus hautes à cause du /t/ dental le précédent ; cette occlusive est elle-même produite plus proche d'une fricative, avec une barre de relâchement peu visible et du bruit dans le signal.

Sur la Figure 14, on voit que l'image (a) a une moins bonne définition que la (b) car le signal est plus court, cependant les fréquences sont amplifiées de façon similaire à la (b) – on distingue notamment les formants couramment décrits pour /r/.

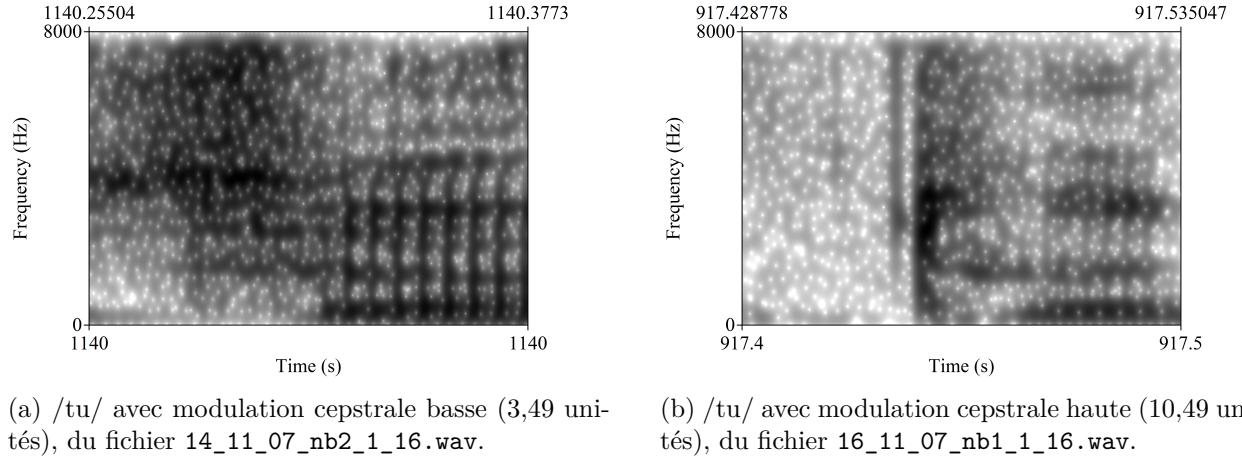


FIGURE 13 – /tu/ selon la même durée ($\sim 100\text{ms}$) mais des modulations céphrales différentes.

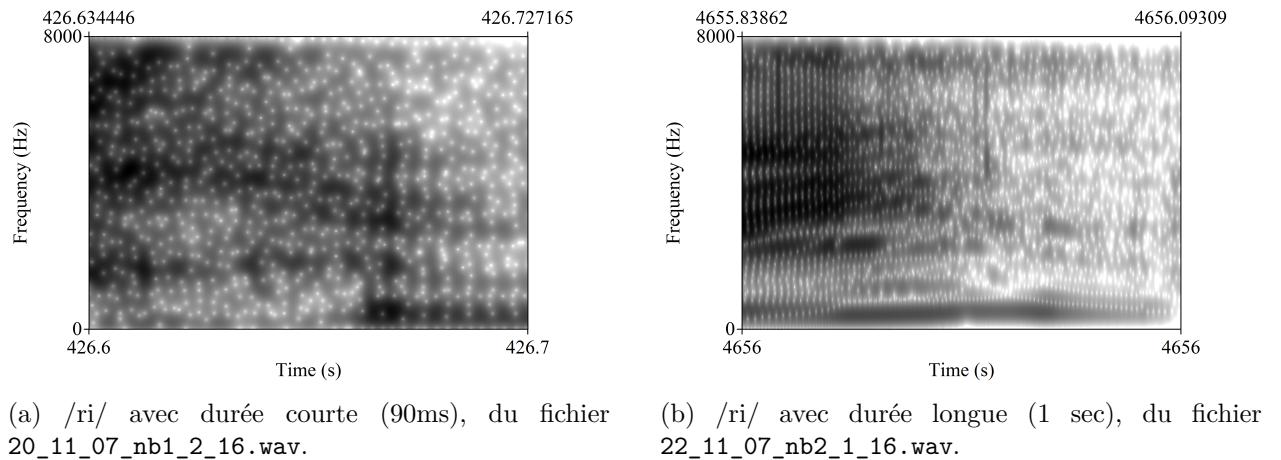


FIGURE 14 – /ri/ selon la même modulation céphrale ($\sim 9,10$ unités) mais des durées différentes.

Le taux de modulation céphrale peut être ainsi comparé, mais il peut être intéressant de voir les moyennes de cette mesure afin de l’appréhender. Nous pouvons par exemple extraire les boîtes à moustaches des moyennes et dispersions de modulation céphrale en fonction des locuteurs de NCCFr (Figure 15) : nous voyons que les moyennes de la dispersion se situent entre 4,5 et 6 unités, tandis que celles de la moyenne sont entre 9,5 et 10 unités – ces valeurs ne sont donc pas comparables. Elles permettent de voir différentes répartitions des locuteurs : sur la Figure 15a, on voit qu’un locuteur en particulier – le douzième en partant de la droite – a une moyenne de dispersion extrêmement basse ;

la dispersion du taux de changement cepstral est basse. Cependant, sa moyenne de modulation cepstrale n'est pas excessivement basse pour autant, ce qui peut s'expliquer ainsi : la dispersion sera la même que le maximum et le minimum soient hauts (10 et 8 par exemple) ou bas (4 et 2 par exemple), alors que la moyenne sera plus haute dans le premier cas – c'est ce dont il pourrait s'agir ici.

Selon la mesure utilisée, les conclusions ne seront pas les mêmes, et il sera important de les comparer. La dispersion est à première vue plus fiable, puisqu'elle mesure à quel point la différence entre le début et la fin du diphone est grande. Cela est important car l'énergie spectrale est distribuée de façons différentes selon la diphone : des fricatives ont une énergie distribuée dans les hautes fréquences par exemple.

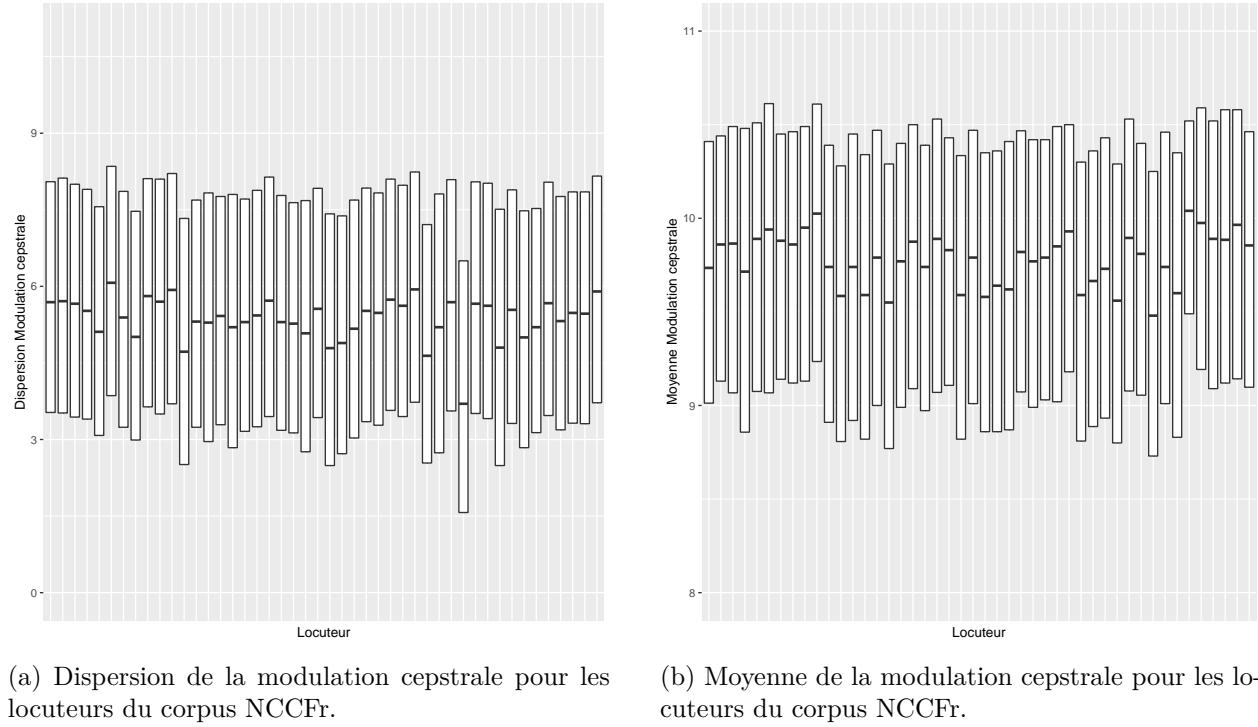


FIGURE 15 – Boîtes à moustache de la dispersion et la moyenne de la modulation cepstrale dans les corpus NCCFr.

Enfin, notre analyse s'effectuant sur R, nous extrayons toutes les données obtenues dans un fichier .txt à l'aide du script Praat `extraction_textgrid.praat` : on récupère les éléments nécessaires mentionnés plus haut présents dans les TextGrids (durée, données syntaxiques, f_0) en prenant soin d'écrire `undefined` dans le tableau si la f_0 n'est pas obtenable (dans le cas de phonèmes sourds par exemple). On prend aussi des informations sur la position du diphone selon sa syllabe, son mot d'appartenance. On ajoute par la suite les mesures de modulation cepstrale obtenues, normalisées

et non normalisées.

Les résultats sont analysés avec R et RStudio. On utilise notamment le script `donnees_table.R` afin d'obtenir un tableau avec les données de durée et de modulation cepstrale selon la frontière prosodique (séquence ou mot), pour un diphone dans un même mot (/si/ dans <cassis>) ou dans deux mots différents (/si/ dans <course hippique>). On utilise les *packages* `languageR`, `readxl`, `dplyr`, `stringr`, `ggplot2`, ainsi que les fonctions `mean()`, `aov()`, `lm()` et `TukeyHSD()`.

Résumé

- Les deux corpus en français utilisés sont NCCFr (parole spontanée, 36h d'enregistrement) et Ester (parole journalistique, 100h d'enregistrement).
- On en extrait la durée, la modulation cepstrale, la f_0 , les parties du discours et les dépendances syntaxiques.
- Les langages Python, Praat et R sont utilisés.
- Si la modulation cepstrale est haute, alors il y a un changement brutal entre le début et la fin du diphone : il est bien articulé. Au contraire, si elle est basse, il y a une forte coarticulation du diphone.

5 Résultats

Dans cette section, nous présentons nos résultats par ordre d'importance : d'abord la relation entre frontières prosodiques, durée et modulation cepstrale, puis en approfondissant en comparant mots grammaticaux et lexicaux.

5.1 Durée et modulation cepstrale aux frontières prosodiques

Notre première hypothèse était la suivante :

(i) *Les mesures de durée et de modulation cepstrale permettent de différencier les éléments initiaux et finaux des constituants prosodiques du français. Cette distinction sera d'autant plus visible pour le corpus journalistique qui emploie un français plus normé et mieux articulé.*

Nous filtrons nos résultats de sorte à retirer les hésitations des corpus, annotées en diphones comme des combinaisons de @, & et _ (@_, @@, _&, etc...). En effet, ces hésitations sont traitées comme des diphones à part entière, et augmentent la longueur moyenne de ceux-ci : la durée moyenne des diphones du corpus Ester est réduite de 10ms quand on enlève les hésitations (de 90ms à 82ms), par exemple, tandis qu'il y a une réduction de presque 100ms entre le corpus NCCFr avec hésitations, et sans hésitations (de 169ms à 71ms).

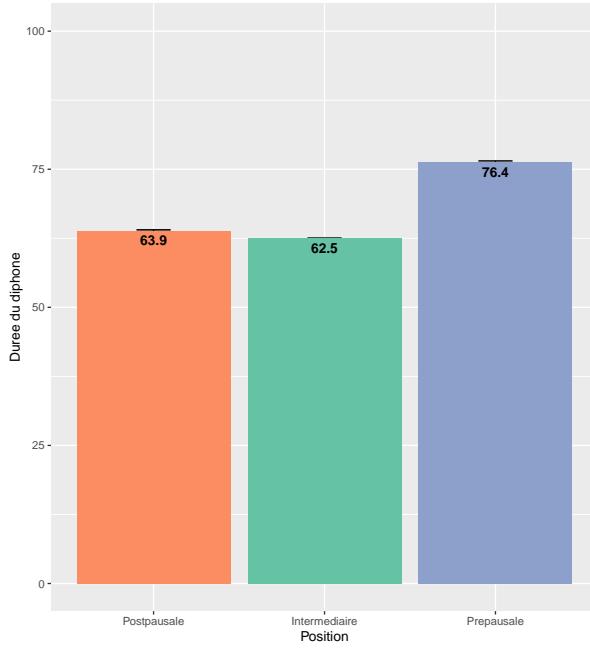
Nous mettons également à l'écart les diphones appartenant à des mots différents, en effet cela biaise nos mesures en les rendant plus importantes, car les effets de jointure entre différents mots provoquent entre autre une hausse de la durée et une meilleure articulation.

Nous obtenons un aperçu de ces résultats dans le Tableau 2 qui répertorie la moyenne de la durée et la dispersion de la modulation cepstrale du corpus selon la position du diphone dans la séquence (avant ou après une pause, ou dans la séquence directement. La catégorie « mono » fait référence aux cas où le diphone est dans un mot entre deux pauses) ou dans le mot (au début, au milieu, à la fin, ou dans un mot monosyllabique que nous traitons comme une catégorie à part entière). Le nombre d'occurrences des diphones dans chaque catégorie est aussi donné.

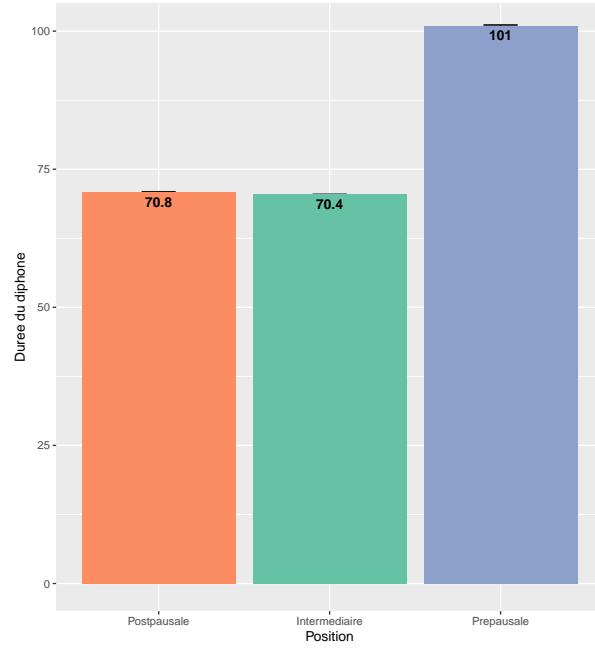
Modalité	Position	Occurrences NCCFr	NCCFr	Occurrences Ester	Ester
<i>Durée</i>					
Niveau Séquence	Début	38945	64	51022	70
	Inter	566837	62	1465846	70
	Fin	48107	78	105558	101
	Mono	11928	78	3446	106
Niveau Mot	Début	95833	64	253834	72
	Inter	183452	61	710232	69
	Fin	83192	70	212784	82
	Mono	335530	64	484683	74
<i>Modulation cepstrale</i>					
Niveau Séquence	Début		8.8		9.8
	Inter		8.3		9.3
	Fin		9.5		13.1
	Mono		8.7		12.8
Niveau Mot	Début		8.3		9.2
	Inter		8.2		9.1
	Fin		9.2		11.1
	Mono		8.5		9.7

TABLE 2 – Tableau contenant les durées et modulations cepstrales de dipphones appartenant au même mot dans NCCFr et Ester, dans différentes structures prosodiques.

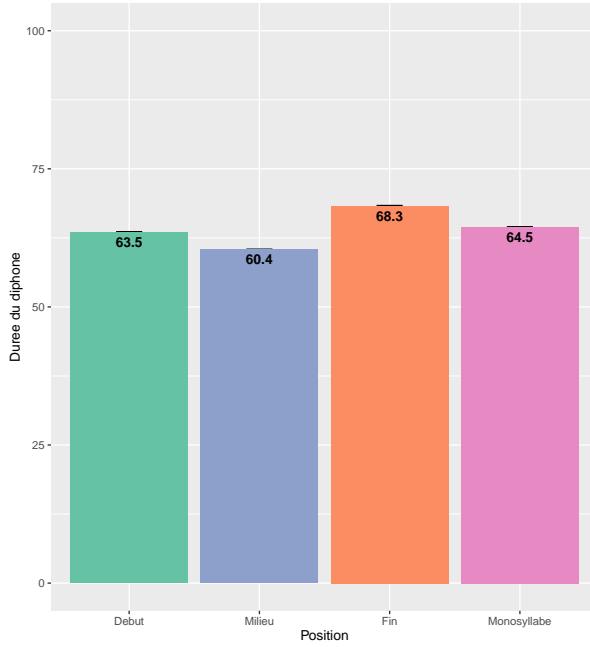
Afin d'illustrer plus clairement notre propos, nous avons utilisé `ggplot2` afin de créer des diagrammes en barres (Figures 16 et 17).



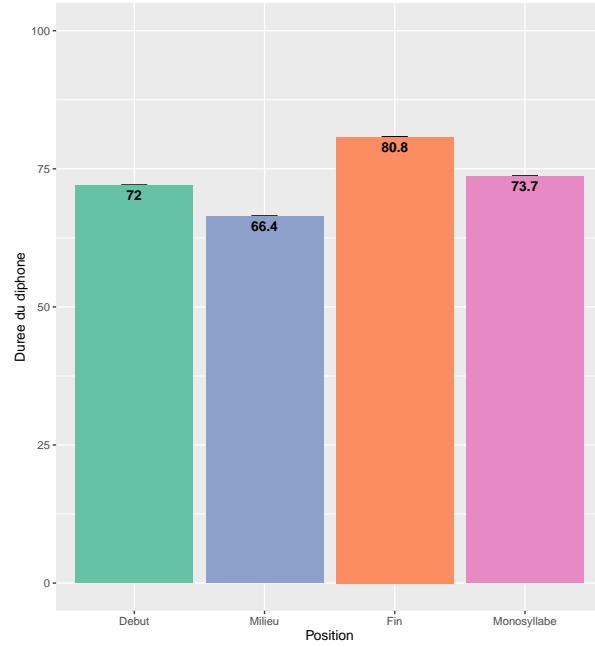
(a) Durée du diphone selon sa position dans une séquence du corpus NCCFr : après une pause, dans la séquence, et avant une pause.



(b) Durée du diphone selon sa position dans une séquence du corpus Ester : après une pause, dans la séquence, et avant une pause.

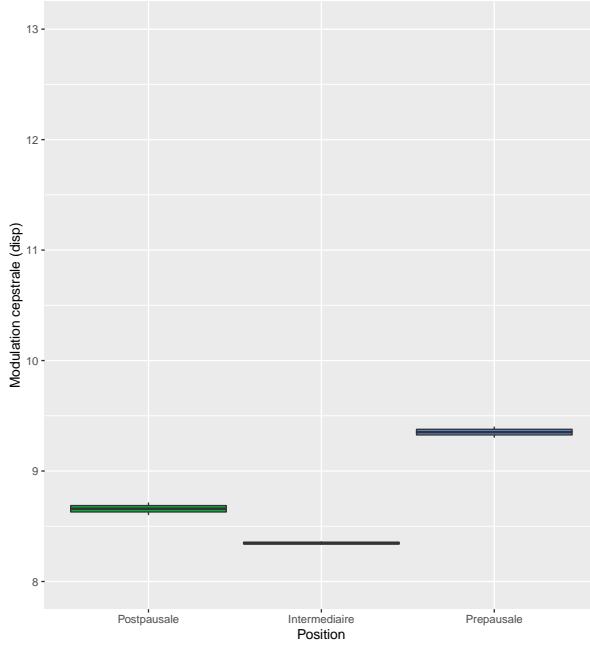


(c) Durée du diphone selon sa position dans un mot du corpus NCCFr : au début du mot, au milieu, à la fin ou dans un mot monosyllabique.

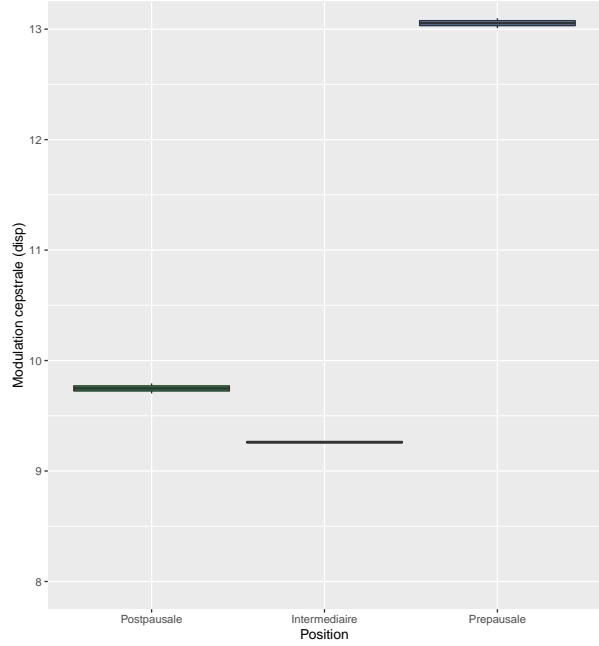


(d) Durée du diphone selon sa position dans un mot du corpus Ester : au début du mot, au milieu, à la fin ou dans un mot monosyllabique.

FIGURE 16 – Durée du diphone selon sa position dans une frontière prosodique.



(a) Dispersion de la modulation cepstrale du diphone selon sa position dans une séquence du corpus NCCFr : après une pause, dans la séquence, et avant une pause.



(b) Dispersion de la modulation cepstrale du diphone selon sa position dans une séquence du corpus Ester : après une pause, dans la séquence, et avant une pause.

FIGURE 17 – Dispersion de la modulation cepstrale du diphone selon sa position dans une frontière prosodique.

Nous voyons que, au niveau de la séquence, les diphones prépausals ont une durée plus grande que les postpausals et intermédiaires. Ces derniers ne se diffèrent pas l'un de l'autre par leur durée. Enfin, le diphone dans une séquence entre deux pauses (*mono*, visible sur le Tableau 2) possède les mêmes caractéristiques que le diphone prépausal. Ces observations sont similaires dans le niveau prosodique du mot, sauf pour le diphone dans un mot monosyllabique : cette fois-ci, il est plus proche des diphones en début ou milieu de syllabe.

Concernant la modulation cepstrale, elle agit de façon similaire à la durée pour le niveau du mot, mais a une autre propriété pour le niveau de la séquence : un diphone postpausal a une modulation cepstrale plus importante qu'un diphone en milieu de séquence.

Comme notre hypothèse l'envisageait, les différences sont plus saillantes pour le corpus journalistique Ester que pour celui de parole spontanée NCCFr.

Étant donné que nous travaillons sur de grands corpus de données, nous utilisons des modèles mixtes incluant des effets aléatoires afin de contrôler la significativité des résultats : en plus des

prédicteurs, qui sont les variables qui prédisent le mieux la distribution des données, on ajoute des variables aléatoires dont on sait qu'elles ont un impact sur les données.

Les prédicteurs que nous testons sont la durée et la modulation cepstrale sur la position du diphone. Les variables aléatoires sont le diphone concerné, et le locuteur. Nous utilisons la fonction `lmer()` sur les deux prédicteurs (ici la durée) et les variables aléatoires :

```
lmer(duree ~ position+(1|diphone)+(1|locuteur), data=corpus) (4)
```

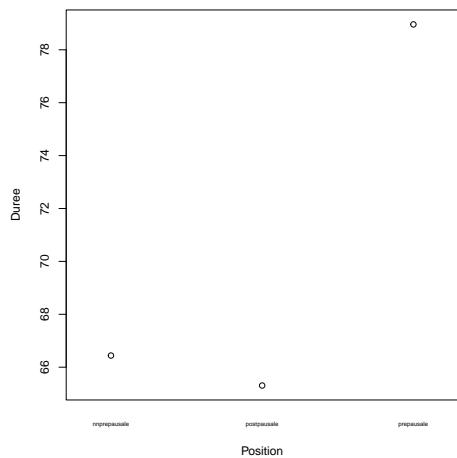
Pour tester la pertinence des variables aléatoires sélectionnées, nous effectuons le test statistique avec et sans chacune d'elle – ensuite, on effectue une ANOVA entre les deux versions pour voir s'il était pertinent d'utiliser la variable.

Variable	Significativité
Diphone	$p = 2.2e - 16 * **$
Locuteur	$p = 2.2e - 16 * **$
Appartenance au même mot	$p > 0,05$

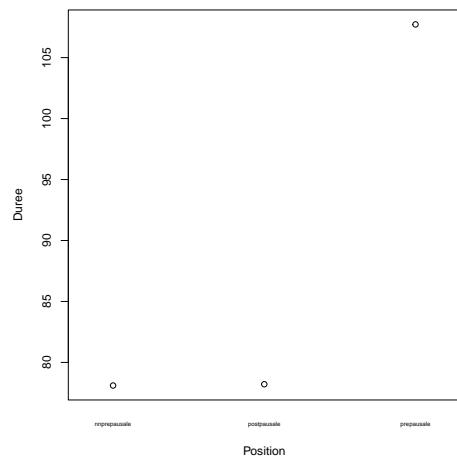
TABLE 3 – ANOVA utilisée pour comparer l'importance des variables aléatoires *diphone*, *locuteur* et *appartenance au même mot ou à des mots différents* sur le modèle mixte.

D'après les résultats, les deux variables aléatoires ont un impact significatif sur le modèle et sont conservées pour la suite. L'appartenance du diphone au même mot ou non a aussi été testée, mais elle n'impactait significativement pas le modèle et n'a donc pas été gardée : dans les données telles que le Tableau 13, on voit que les durées sont plus longues, mais conservent la même tendance que le Tableau 2 malgré tout.

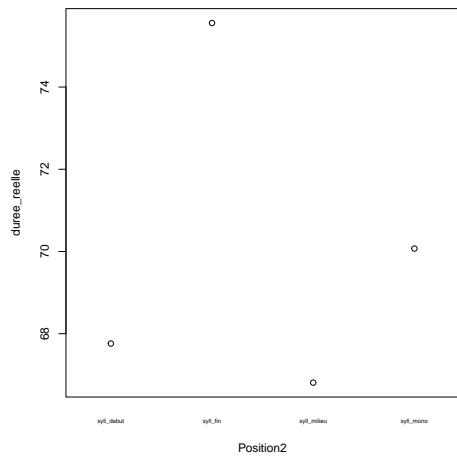
Ci-dessous sont les plots obtenus en sélectionnant les prédicteurs *durée* (Figure 18), *dispersion de modulation cepstrale* et *moyenne de modulation cepstrale* (Figure 19).



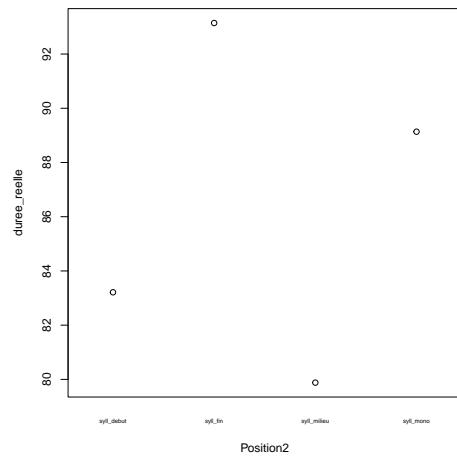
(a) `plotLMER.fnc()` pour la durée en fonction de la position (nnprepausale = milieu de séquence) du diphone dans la séquence, corpus NCCFr.



(b) `plotLMER.fnc()` pour la durée en fonction de la position (nnprepausale = milieu de séquence) du diphone dans la séquence, corpus Ester.



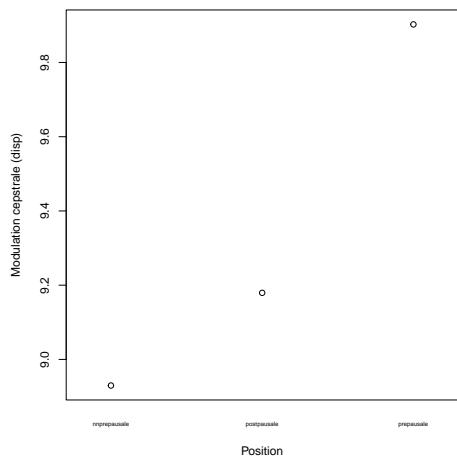
(c) `plotLMER.fnc()` pour la durée en fonction de la position (mono = mot monosyllabique) du diphone dans le mot, corpus NCCFr.



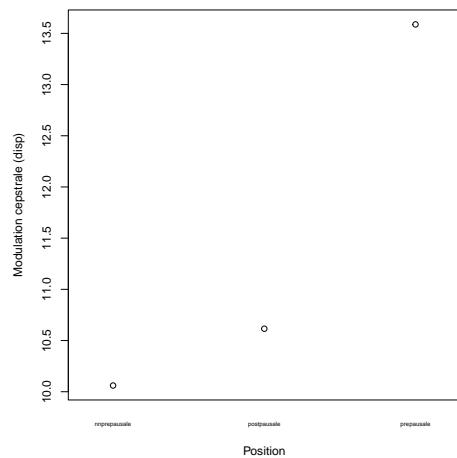
(d) `plotLMER.fnc()` pour la durée en fonction de la position (mono = mot monosyllabique) du diphone dans le mot, corpus Ester.

FIGURE 18 – `plotLMER.fnc()` pour la durée en fonction de la position du diphone dans la séquence et le mot.

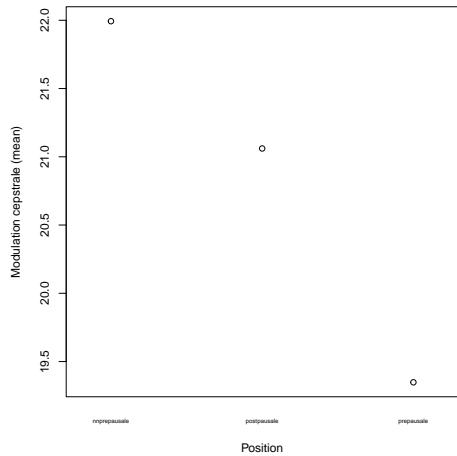
On voit, au niveau de la séquence, que la durée différencie significativement les diphones pré-pausals des autres diphones, mais elle peine à différencier l'opposition entre diphone postpausal et intermédiaire. Cependant, le modèle est plus constant pour le niveau prosodique du mot : la fin du mot est plus longue que le début, lui-même plus long que le milieu du mot. Enfin, le mot monosyllabique est situé entre la fin et le début du mot.



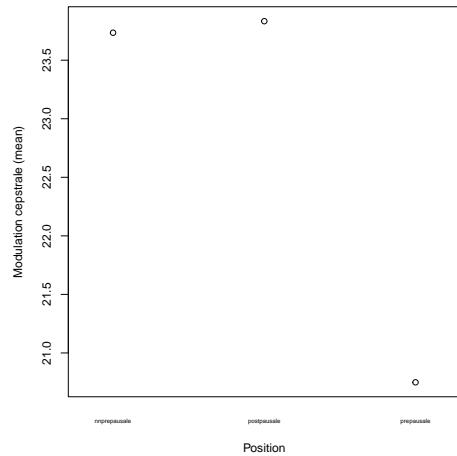
(a) `plotLMER.fnc()` pour la dispersion de la modulation cepstrale en fonction de la position (`nnpause` = milieu de séquence) du diphone dans la séquence, corpus NCCFr.



(b) `plotLMER.fnc()` pour la dispersion de la modulation cepstrale en fonction de la position (`nnpause` = milieu de séquence) du diphone dans la séquence, corpus Ester.



(c) `plotLMER.fnc()` pour la moyenne de la modulation cepstrale en fonction de la position (`nnpause` = milieu de séquence) du diphone dans la séquence, corpus NCCFr.



(d) `plotLMER.fnc()` pour la moyenne de la modulation cepstrale en fonction de la position (`nnpause` = milieu de séquence) du diphone dans la séquence, corpus Ester.

FIGURE 19 – `plotLMER.fnc()` pour la dispersion et la moyenne de modulation cepstrale en fonction de la position du diphone dans la séquence.

Concernant la modulation cepstrale, les trois positions sont très bien différenciées par la dispersion pour les deux corpus. La moyenne de modulation cepstrale remplit ce rôle de façon similaire pour NCCFr, mais peine à différencier les positions postpausale et intermédiaire pour Ester.

Nous pouvons utiliser un autre test statistique afin de comprendre les interactions entre position,

durée et modulation cepstrale : une ANOVA suivie d'un test de Tukey nous donnent l'interaction deux-à-deux des positions selon une variable. Par exemple : `aov(formula = duree ~ position, data = corpus)` suivie de `TukeyHSD()`. L'inconvénient est que l'ANOVA nous présente toujours les résultats comme significatifs car la taille des corpus est grande, mais le test de Tukey nous permet de voir les détails des interactions. Les résultats sont dans le Tableau 4 (NCCFr) et 5 (Ester).

Modalité	Position	Différence
Durée	postpausale-intermédiaire	1.400414
Durée	prepausale-intermédiaire	13.870364
Durée	prepausale-postpausale	12.469949
Dispersion modulation cepstrale	postpausale-intermédiaire	1.104
Dispersion modulation cepstrale	prepausale-intermédiaire	1.301
Dispersion modulation cepstrale	prepausale-postpausale	0.197
Moyenne modulation cepstrale	postpausale-intermédiaire	-1.388035
Moyenne modulation cepstrale	prepausale-intermédiaire	-3.064370
Moyenne modulation cepstrale	prepausale-postpausale	-1.676335

TABLE 4 – Test de Tukey pour les modalités de la durée et la dispersion de la modulation cepstrale selon la position du diphone dans la séquence, corpus NCCFr.

Modalité	Position	Différence
Durée	postpausale-intermédiaire	0.3166303
Durée	prepausale-intermédiaire	30.6075027
Durée	prepausale-postpausale	30.2908724
Dispersion modulation cepstrale	postpausale-intermédiaire	2.609
Dispersion modulation cepstrale	prepausale-intermédiaire	3.087
Dispersion modulation cepstrale	prepausale-postpausale	0.478
Moyenne modulation cepstrale	postpausale-intermédiaire	0.06816152
Moyenne modulation cepstrale	prepausale-intermédiaire	-2.95744741
Moyenne modulation cepstrale	prepausale-postpausale	-3.02560893

TABLE 5 – Test de Tukey pour les modalités de la durée et la dispersion de la modulation cepstrale selon la position du diphone dans la séquence, corpus Ester.

On voit que, bien que toutes les positions soient significatives entre elles, les positions postpausale et intermédiaires ne sont significativement différentes que de 1ms dans NCCFr et moins de 1ms dans Ester : cela n'est pas une différence perceptible par un humain, et nous considérons ce résultat comme non significatif, humainement parlant. Ces deux positions se démarquent très bien de la position prépausale, plus de 10ms plus longue en moyenne que les deux autres pour NCCFr et 30ms pour

Ester.

Concernant la dispersion de modulation cepstrale, celle-ci nous fait parvenir un autre rapprochement : cette fois-ci, ce sont les positions prépausale et postpausale qui sont proches l'une de l'autre (0,2 unité pour NCCFr et 0,5 pour Ester), et très éloignées de la position intermédiaire qui est plus basse d'une unité pour NCCFr et environ 3 pour Ester. La moyenne de modulation cepstrale fait, pour Ester comme plus haut pour le modèle mixte Figure 19d, état d'une différence moins grande entre les positions postpausale et intermédiaire que celles-ci avec la position prépausale, bien que tous les résultats soient significatifs. Les résultats pour NCCFr mettent les trois positions à équidistance l'une de l'autre, comme on le voit sur la Figure 19c.

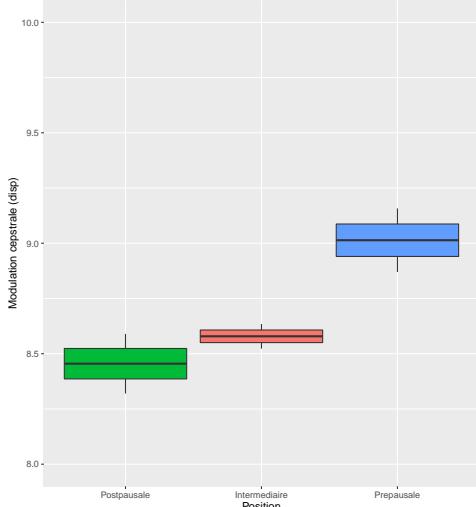
On voit que la différence entre diphones prépausals et postpausals est importante selon la durée du diphone, pour les deux corpus. Cependant, en ce qui concerne la modulation cepstrale, ces deux positions ont très peu de différence entre elles : elles ont une modulation cepstrale proche mais une durée différente – les diphones postpausals sont plus courts, et non allongés comme les diphones prépausals.

5.1.1 Analyse d'un sous-ensemble de diphones

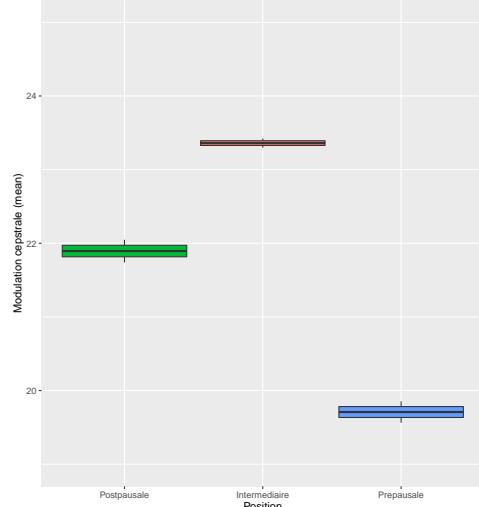
Nous avons vu que le diphone provoquait des variations dans la modélisation de nos données, il peut donc être intéressant de faire une étude plus spécialisée sur certains d'entre eux. Le corpus NCCFr étant plus petit, nous nous basons sur ses contraintes afin de sélectionner les diphones à étudier. Nous filtrons ainsi selon six des diphones les plus fréquents dans chaque position (au moins 2000 occurrences) : /pa, la, sa, se, me, ty/ – les autres diphones étant souvent dans des proportions moindres selon la position dans la séquence notamment, cela donnait des résultats peu représentatifs du corpus.

Nous avons d'abord observé les durées (voir Figure 20c) : il n'y a pas de différence significative entre elles d'après une ANOVA ($p > 0,05$).

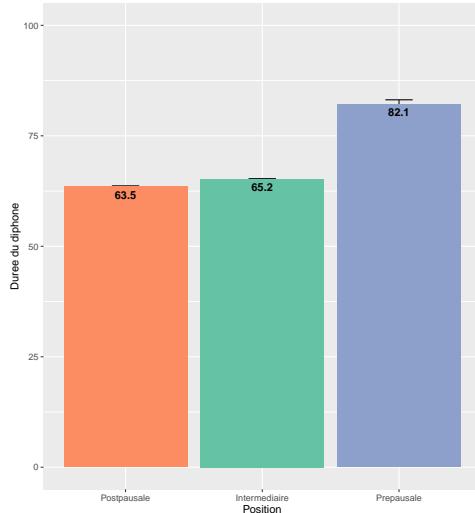
Une ANOVA sur les modulations cepstrales révèle que, cette fois-ci, la moyenne montre des interactions significatives ($p < 0,001$) – bien plus que la dispersion qui ne possède pas d'interaction significative ($p > 0,05$). Cela est visible Figures 20a et 20b.



(a) Dispersion de la modulation cepstrale pour le sous-corpus issu de NCCFr.



(b) Moyenne de la modulation cepstrale pour le sous-corpus issu de NCCFr.

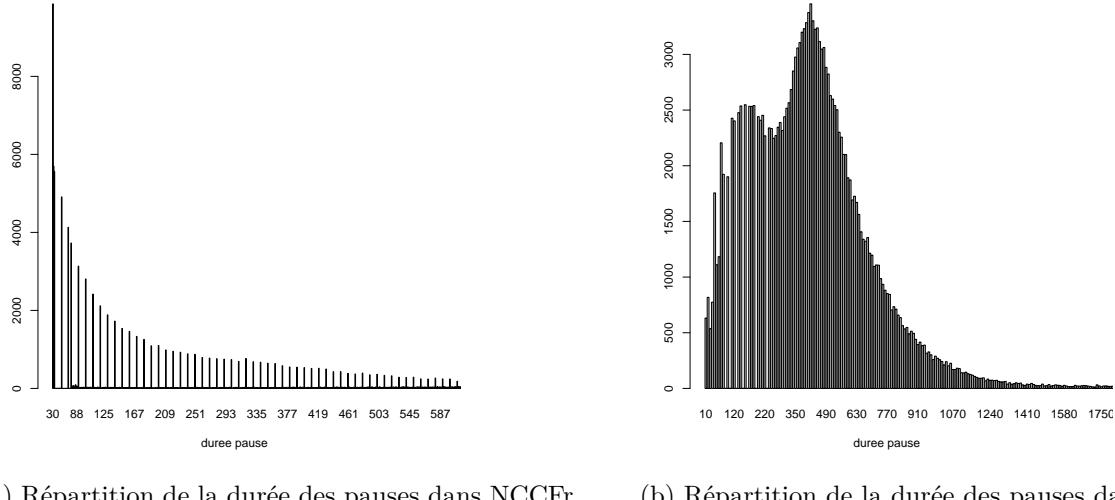


(c) Durée des diphones du sous-corpus issu de NCCFr.

FIGURE 20 – Modulations cepstrales et durée pour le sous-corpus issu de NCCFr.

Toutefois, nous avons décidé d'approfondir nos recherches concernant ces deux mesures en investiguant la durée des pauses pour les diphones pré- et postpausals. Sur R, nous avons étudié la répartition de la durée des pauses pour NCCFr d'abord : la moyenne de durée des pauses est de 1812ms alors que sa médiane est de 403ms – les données sont réparties de façon très inégale. Cela se confirme en créant un barplot, Figure 21a. On voit que 100ms délimite les pauses courtes – extrêmement nombreuses dans le corpus mais qui constituent des pauses trop courtes pour terminer une séquence –, des pauses plus longues. C'est donc une bonne borne inférieure pour notre limite de durée de pauses de séquence. Au contraire, les durées d'Ester sont bien mieux réparties (Figure

21b) : le nombre de pauses ne monte jamais au-dessus de 4000 pour une durée fixée contrairement à NCCFr. Nous lui appliquons aussi la borne de 100ms – si nous mettons la borne inférieure au-delà, il y a trop peu d'occurrences pour avoir un échantillon significatif.



(a) Répartition de la durée des pauses dans NCCFr. (b) Répartition de la durée des pauses dans Ester.

FIGURE 21 – Répartition de la durée des pauses dans les deux corpus.

Nous avons décidé de faire un subset de nos corpus en ne gardant que les pauses supérieures à 100ms. Cette fois-ci, en effectuant de nouveau une ANOVA, les données montrent que la dispersion de la modulation cepstrale est plus efficace pour différencier significativement les positions des diphones ($p = 0,00067$) que la moyenne de la modulation cepstrale ($p > 0,05$) : les pauses courtes, annotées comme des pauses par le corpus mais n'étant pas des pauses de l'ampleur de pauses de séquence, jouaient en sa défaveur et cette mesure fonctionne mieux avec les pauses longues. L'annotation des pauses dans les corpus sera étudiée dans la partie [Discussion](#). La dispersion de la modulation cepstrale est donc une meilleure mesure à utiliser dans le corpus.

Enfin, nous avons vérifié la répartition des mots du corpus afin de s'assurer que les mesures soient significatives pour tous les diphones. /pa/ et /sa/ possédaient tous deux des mots surreprésentés qui rendaient leurs interactions en termes de durée moins significatives – « pas » et « ça ». De plus, /ty/ est presque entièrement composé du mot « tu », et possède une durée prépausale moins longue que le reste du corpus. Ces cas mis à part, le reste du corpus est significatif face aux tests appliqués.

5.1.2 Impact des hésitations

Les hésitations ont été précédemment retirées des données étudiées car elles étaient considérées comme des diphones à part entière et allongeaient la mesure de durée, mais il peut être intéressant de voir si elles impactent le comportement d'autres diphones. Nous nous intéressons ici à l'interaction entre hésitations et durée d'un diphone pré- ou postpausal.

Nous avons séparé les deux corpus en prépausal et postpausal, puis en position d'hésitation (précédente ou suivante) afin d'avoir tous les cas possibles. Les hésitations sont constituées de *fillers* « euh » mais aussi « bah » ou « hm ». Nous n'avons pas trouvé d'hésitation précédant un diphone postpausal.

Pour le corpus NCCFr, les diphones en frontière de séquence précédés par une hésitation n'ont pas une durée différente des autres. Cependant, ces diphones suivis d'une hésitation ont une durée plus longue de 30ms en moyenne : il y a donc un lien entre la durée de diphones en frontière de séquence et la présence d'une hésitation après eux. L'impact des hésitations suivant un diphone en positions postpausale ou prépausale est visible Figure 22.

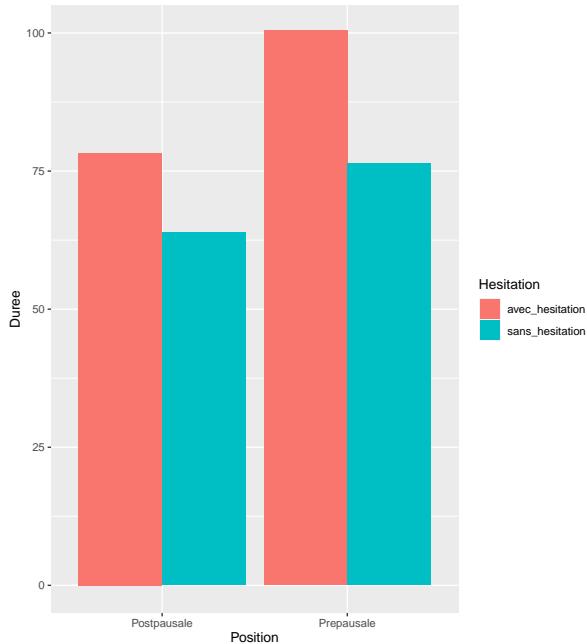


FIGURE 22 – Impact des hésitations suivant un diphone en positions postpausale ou prépausale, corpus NCCFr.

En ce qui concerne le corpus Ester, les diphones en frontière de séquence sont plus longs quand ils sont précédés d'hésitations (20ms de plus), mais aussi quand ils en sont suivis (plus de 50ms de

plus). Cependant, ce corpus ne comporte que quelques centaines d'occurrences pour ces cas, et est donc très peu représentatif.

Résumé

- Les deux corpus ont une durée plus importante pour le diphone prépausal, et une modulation cepstrale plus haute pour les diphones prépausals et postpausals.
- La modulation cepstrale permet donc de différencier les diphones postpausals et intermédiaires, alors que la durée ne distingue que le diphone prépausal des deux autres.
- Le niveau prosodique du mot est aussi touché par ce phénomène, bien que plus faiblement.
- Les durées des diphones du corpus journalistique sont plus élevées que celles du corpus de parole spontanée.
- La mesure de modulation cepstrale basée sur la dispersion de celle-ci est plus performante que celle basée sur sa moyenne lorsque les pauses sont plus longues que 100ms.

5.2 Durée et modulation cepstrale selon les parties du discours

Nous rappelons notre seconde hypothèse :

(ii) Les différences de durée et de modulation cepstrale entre éléments initiaux et finaux des constituants prosodiques du français sont éprouvées peu importe la catégorie grammaticale du diphone. Nous prévoyons toutefois que les mesures aient des performances moins soutenues sur les catégories des mots grammaticaux (pronoms, déterminants, etc...).

L'analyse des sous-corpus révèle que certains mots peuvent perturber les résultats obtenus, et que ceux-ci sont souvent des mots grammaticaux très usités. De nombreuses études traitent régulièrement les mots grammaticaux à part sous couvert qu'ils seraient moins impactés par les allongements prépausals. Analyser les résultats selon les parties du discours (catégories grammaticales) des corpus, afin de voir si la tendance observée dans la section précédente est généralisée ou non, est notre objectif dans cette sous-partie.

Nous séparons nos corpus en quatre sous-corpus contenant chacun les mots lexicaux ou grammaticaux des corpus respectifs. Les sous-corpus de mots lexicaux sont composés des parties du discours suivantes : Nom, Verbe, Adjectif, Adverbe, Nom Propre, Abbréviation, et Interjection. Les

sous-corpus de mots grammaticaux sont composés des parties du discours suivantes : Adposition, Auxiliaire, Conjonction de coordination, Déterminant, Numéral, Participe, Pronom, et Conjonction de subordination. Nous compilons les moyennes des résultats dans les Tableau 6 (mots lexicaux) et 7 (mots grammaticaux).

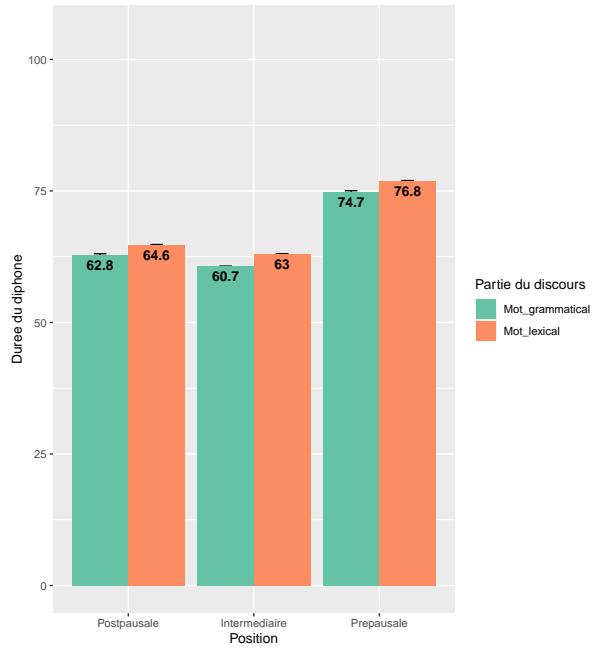
Modalité	Position	Occurrences NCCFr	NCCFr	Occurrences Ester	Ester
<i>Durée</i>					
Niveau Séquence	Début	22221	65	19776	75
	Inter	440089	63	1170492	71
	Fin	39353	78	89554	102
	Mono	8838	79	1053	104
Niveau Mot	Début	84393	63	224763	72
	Inter	173436	61	656937	69
	Fin	75395	71	190186	83
	Mono	198523	67	228254	81
<i>Modulation cepstrale</i>					
Niveau Séquence	Début		8.9		10.2
	Inter		8.5		9.4
	Fin		9.5		13.4
	Mono		8.7		13.3
Niveau Mot	Début		8.4		9.2
	Inter		8.2		9.1
	Fin		9.3		11.3
	Mono		8.8		10.7

TABLE 6 – Tableau contenant les durées et modulations cepstrales de dipones appartenant au même mot lexical dans NCCFr et Ester, dans différentes structures prosodiques.

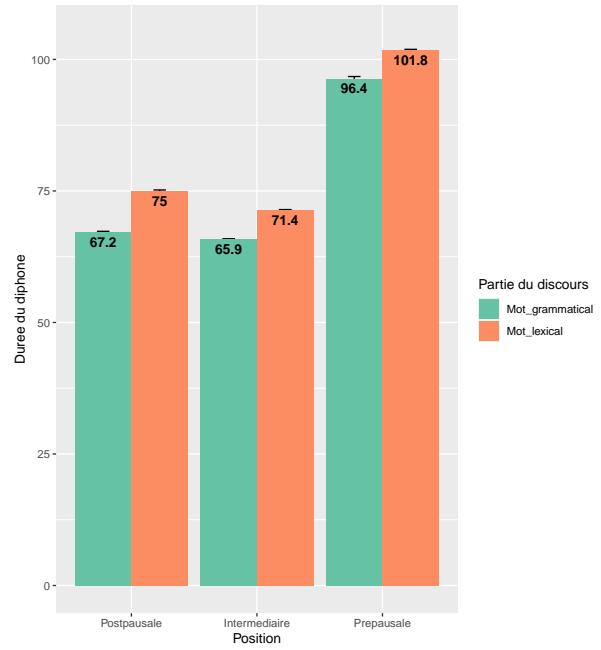
Modalité	Position	Occurrences NCCFr	NCCFr	Occurrences Ester	Ester
<i>Durée</i>					
Niveau Séquence	Début	16668	63	30826	66
	Inter	126013	61	273902	66
	Fin	8632	79	13418	95
	Mono	3065	75	2341	106
Niveau Mot	Début	11324	66	24811	73
	Inter	9774	68	41241	68
	Fin	7704	68	19010	74
	Mono	136442	61	251241	66
<i>Modulation cepstrale</i>					
Niveau Séquence	Début		8.6		9.6
	Inter		7.8		8.5
	Fin		9.4		11.4
	Mono		8.6		12.6
Niveau Mot	Début		8.1		9.1
	Inter		8.4		8.4
	Fin		8.4		8.5
	Mono		8		8.8

TABLE 7 – Tableau contenant les durées et modulations cepstrales de diphones appartenant au même mot grammatical dans NCCFr et Ester, dans différentes structures prosodiques.

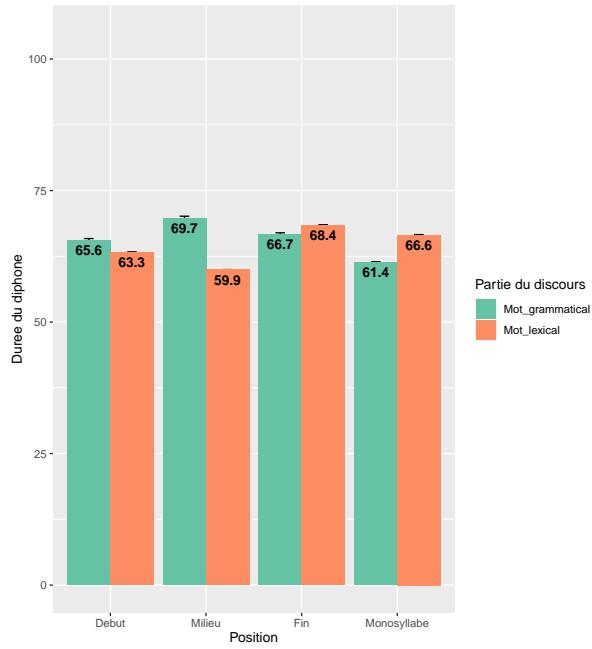
Nous commençons par obtenir des figures des données.



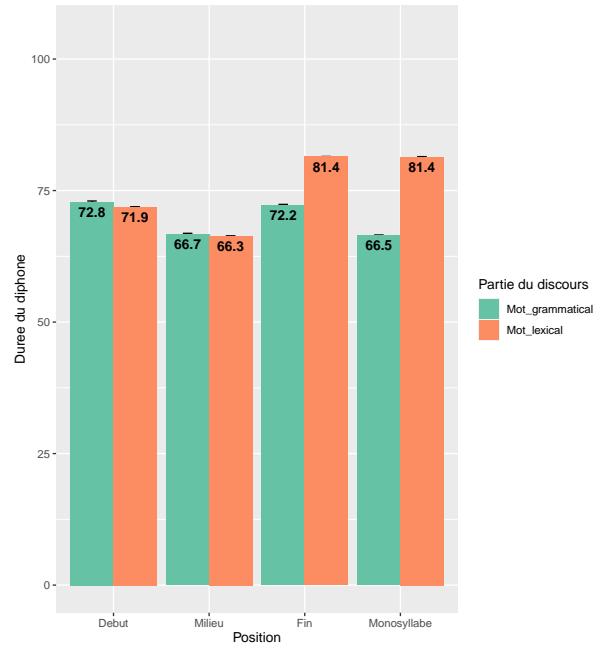
(a) Durée du diphone selon sa partie du discours et sa position dans une séquence du corpus NCCFr : après une pause, dans la séquence, et avant une pause.



(b) Durée du diphone selon sa partie du discours et sa position dans une séquence du corpus Ester : après une pause, dans la séquence, et avant une pause.

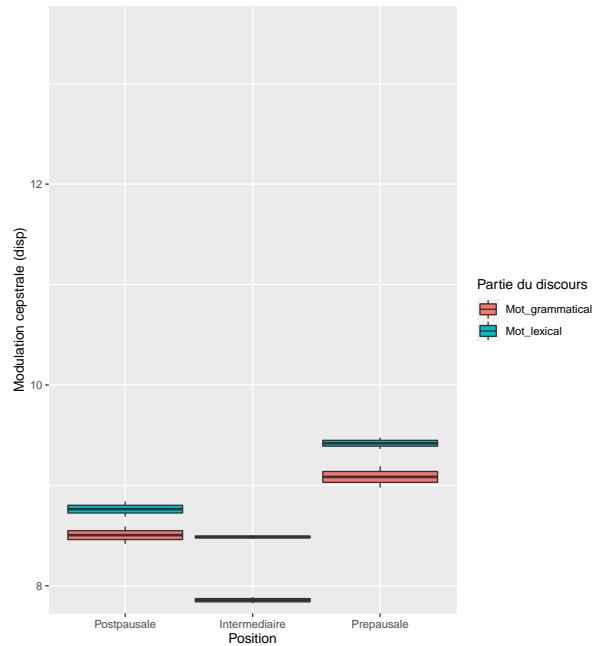


(c) Durée du diphone selon sa partie du discours et sa position dans un mot du corpus NCCFr : au début du mot, au milieu, à la fin ou dans un mot monosyllabique.

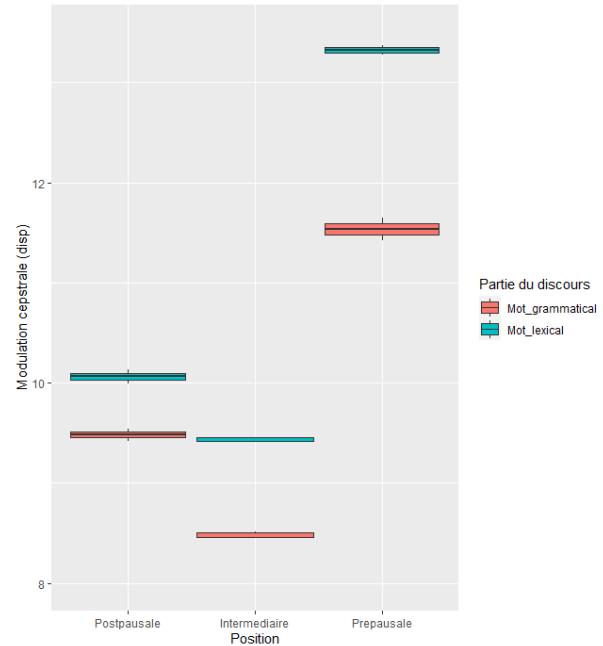


(d) Durée du diphone selon sa partie du discours et sa position dans un mot du corpus Ester : au début du mot, au milieu, à la fin ou dans un mot monosyllabique.

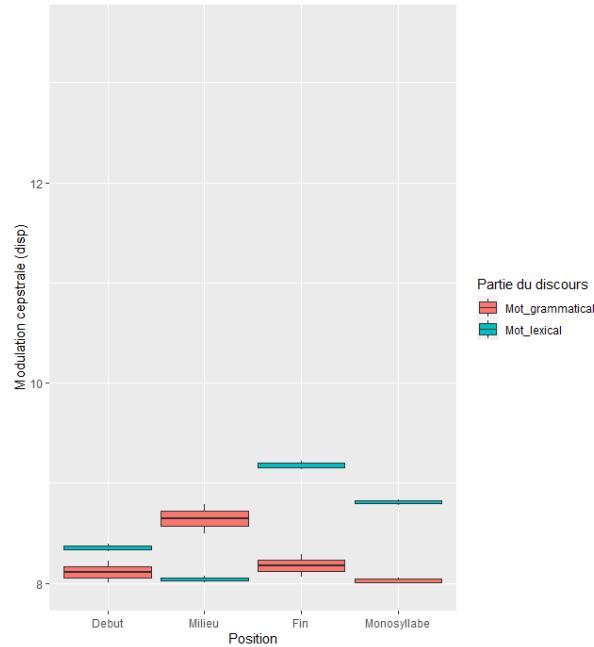
FIGURE 23 – Durée du diphone selon sa position dans une frontière prosodique, et sa partie du discours.



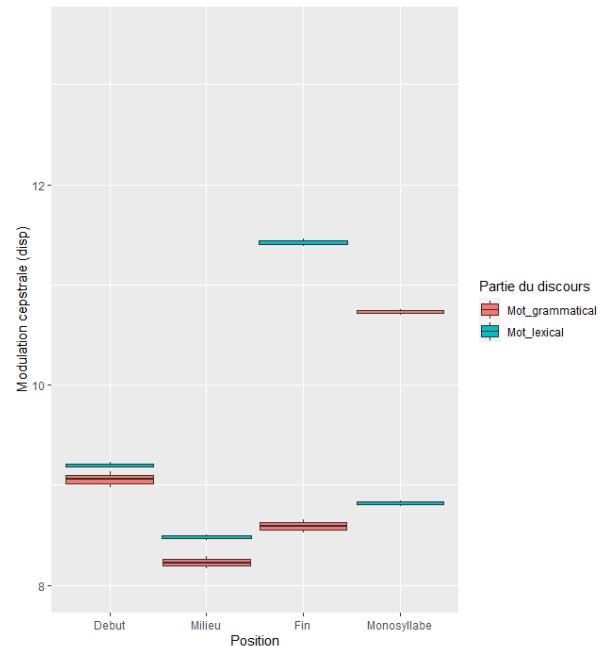
(a) Modulation cepstrale du diphone selon sa position dans une séquence du corpus NCCFr : après une pause, dans la séquence, et avant une pause.



(b) Modulation cepstrale du diphone selon sa position dans une séquence du corpus Ester : après une pause, dans la séquence, et avant une pause.



(c) Modulation cepstrale du diphone selon sa position dans un mot du corpus NCCFr : au début du mot, au milieu, à la fin ou dans un mot monosyllabique.



(d) Modulation cepstrale du diphone selon sa position dans un mot du corpus Ester : au début du mot, au milieu, à la fin ou dans un mot monosyllabique.

FIGURE 24 – Modulation cepstrale du diphone selon sa position dans une frontière prosodique, et sa partie du discours.

Les données suivent les observations que nous avions effectuées dans la sous-partie précédente : les diphones prépausals sont plus longs que les postpausals et les intermédiaires. La modulation cepstrale est la mesure qui permet de différencier ces deux derniers. Cependant, dans le Tableau 7 des mots grammaticaux, on voit que le niveau du mot ne possède pas de différence de durée ou de modulation cespstrale (celle de Ester pour le début de mot est même plus élevée que la fin de mot) : des ressemblances entre les deux catégories de parties du discours cachent donc des phénomènes que nous allons explorer.

Sur les Figures de durée 23, les mots grammaticaux et lexicaux suivent la même tendance même si les premiers sont souvent plus courts. On voit sur les Figures de modulation cepstrale 24 que pour la séquence, les parties du discours lexicales et grammaticales sont proches, mais pour les fins de mots les mots grammaticaux ont des diphones moins bien articulés.

Afin de tester la significativité des résultats, nous effectuons une ANOVA suivie d'un test de Tukey sous la forme `aov(formula = duree ~ position * PoS, data = corpus) > TukeyHSD()`.

Sur le Tableau 14 en Annexe, on voit que toutes les interactions entre position dans la séquence et partie du discours sont significatives. Cependant, il faut garder à l'esprit qu'elles peuvent être dues à la taille du corpus.

Quelques résultats dans le mot ne sont pas significatifs, mais concernent une minorité de cas : on y compte les diphones dans des mots monosyllabiques et les diphones en fin de mots par exemple ($p > 0,1$ pour la durée comme pour la modulation cepstrale).

Si l'on veut étudier les différences entre parties du discours, il va nous falloir effectuer une analyse plus fine.

5.2.1 Analyse d'un sous-ensemble de diphones

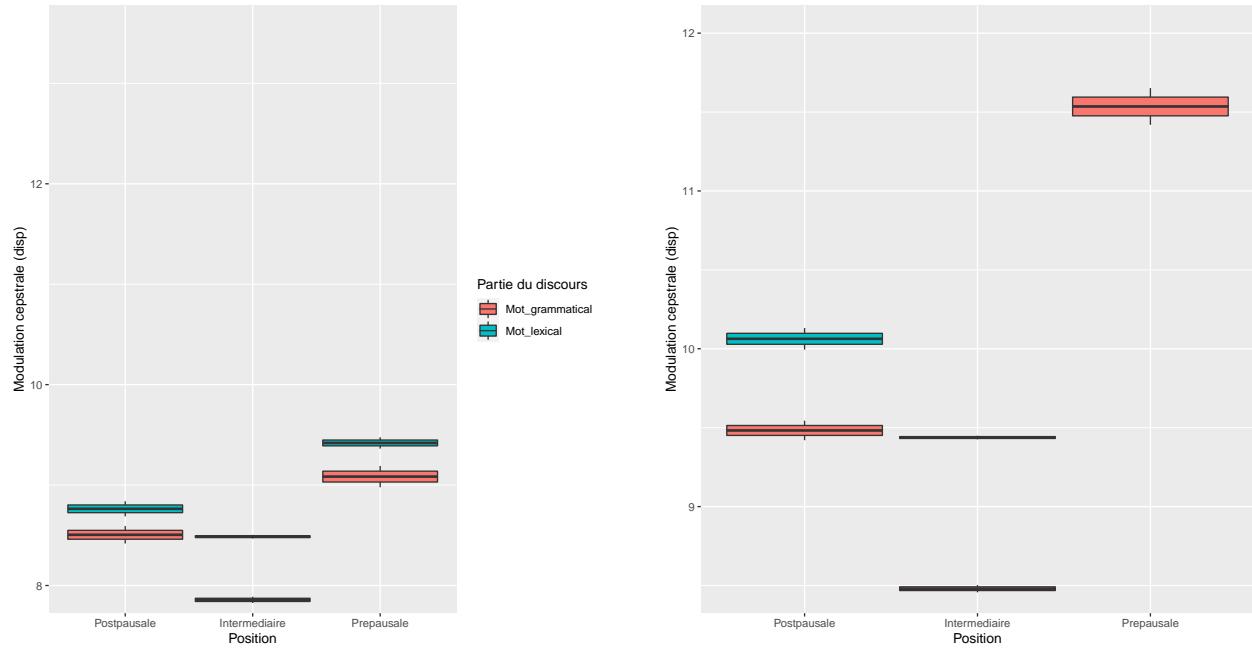
Pour cette section, nous prenons le sous-corpus composé de six diphones fréquents du corpus NCCFr.

Position	Tout	Verbe	Nom	Adverbe	Pronom
Durée postpausale	63	+	++	+	+
Durée intermédiaire	64	+	++	=	=
Durée prépausale	72	+	++	++	-
Modulation cepstrale (disp) postpausale	9,059	++	++	+	+
Modulation cepstrale (disp) intermédiaire	8,757	++	++	++	+
Modulation cepstrale (disp) prépausale	9,036	++	++	++	-

TABLE 8 – Aperçu de la durée et la modulation cepstrale des diphones en frontière de séquence, y compris pour les parties du discours Verbe, Nom, Adverbe et Pronom.

Nous commençons par regarder un aperçu des catégories grammaticales comportant le plus d'occurrences dans les corpus : les noms, les verbes, les adverbes et les pronoms. Nous comparons les moyennes des durées, pour chaque position des diphones /pa, la, sa, sɛ, mɛ, ty/ dans la séquence, des parties du discours par rapport aux moyennes globales. On voit que les pronoms sont moins longs et plus coarticulés que les autres parties du discours, bien plus hautes que les moyennes globales. Mais il peut être intéressant d'étudier les pronoms représentés par les diphones dans le sous-corpus. Deux des diphones ne donnaient pas lieu à des pronoms : /pa/ et /mɛ/ ; les mots représentés par /la, sa, sɛ, ty/ étaient presque uniquement « la », « ça », « c'est » et « tu ».

Nous avons comparé la durée et la modulation cepstrale de la catégorie grammaticale entière des pronoms avec celle des quatre diphones précédents, et ces premières sont plus longues que celles de nos quatre diphones : il y a donc un élément qui fait baisser les mesures. La Figure 25 montre cette différence de durées. Après comparaison des durées entre les mots les plus fréquents, il s'avère que c'est « c'est » qui faisait baisser ces mesures avec une moyenne de 44ms, ce qui est très bas pour notre corpus : en l'enlevant, les mesures rejoignent celles du corpus entier des pronoms. Certains pronoms très répandus peuvent donc impacter les mesures du corpus.



(a) Durée des pronoms pour le sous-corpus de NCCFr.

(b) Durée des pronoms pour le corpus entier de NCCFr.

FIGURE 25 – Durée des pronoms pour le sous-corpus et le corpus entier de NCCFr.

Concernant le corpus Ester, nous avons essayé de reproduire l’expérience, mais n’avons pas trouvé de différence entre « c’est » et le reste des pronoms. Nous souhaitons cependant remarquer que les pronoms les plus courants ne sont pas les mêmes dans ce corpus, et il est normal que les résultats ne soient pas retrouvés. Par exemple « cela » est le deuxième pronom le plus fréquent du sous-corpus de Ester, alors que ce pronom n’est mentionné que six fois dans NCCFr. Inversement, « tu » est le pronom le plus fréquent de NCCFr, alors qu’il est bien moins fréquent dans Ester. De ce fait, le diphone /ty/ n’a pas le même impact dans NCCFr et Ester. Il y a également du bruit dans les pronoms de Ester : des mots tels que « sarkozy » ou « mais » ont été classés comme des pronoms par l’annotateur syntaxique. Toutefois, ils ne représentent qu’une petite partie des données, seulement 0,08% du sous-corpus de Ester.

La deuxième partie de cette étude de sous-ensembles des corpus s’intéresse à la présence de « tu » dans le sous-corpus de NCCFr. En effet, alors que nous nous attendions à observer un phénomène de réduction sur ce pronom car il est très utilisé, nous avons vu que sa durée et sa modulation cepstrale étaient semblables aux autres pronoms mis à part « c’est ». Nous souhaitons voir si ce pronom est bien affecté par l’hypertarticulation prépausale au même titre que les autres.

Une mesure de moyenne sur la durée nous indique que le « tu » prépausal est plus long que le non prépausal de 7ms. Cependant, cette différence est considérée comme significative par une ANOVA effectuée sur le corpus ($p = 0,0067747$). La modulation cepstrale, elle, n'a pas de différence significative entre les différentes positions ($p = 0,1611788$). Le « tu », bien que faiblement, est tout de même l'objet d'un phénomène d'allongement prépausal dans le corpus NCCFr.

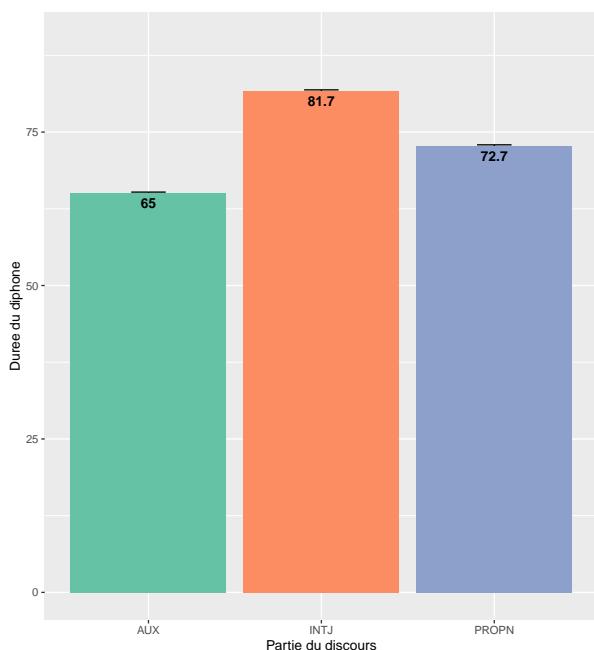
5.2.2 Étude de quelques parties du discours

Dans cette partie, nous étudions des parties du discours particulières, dont la durée et/ou la modulation cepstrale sont incongruentes aux observations effectuées jusqu'ici. Quelques remarques tout d'abord au sujet de la distribution des parties du discours dans la séquence : on retrouve nombre de déterminants, noms et adpositions dans la catégorie des mots monosyllabiques ; cela est attendu car, mis à part les noms qui sont très courants en toutes positions, les deux autres parties du discours sont souvent très courtes.

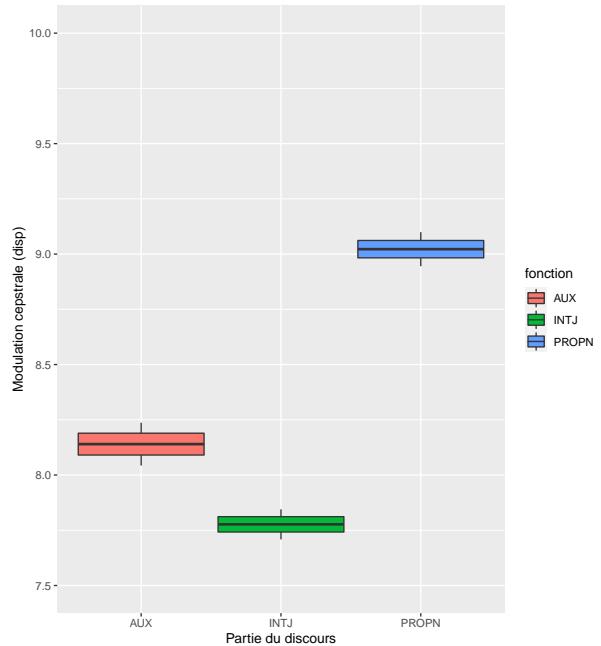
Dans cette catégorie prépausale, la partie du discours *nom* allonge considérablement la durée ; si on l'enlève, la durée et la modulation cepstrale baissent – c'est là encore un phénomène attendu. Cependant, la durée de déterminants fréquents (« le », « la », « un »...) et leur modulation cepstrale ne sont pas particulièrement plus basses que les autres déterminants. La seule exception est « un » : il est plus long et mieux articulé que les autres déterminants avec une durée moyenne de 98ms et une dispersion de modulation cepstrale moyenne de 11.

Nous avons remarqué un effet de l'appartenance d'un diphone à un même mot ou deux mots différents avec la partie du discours des conjonctions de coordination : leur durée était plus élevée que celle des verbes, noms, adverbes et pronoms. Enlever les mots les plus fréquents comme « mais » et « et » ne faisait qu'augmenter leur durée. Cependant, cela est dû au fait que les diphones n'appartenaient pas au même mot : quand on ne sélectionne que ceux appartenant à un même mot, leur durée baisse. Cet effet est présent pour les deux corpus.

Trois autres parties du discours ont un comportement intéressant ; nous l'avons représenté sur la Figure 26.



(a) Durée de quelques parties du discours de NCCFr.



(b) Dispersion de la modulation cepstrale de quelques parties du discours de NCCFr.

FIGURE 26 – Durée et dispersion de la modulation cepstrale de quelques parties du discours du corpus NCCFr.

Les interjections ont une moyenne de durée plus haute que les valeurs des verbes, noms, adverbes et pronoms, mais une modulation cepstrale plus basse. La durée est principalement augmentée par l'hésitation « euh » : pour les hésitations longues, « euh » est l'interjection privilégiée. Ce résultat n'est pas retrouvé dans Ester car ce corpus ne comporte que très peu d'hésitations, dû à son caractère scripté.

Pour NCCFr uniquement, les auxiliaires ont des relevés bas pour la moyenne de durée comme celle de modulation cepstrale – étant des mots très employés et plutôt courts, c'est un résultat attendu surtout dans un corpus spontané. On retrouve notamment « était » ou d'autres conjugaisons de « être » parmi les mots les plus fréquents, mais les enlever ne change rien aux mesures : ce ne sont pas eux qui portent la réduction observée mais bien la partie du discours dans son ensemble.

Enfin, la dernière partie du discours étudiée ici est le nom propre. En effet, ceux-ci ont une durée moyenne plus importante que les mesures étalons utilisée jusqu'ici (verbes, noms, adverbes et pronoms) alors que leur modulation cepstrale n'est pas plus haute pour autant. Aucun mot particulier n'influence ces mesures : les noms propres sont une catégorie très diverse. Les dépendances syntaxiques les plus saillantes comme les compléments d'objet et les displocations n'ont pas d'influence non plus.

Résumé

- Les mots grammaticaux comme les mots lexicaux subissent le phénomène précédemment décrit : les diphones prépausals et postpausals ont une modulation cepstrale plus importante, et les diphones prépausals sont plus longs.
- La durée des mots grammaticaux en début de mot est même plus longue que celle des mots lexicaux.
- Il est difficile de comparer certaines parties du discours entre les deux corpus car les façons de parler de leurs locuteurs sont très distinctes (style soutenu pour Ester).
- Les noms propres ont une modulation cepstrale plus haute que la normale ; les interjections ont par contre une modulation cepstrale moyenne mais une durée plus longue, surtout l'hésitation « euh ». Dans NCCFr, les auxiliaires subissent une forte réduction.

5.3 Impact du type de syllabe

Voici notre hypothèse pour cette partie :

(iii) La forme de la syllabe (CV, VC...) a un impact sur les mesures de durée et de modulation cepstrale car le phonème touché par les contraintes prosodiques de la frontière de constituant n'est pas le même.

Dans cette sous-partie, nous nous intéressons à l'impact du type de syllabe sur les données des corpus : syllabe ouverte de la forme (C)V et syllabe fermée de la forme V(C)C. Notre but est de voir si le phone recevant l'allongement est toujours la voyelle, ou si la consonne peut aussi être facteur de variation. Ainsi, étant donné que nous avons les mesures de durée pour chaque phone d'un diphone, nous pouvons nous positionner au niveau du phone dans les résultats. Ci-dessous sont les Tableaux correspondant aux durées de la voyelle et de la consonne pour des diphones dans une syllabe ouverte (9) ou fermée (10). La modulation cepstrale n'est ici pas accessible puisqu'elle ne peut pas se mesurer sur un phone seul, seulement sur la transition entre deux phones.

Modalité	Position	Occurrences NCCFr	NCCFr	Occurrences Ester	Ester
<i>Durée</i>	<i>Voyelle</i>				
Niveau Séquence	Début	23005	62	25660	69
	Inter	236752	63	532204	71
	Fin	22307	94	37434	124
	Mono	5536	81	1857	134
Niveau Mot	Début	42738	59	114626	62
	Inter	43982	62	196593	72
	Fin	47080	79	98344	94
	Mono	171553	65	205321	76
<i>Durée</i>	<i>Consonne</i>				
Niveau Séquence	Début		66		70
	Inter		70		78
	Fin		77		95
	Mono		70		81
Niveau Mot	Début		73		86
	Inter		66		73
	Fin		72		84
	Mono		70		78

TABLE 9 – Tableau contenant les durées de phones (syllabe ouverte (C)V) appartenant au même mot dans NCCFr et Ester, dans différentes structures prosodiques.

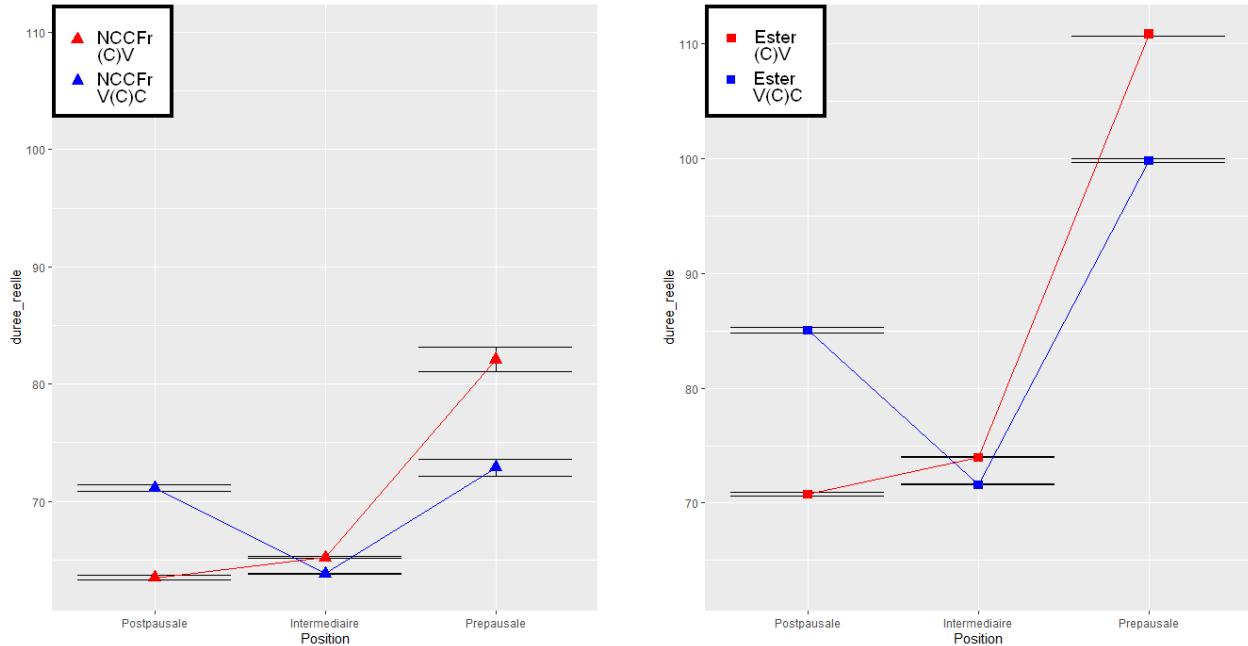
Modalité	Position	Occurrences NCCFr	NCCFr	Occurrences Ester	Ester
<i>Durée</i>	<i>Voyelle</i>				
Niveau Séquence	Début	5562	62	11497	67
	Inter	156833	65	443655	74
	Fin	10556	81	41380	87
	Mono	448	65	586	87
Niveau Mot	Début	29866	68	73269	79
	Inter	69932	67	248999	76
	Fin	16877	67	62912	74
	Mono	63240	63	119353	69
<i>Durée</i>	<i>Consonne</i>				
Niveau Séquence	Début		63		78
	Inter		58		64
	Fin		77		115
	Mono		74		109
Niveau Mot	Début		59		65
	Inter		58		62
	Fin		69		90
	Mono		61		76

TABLE 10 – Tableau contenant les durées de phones (syllabe fermée V (C) C) appartenant au même mot dans NCCFr et Ester, dans différentes structures prosodiques.

Dans le Tableau 9 des syllabes ouvertes, on voit que la consonne est plus longue que la voyelle, sauf dans les cas de fin de séquence ou fin de mot : la voyelle est alors plus longue. L'allongement final est donc surtout effectué sur la voyelle dans une syllabe ouverte.

Dans le Tableau 10 des syllabes fermées, ce phénomène est moins présent : la voyelle est toujours celle qui porte l'allongement lorsque l'on est en fin de groupe pour NCCFr, mais pas pour Ester. De plus, au niveau du mot, c'est la consonne qui porte l'allongement, à tel point que la voyelle est légèrement plus courte. Dans les deux cas – syllabe ouverte ou fermée –, la rime est allongée. Mais la répartition de la durée entre consonne et voyelle est différente.

Un phénomène intéressant a été observé en construisant des graphiques des données, Figure 27.



(a) Durée du diphone dans une syllabe ouverte ou fermée, selon sa position dans une séquence du corpus NCCFr : après une pause, dans la séquence, et avant une pause.

(b) Durée du diphone dans une syllabe ouverte ou fermée, selon sa position dans une séquence du corpus Ester : après une pause, dans la séquence, et avant une pause.

FIGURE 27 – Durée du diphone selon sa position dans une frontière prosodique.

On voit qu'il y a un croisement des données des syllabes ouvertes et fermées dans les deux corpus – les syllabes ouvertes postpausales ont une durée inférieure à celles en milieu de séquence, alors que les syllabes fermées postpausales sont plus longues que leur équivalent intermédiaire. Par exemple, la voyelle fermée /et/ a une durée moyenne plus longue que /sa/ en position postpausale, mais est moins longue qu'elle en position prépausale. Les syllabes fermées ont donc une durée plus longue en position postpausale que les syllabes ouvertes.

5.3.1 Impact de la consonne

Nous avons vu que la consonne portait l'allongement prépausal des syllabes fermées, nous nous proposons maintenant de chercher si ces syllabes réagissent de la même façon à la position prépausale selon leur consonne. Concernant les occlusives sourdes, celles-ci vont de la même durée que la voyelle à une durée plus longue (notamment /ik/ et /et/, sans qu'ils soient influencés par un mot particulièrement fréquent). Les fricatives et les nasales ont aussi une durée plus longue que la voyelle en position prépausale, mais ce n'est pas le cas des occlusives voisées et des liquides. Ces observations sont exactement les mêmes pour les deux corpus.

Seules les occlusives sourdes, les fricatives et les nasales sont donc allongées dans une syllabe fermée prépausale. Dans la partie [Discussion](#) nous aborderons la segmentation des consonnes occlusives, qui peut avoir un impact sur les résultats.

5.3.2 Impact du nombre de syllabes

Enfin nous étudions l'impact du nombre de syllabes sur la durée de l'allongement final. Nous filtrons nos corpus pour ne garder que les diphones prépausals. Ensuite, nous séparons nos corpus en trois parties : les diphone prépausals appartenant à un mot court (jusqu'à trois syllabes), moyen (entre quatre et six syllabes) ou long (entre sept et neuf syllabes). Ensuite, nous mesurons leur durée moyenne. D'après nos résultats, la durée moyenne de diphones prépausals appartenant à un mot de moins de six syllabes est plus longue qu'un mot contenant entre sept et neuf syllabes. Il faut noter que NCCFr, en sa qualité de corpus de parole spontanée, comporte moins de mots prépausals de plus de sept syllabes comparé aux mots moins longs, par rapport à Ester.

Nous filtrons ensuite nos corpus pour différencier les mots lexicaux des mots grammaticaux – cela peut avoir un impact sur les résultats, il pourrait y avoir une différence plus marquée pour les mots lexicaux. Toutefois, il y a peu de mots grammaticaux prépausals de plus de sept syllabes dans les corpus – NCCFr en a 153 et Ester 556 ; ce faible nombre n'est pas surprenant car les mots grammaticaux sont généralement plus courts que les mots lexicaux. Concernant les résultats, les mots grammaticaux ont des durées plus courtes, mais le rapport des durées entre mots courts, moyens et longs est proche des mots lexicaux.

Résumé

- Les syllabes prépausales ouvertes et fermées subissent un allongement de la rime, mais la répartition de la durée dans celle-ci est différente.
- Voici un tableau récapitulatif des caractéristiques de chaque type de syllabe :

Syllabe ouverte	Syllabe fermée
Allongement sur la voyelle	<i>Si occlusive sourde, fricative ou nasale</i> ⇒ allongement sur la consonne <i>Sinon</i> ⇒ allongement réparti équitablement entre consonne et voyelle

- Les syllabes ouvertes postpausales ont une durée inférieure à celles en milieu de séquence, alors que les syllabes fermées postpausales sont plus longues que leur équivalent intermédiaire.

5.4 Impact des dépendances syntaxiques

5.4.1 Sujets et compléments d'objet

Cette dernière sous-partie porte sur l'apport des dépendances syntaxiques aux mesures utilisées dans cette étude. Nous n'allons étudier que certaines dépendances syntaxiques dont nous pensons qu'elles peuvent avoir un impact sur les données : dans un premier temps, les groupes nominaux sujets ou compléments d'objet d'une racine dans un énoncé. Notre hypothèse était :

(iv) *Les dépendances syntaxiques ont également un rôle à jouer dans les résultats : des mots porteurs d'une nouvelle information seront hyperarticulés comparés à ceux portant une information déjà connue (complément d'objet et sujet respectivement par exemple, dans le cas d'une phrase conservant la structure informationnelle).*

Sur la Figure 28, nous voyons un exemple d'énoncé issu du corpus Ester annoté en dépendances syntaxiques : on remarque notamment les dépendances *subj* (sujet) et *comp :obj* (complément d'objet) – ces derniers, dépendances du verbe, sont distincts des *mod* (modificateur) car ils ne sont employés que par des verbes transitifs alors que les modificateurs sont utilisés avec des verbes intransitifs et sont généralement des adverbes.

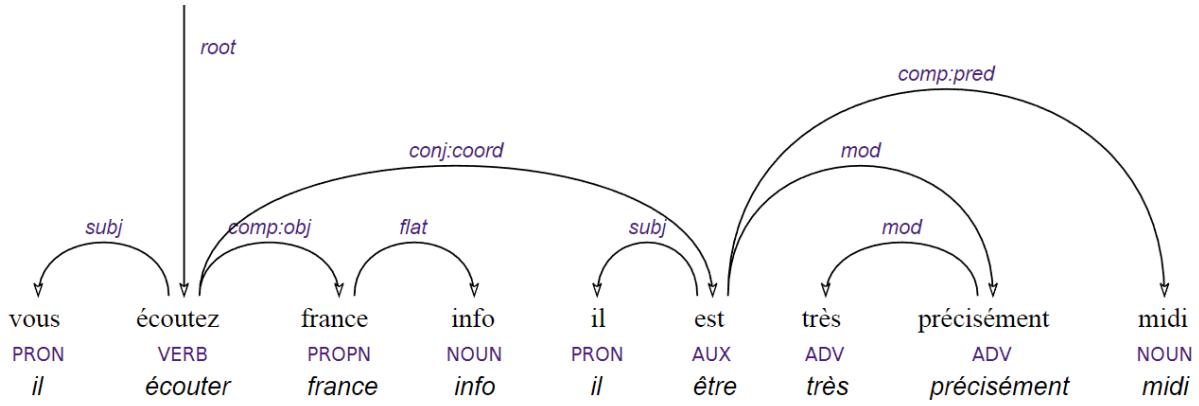
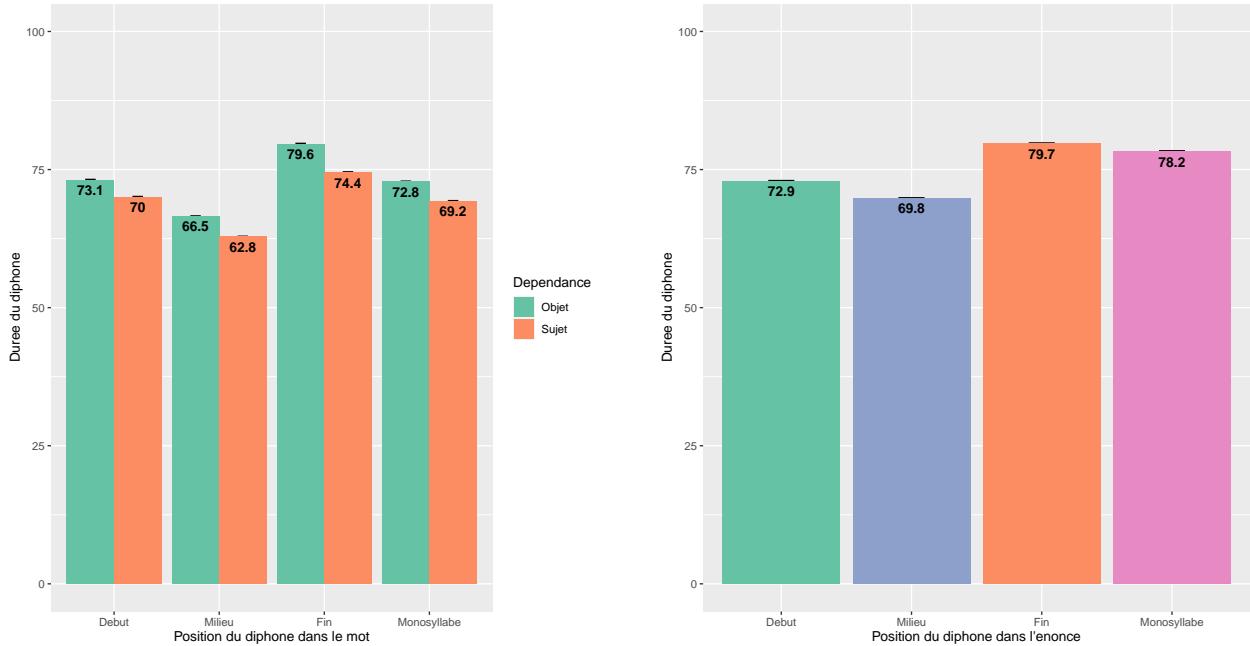


FIGURE 28 – Arbre des dépendances syntaxiques de l'énoncé « Vous écoutez France Info, il est très précisément midi. » construit avec Arborator (Guibon et al., 2020).

Les données sont filtrées afin de conserver les dépendances sujet et objet. On répertorie aussi le nombre de sujets et objets dans la phrase ainsi que la position du mot analysé dans la séquence du sujet ou de l'objet et la taille de l'énoncé. Les positions du diphone dans la séquence ne sont pas prises en compte car pour les sujets, il y a très peu de diphones prépausals (4000 contre 15000 postpausals, par exemple). D'abord, nous étudions les différences entre les deux dépendances.

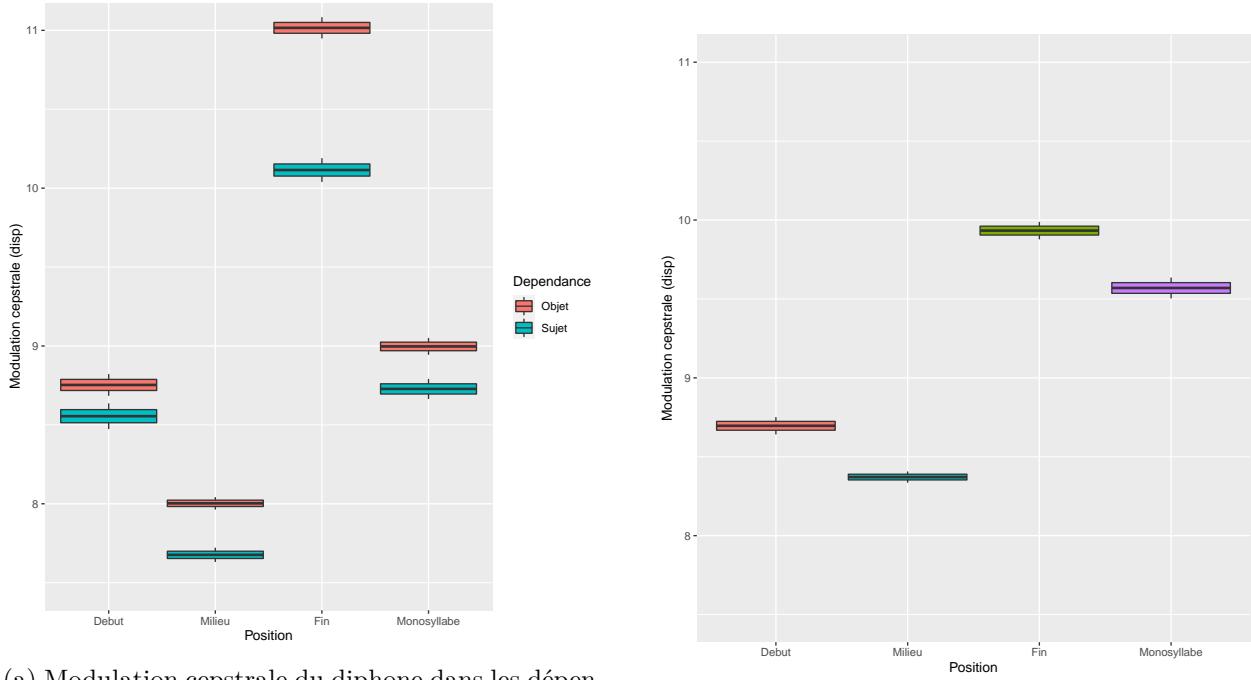


(a) Durée du diphone dans les dépendances *sujet* et *complément d'objet* du corpus Ester : au début du mot, au milieu, à la fin ou en monosyllabique.

(b) Durée du diphone dans la partie du discours *verbe* du corpus Ester : au début du mot, au milieu, à la fin ou dans un mot monosyllabique.

FIGURE 29 – Durée du diphone dans des dépendances syntaxiques selon sa position dans un mot.

La Figure 29a montre que les sujets et les objets ont des durées très proches pour les quatre positions possibles du diphone dans le mot, mais les compléments d'objet ont des durées plus longues pour toutes les positions. On peut la comparer avec celle des verbes afin de mettre ces valeurs en perspective : si l'on regarde la Figure 29b, on voit que les durées du complément d'objet sont très similaires à celles des verbes du corpus.



(a) Modulation cepstrale du diphone dans les dépendances *sujet* et *complément d'objet* du corpus Ester : au début du mot, au milieu, à la fin ou en monosyllabique.

(b) Modulation cepstrale du diphone dans la partie du discours *verbe* du corpus Ester : au début du mot, au milieu, à la fin ou dans un mot monosyllabique.

FIGURE 30 – Durée du diphone dans des dépendances syntaxiques selon sa position dans un mot.

La Figure 30a représente nos résultats pour la dispersion de la modulation cepstrale : on voit que les deux dépendances ont des résultats similaires comme pour la durée, avec la dépendance objet qui a des mesures plus hautes ; comparée aux mesures pour la catégorie des verbes (Figure 30b), c'est cette fois-ci la dépendance sujet qui lui ressemble. En effet, la mesure de modulation cepstrale pour la fin de mot est plus haute pour les objets.

Nous effectuons une ANOVA suivie d'un test de Tukey afin de comparer les deux dépendances entre elles.

$$\text{aov}(\text{duree} \sim \text{position} * \text{dependance}, \text{data}=\text{corpus}) \quad (5)$$

D'après les résultats, les deux dépendances sont souvent significativement différentes ($p < 0,001$), mais on retrouve quelques positions pour lesquelles ce n'est pas le cas : les diphones de syllabes de début de mot ne sont pas significativement différentes pour les deux dépendances ($p = 0,021$), ainsi que pour la syllabe de début de mot du sujet et celle de milieu de mot de l'objet ($p = 0,039$). Enfin, les syllabes finales et celles appartenant à des mots monosyllabiques sont confondues entre les deux dépendances ($p = 0,999$). Ce sont donc surtout les syllabes finales qui se démarquent entre les deux dépendances, ce qui est congru avec le fait que la durée différencie la position finale des autres positions.

Nous utilisons ensuite `lmer()` afin de créer des modèles mixtes pour savoir si nos résultats sont significativement différents en prenant en compte les variables aléatoires ayant un impact sur les données. Dans un premier temps, nous créons un modèle pour chaque dépendance afin de voir si elles présentent bien le même phénomène constaté dans le reste de nos données.

Les prédicteurs que nous testons sont la durée et la modulation cepstrale (dispersion) sur la position du diphone. Les variables aléatoires sont le diphone concerné et le locuteur, ainsi que d'autres variables dont nous suspectons qu'elles auront un impact : le nombre de mots dans l'énoncé (lg), la position du mot dans le sujet ou l'objet (pos), et le nombre de sujets ou d'objets dans l'énoncé (nb).

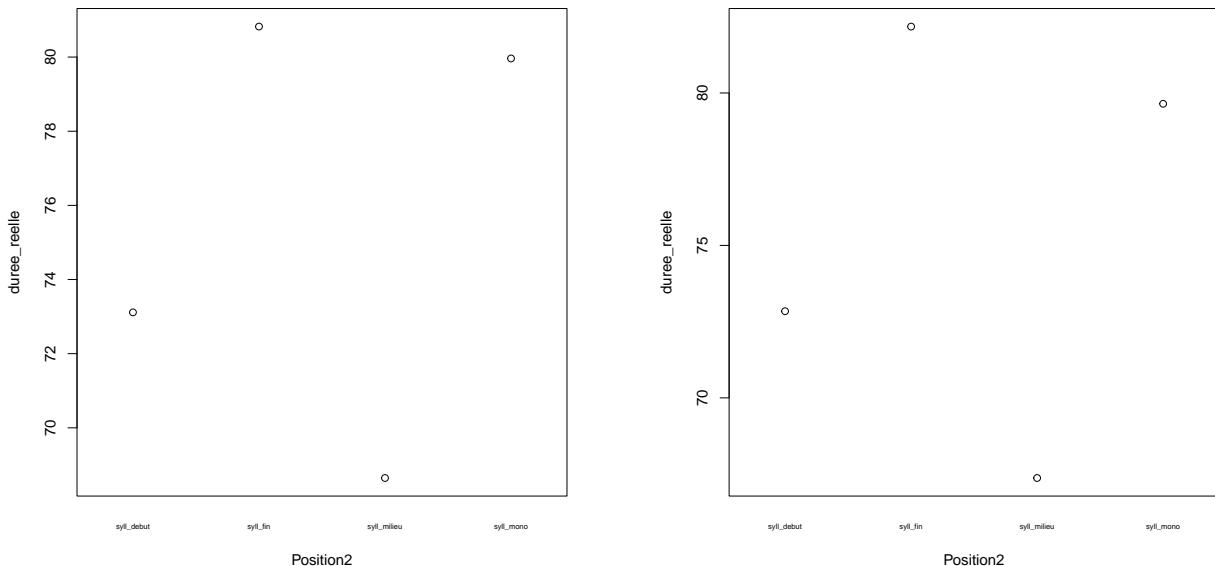
```
lmer(duree ~ position+(1|diphone)+(1|locuteur)+(1|lg)+(1|pos)+(1|nb), data=corpus)
```

(6)

Random effects :			
Groups	Name	Variance	Std.Dev.
diphone	(Intercept)	209.140	14.462
longueur_enonce	(Intercept)	21.639	4.652
locuteur	(Intercept)	3.110	1.764
pos_dependance	(Intercept)	6.063	2.462
nb_dependance	(Intercept)	3.164	1.779
Residual		545.389	23.354

Fixed effects :						
	Estimate	Std.Error	df	t value	Pr(> t)	
(Intercept)	7.284e+01	9.193e-01	1.438e+02	79.24	<2e-16	***
Position2syll_fin	9.335e+00	1.946e-01	2.257e+05	47.96	<2e-16	***
Position2syll_milieu	-5.470e+00	1.774e-01	2.256e+05	-30.84	<2e-16	***
Position2syll_mono	6.802e+00	1.847e-01	2.259e+05	36.82	<2e-16	***

TABLE 11 – Modèle mixte (durée) pour la dépendance *objet*.



(a) `plotLMER.fnc()` pour la durée en fonction de la position (mono = mot monosyllabique) du diphone dans le mot, dépendance *sujet*.

(b) `plotLMER.fnc()` pour la durée en fonction de la position (mono = mot monosyllabique) du diphone dans le mot, dépendance *complément d'objet*.

FIGURE 31 – Durée du diphone selon sa position dans une frontière prosodique.

La dépendance *sujet* a des résultats similaires à ceux des objets pour la Table 11. On constate

que les positions sont correctement différencierées pour les deux dépendances et que leurs résultats sont similaires ; seuls les mots monosyllabiques se confondent avec les syllabes finales, mais nous avons vu que cela était courant lors de nos analyse. En ce qui concerne les effets aléatoires sur le Tableau 11, nous voyons que la variance est particulièrement importante pour le diphone et la longueur de l'énoncé. Une ANOVA comparant l'efficacité des autres variables aléatoires est effectuée, et indique que les autres variables ont également un impact significatif sur le modèle ($p = 0$).

Nous effectuons entre une ANOVA suivie d'un test de Tukey afin de comparer les deux dépendances entre elles pour la mesure de modulation cepstrale. D'après les résultats, il n'y a que deux combinaisons dont les interactions ne montrent pas de différence significative, et elles concernent toutes deux les débuts de mots et les sujet monosyllabiques, qui sont une catégorie à part entière – respectivement, les débuts d'objets ($p = 0,999$) et les débuts de sujets ($p = 0,024$).

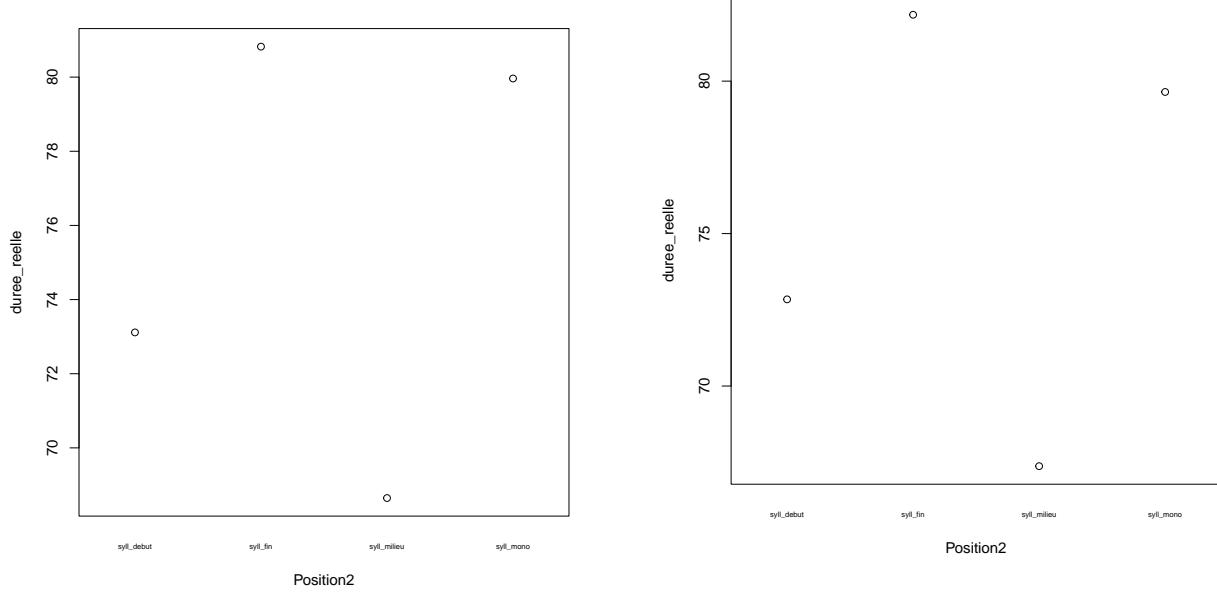
En résumé, les durées de débuts de mots sont similaires entre les deux dépendances, mais les fins de mots sont plus longues pour les objets, ce qui le rapproche des verbes. La modulation cepstrale est plus importante pour toutes les positions de la dépendance objet ; es positions de début et de milieu sont proches de celles des verbes, mais les fins de verbes ont une modulation cepstrale réminiscente des sujets.

Ci-dessous, nous présentons les résultats de notre modèle mixte pour la modulation cepstrale.

Random effects :					
Groups	Name	Variance	Std.Dev.		
diphone	(Intercept)	14.19479	3.7676		
longueur_enonce	(Intercept)	0.44388	0.6662		
locuteur	(Intercept)	0.11629	0.3410		
pos_dependance	(Intercept)	0.09378	0.3062		
nb_dependance	(Intercept)	0.05718	0.2391		
Residual		30.77396	5.5474		

Fixed effects :						
	Estimate	Std.Error	df	t value	Pr(> t)	
(Intercept)	9.130e+00	1.825e-01	3.426e+02	50.04	<2e-16	***
Position2syll_fin	1.678e+00	4.624e-02	2.258e+05	36.28	<2e-16	***
Position2syll_milieu	-6.083e-01	4.215e-02	2.258e+05	-14.43	<2e-16	***
Position2syll_mono	1.121e+00	4.388e-02	2.260e+05	25.56	<2e-16	***

TABLE 12 – Modèle mixte (modulation cepstrale) pour la dépendance *objet*.



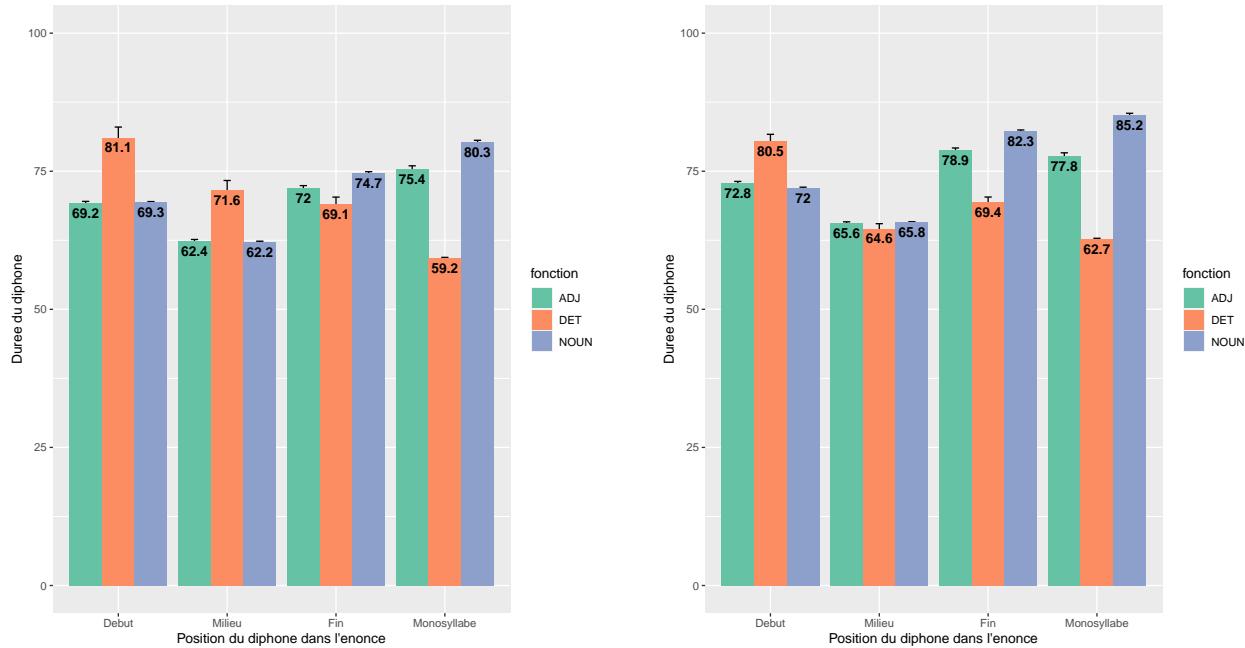
(a) `plotLMER.fnc()` pour la modulation cepstrale en fonction de la position (mono = mot monosyllabique) du diphone dans le mot, dépendance *sujet*.

(b) `plotLMER.fnc()` pour la modulation cepstrale en fonction de la position (mono = mot monosyllabique) du diphone dans le mot, dépendance *complément d'objet*.

FIGURE 32 – Modulation cepstrale du diphone selon sa position dans une frontière prosodique.

La dépendance *sujet* a de nouveau des résultats similaires à ceux des objets pour la Table 12. Les résultats ressemblent énormément à ceux obtenus pour la durée. Toutefois, on voit dans le Tableau 12 que la variance des effets aléatoires est bien moins importante : seuls les diphones ont une variance et un écart-type supérieurs à 0.

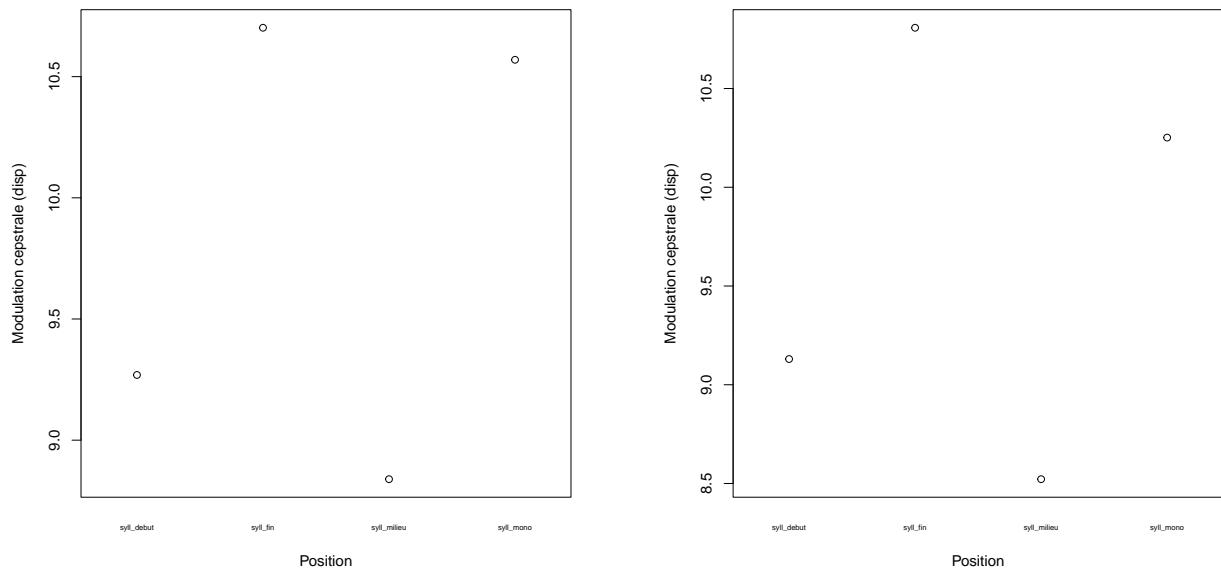
Nous effectuons maintenant une comparaison intra-dépendance, en étudiant quelques parties du discours de chaque dépendance. En regardant les parties du discours les plus fréquentes dans les dépendances, nous choisissons les catégories grammaticales suivantes : nom, adjetif et déterminant.



(a) Durée du diphone appartenant à un nom, un adjectif ou un déterminant, dans les dépendances syntaxiques *sujet* du corpus Ester : au début du mot, au milieu, à la fin ou dans un mot monosyllabique.

(b) Durée du diphone appartenant à un nom, un adjectif ou un déterminant, dans les dépendances syntaxiques *objet* du corpus Ester : au début du mot, au milieu, à la fin ou dans un mot monosyllabique.

FIGURE 33 – Durée du diphone selon sa position dans une frontière prosodique pour trois catégories grammaticales différentes.



(a) Modulation cepstrale du diphone appartenant à un nom, un adjetif ou un déterminant, dans les dépendances syntaxiques *sujet* du corpus Ester : au début du mot, au milieu, à la fin ou dans un mot monosyllabique.

(b) Modulation cepstrale du diphone appartenant à un nom, un adjetif ou un déterminant, dans les dépendances syntaxiques *objet* du corpus Ester : au début du mot, au milieu, à la fin ou dans un mot monosyllabique.

FIGURE 34 – Modulation cepstrale du diphone selon sa position dans une frontière prosodique pour trois catégories grammaticales différentes.

Concernant la durée, on voit Figure 33 que le déterminant est plus long que les autres parties du discours en début de mot : cela est congru avec les observations que nous avions déjà effectuées entre mots grammatical et lexical. Cependant, pour la dépendance *sujet*, le déterminant est également plus long en milieu de mot que les autres catégories grammaticales. C'est là la seule différence entre les deux dépendances.

Pour chaque dépendance syntaxique, nous effectuons une ANOVA suivie d'un test de Tukey afin d'observer les interactions entre chaque partie du discours, pour la durée (représentée ci-après) et la modulation cepstrale : `aov(formula = duree ~ Position * cat_gramm, data = corpus) → TukeyHSD()`. Pour la durée, quelques interactions ne sont pas significativement différentes entre les parties du discours. Certaines de ces interactions sont retrouvées dans les deux dépendances : le début des adjetifs et des noms ($p > 0,5$), leur milieu ($p = 0.999$), et la fin des déterminants avec le début des noms et des adjetifs ($p > 0,5$). D'autres ne sont retrouvées que pour une seule dépendance : parmi les résultats les plus intéressants (on exclut par exemple les fins de mots avec

les mots monosyllabiques, qui sont souvent très similaires et donc significativement non différents), on trouve pour les compléments d'objet une similarité entre le milieu des déterminants, et celui des adjectifs ($p = 0.982$) et des noms ($p = 0.949$) (cela est effectivement visible sur la Figure 33b). Pour les sujets, on trouve au contraire que ces interactions sont significativement différentes ($p > 0,001$) ; c'est la milieu et la fin des déterminants qui ne sont pas significativement différents ($p = 0.840$).

Pour la modulation cepstrale, nous observons des différences non significatives relatives aux deux dépendances : elles concernent les milieux d'adjectifs et de noms ($p > 0,814$). Pour la dépendance *objet* uniquement, il n'y a qu'une autre interaction non significative : les débuts de mots pour toutes les parties du discours ($p > 0,349$), encore une fois visible sur la figure. Pour la dépendance *sujet* uniquement, on retrouve plus de positions non significatives entre elles : elles concernent toutes des débuts et fins de mots $p > 0,999$. Cela est congru à nos observations, en effet la modulation cepstrale différence les positions postpausale et intermédiaire (ou début et milieu de mot), mais pas forcément ces premières avec la position prépausale (ou fin de mot).

5.4.2 Dépendance syntaxique des déterminants

Finalement, nous étudions les deux parties du discours comportant la dépendance syntaxique *déterminant (det)* : il s'agit de la catégorie DET (déterminant) et la catégorie NUM (déterminant numéral). Ces deux catégories sont distinctes dans l'annotation syntaxique et permettent de comparer deux classes de déterminants. La dépendance déterminant réfère à tous les mots produits en position de déterminant dépendant d'un autre mot, dont font partie les déterminants numéraux comme « deux » ou « trois ». Ces derniers ne sont pas utilisés qu'en dépendance déterminant, ce qui les met ainsi à part des autres déterminants DET : un exemple de numéral en dépendance déterminant est « *Cinq petit cochons* » alors qu'un numéral en dépendance modifieur est « *Les deux minutes* » et un numéral en dépendance complément d'objet est « *Ça fait trois.* ». Les pronoms numéraux sont des cas connus de déterminants numéraux sans nom : « *Trois* ont repeint la façade ».

En plus de propriétés syntaxiques distinctes, les déterminants et déterminants numéraux ont des propriétés différences en termes sémantiques : par exemple, les déterminants numéraux sont toujours distributifs dans un énoncé, alors que ce n'est pas toujours le cas des autres déterminants (« Il y a *une* femme que tout homme aime, à savoir sa mère ») (Corblin and De Swart, 2004). Ils comportent également la notion de cardinalité qui permet de compter les noms déterminés par les numéraux (Beysson, 2017). Le comportement sémantique différent des déterminants numéraux leur

confère plus de qualités informationnelles. Or, de nombreuses études montrent que l'informativité influence la production des mots : ils sont mieux articulés si plus informatifs (Goodman et al., 1990), et accentués différemment dans le cas de T-Marking, c'est-à-dire lorsqu'une réponse à une question est plus précise que la question elle-même (par exemple, dans « Que fumaient les chanteurs de rock ? – Les chanteurs de rock anglais fumaient des cigarettes », la réponse a un accent prosodique sur « anglais » car ce mot apporte une information nouvelle qui n'était pas contenue dans la question (Marandin et al., 2002)). Il est donc intéressant de voir si, ici, les déterminants numéraux sont produits différemment des autres déterminants. Pour étudier cela, nous filtrons dans nos deux corpus les déterminants et numéraux ayant la dépendance *déterminant*, et nous faisons un graphique représentant les moyennes de durée des deux parties du discours dans différentes positions de l'énoncé.

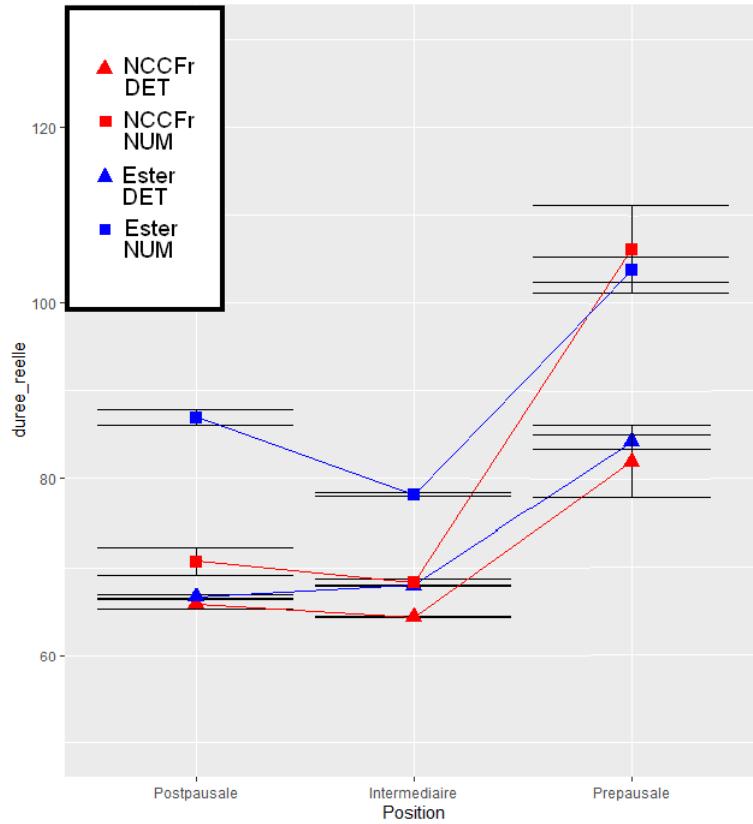


FIGURE 35 – Durée du diphone selon sa catégorie grammaticale (déterminant ou numéral) et sa position dans une séquence : après une pause, dans la séquence, et avant une pause.

Sur la Figure 35, on voit que la position du diphone en fin d'énoncé provoque une durée plus haute pour les numéraux, et ce pour les deux corpus. Les diphones prépausals pour les déterminants sont plus courts, même s'ils sont tout de même plus longs que pour les autres positions des

déterminants. Pour les numéraux de Ester, on voit que les position postpausale et intermédiaire sont bien différenciées. Une remarque intéressante est qu'Ester a des durées plus longues que NCCFr, surtout pour les numéraux, cependant en position prépausale, les deux corpus ont tout de même les mêmes mesures.

Une ANOVA suivie d'un test de Tukey révèlent que, comme on peut le voir sur la Figure 35, les positions postpausale et intermédiaire ne se différencient pas ($p > 0,5$) : cela est attendu, car nous avons vu que la durée ne différenciait pas ces deux positions. Une analyse sur la modulation cepstrale montre que cette fois, ce sont les positions postpausale et prépausale qui ne se différencient pas – cela est aussi attendu selon nos observations précédentes. Le seul résultat remarquable est que les numéraux ont une modulation cepstrale très élevée comme leur durée.

Il y a donc une différence marquée entre la production des deux parties du discours déterminant et numéral : ce dernier a une longueur et une modulation cepstrale en fin d'énoncé plus importante que le déterminant, et ce de façon similaire pour les deux corpus. Des exemples de numéraux avec une durée très élevée (entre 120 et 160ms) incluent « un » ou « cinq ».

Résumé

- On observe des différences entre les dépendances *sujet* et *objet*, notamment entre les positions prépausales :

Durée	Modulation cepstrale
Fins de mots plus longues pour les objets ⇒ Ressemblent à celles des verbes	Plus élevée pour les objets ⇒ Début et de milieu de mot proches de ceux des verbes Fins de sujets proches de celles des verbes

- Lorsque l'on observe les parties du discours *déterminant*, *nom* et *adjectif* dans ces dépendances, le déterminant est plus long en début de mot, mais plus court en fin de mot.
- Comme pour nos autres résultats, la durée ne différencie que la position de fin de mot, et c'est la modulation cepstrale qui différencie la position de début de mot du milieu.
- Certaines catégories grammaticales agissent différemment dans la même dépendance, comme les déterminants et les numéraux en dépendance *déterminant*.

6 Discussion

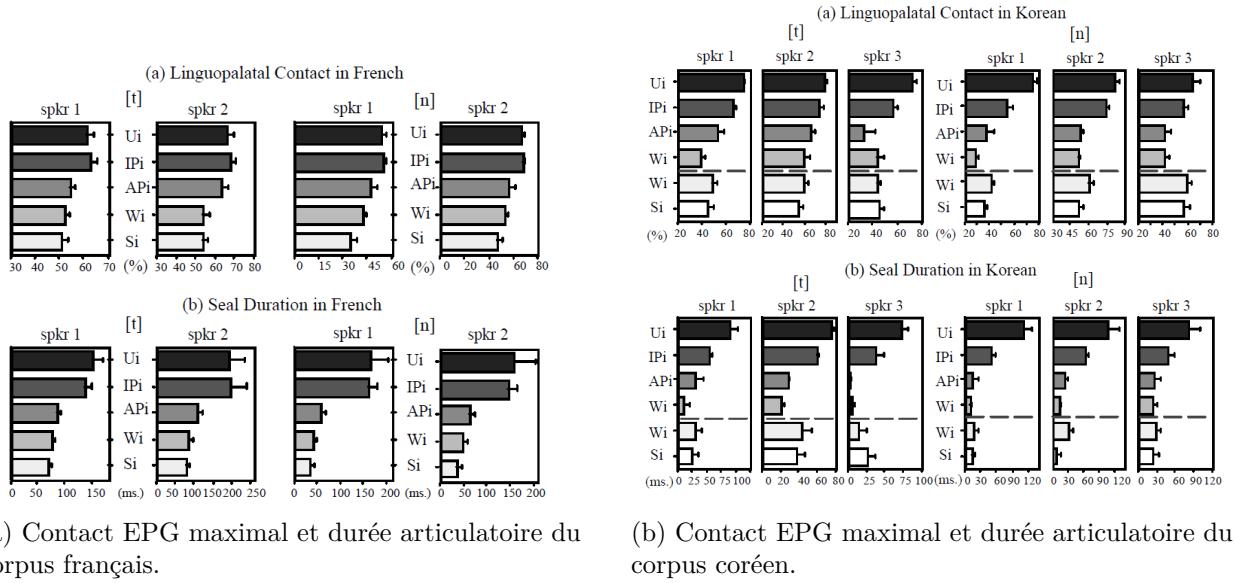
6.1 Mesures de durée et modulation cepstrale

Dans nos Résultats, nous avons maintes fois évoqué les valeurs de modulation cepstrale obtenues. Dans notre Méthode, nous avons fait l'hypothèse – corroborée par notre État de l'art – que cette valeur est liée à l'articulation des phonèmes par les locuteurs. Ainsi, notre analyse de la durée et modulation cepstrale aux frontières prosodiques a montré les points suivants : en position prépausale de fin de séquence, le diphone a une durée plus longue et une modulation cepstrale importante – il y a hyperarticulation du diphone final –, en position postpausale de début de séquence, le diphone n'a pas de durée plus longue mais tout de même une modulation cepstrale importante – il y a renforcement du diphone initial. Ces résultats se retrouvent au sein du mot également.

Les corpus présentent tous deux les résultats retrouvés plus haut, cependant le corpus de parole journalistique Ester possède des durées de diphones en moyenne plus longues. Cela est attendu, car les journalistes parlent plus lentement. Nous obtenons tout de même des résultats significatifs pour NCCFr également. L'impact des hésitations a été observé dans NCCFr : lorsque les diphones sont suivis par une hésitation, ils sont plus longs de 30ms en moyenne ; une hésitation précédant un diphone n'impacte pas celui-ci. Ce résultat n'a pas été retrouvé dans Ester car ce corpus comporte très peu d'hésitations : nous attribuons cela à sa nature journalistique, en effet les présentateurs de radio préparent les points qu'ils vont aborder à l'antenne alors qu'une discussion entre amis est spontanée, autant dans son thème que les interactions entre participants.

Les résultats obtenus – l'hyperarticulation finale et le renforcement initial – sont retrouvés dans les études déjà menées sur des corpus contrôlés comme Keating et al. (2004) ou Tabain (2003) ; Tabain and Perrier (2005) : ils montrent que les phénomènes d'hyperarticulation finale et de renforcement initial observés en phonologie de laboratoire se retrouvent aussi dans la parole spontanée. Par exemple, nous pouvons comparer nos résultats à ceux de Keating et al. (2004) Figure 36a : on voit que la durée entre le niveau de la séquence et celui du mot change beaucoup et est moins importante pour les mots, alors que le contact linguopalatal est similaire entre les deux niveaux ; c'est bien les résultats que nous observons Tableau 2 si nous assimilons le contact linguopalatal à une mesure d'articulation au même titre que notre modulation cepstrale. Comparés aux résultats du coréen, on voit une différence : le contact linguopalatal est moindre dans le niveau du mot, ce qui est effectivement différent de nos résultats (Figure 36b). En plus de corroborer nos résultats, cette comparaison des données permet de montrer l'intérêt de la modulation cepstrale qui ne nécessite

pas d'instrumentation physiologique pour fonctionner.



(a) Contact EPG maximal et durée articulatoire du corpus français.

(b) Contact EPG maximal et durée articulatoire du corpus coréen.

FIGURE 36 – Résultats pour les corpus français et coréen de Keating et al. (2004).

Analysés en termes de phonologie articulatoire, les phénomènes d'hyperarticulation finale et de renforcement initial sont deux processus gestuels distincts, expliqués par l'utilisation de stratégies articulatoires différentes : comme le suggère Fujimura (1990), le renforcement initial implique des gestes articulatoires plus puissants. Cho (2001) illustre, Figure 37, les actions de quatre paramètres pouvant provoquer des changements dans la production des gestes : un paramètre pouvant correspondre à un renforcement est la rigidité – le geste est plus rigide et nécessite ainsi moins de temps d'articulation tout en atteignant sa cible.

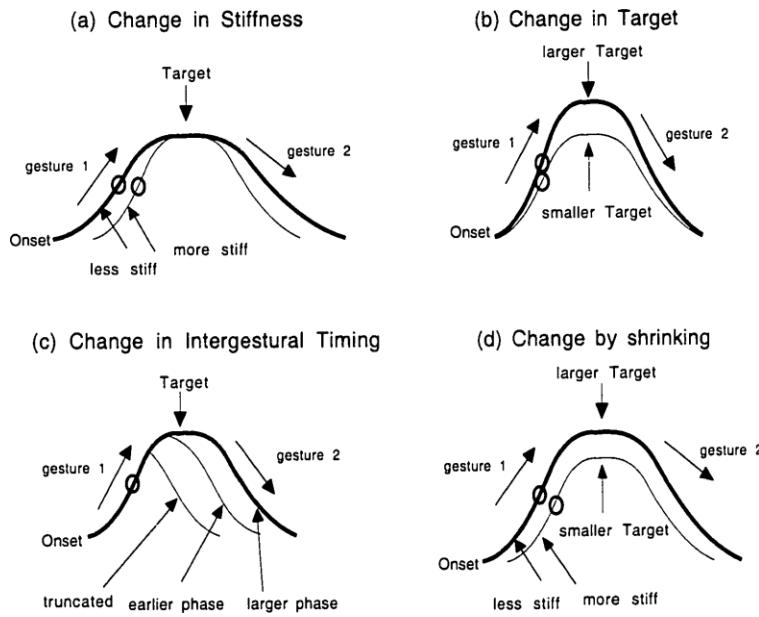


FIGURE 37 – Trajectoires hypothétiques des gestes qui correspondent à un changement de paramètre. (a) montre un changement de rigidité; (b) un changement de cible; (c) un changement de timing intergestuel; et (d) un changement par rétrécissement. Les cercles vides indiquent le point temporel auquel est atteint le pic de vélocité. D'après Cho (2001).

Cela est corroboré par la Figure 38, qui montre qu'une rigidité des gestes plus importante n'affecte pas l'atteinte de la cible articulatoire mais réduit la durée des gestes, comme on constate lors d'un renforcement initial.

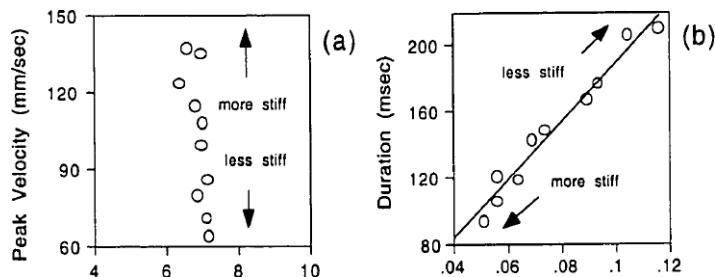


FIGURE 38 – Illustration du pic de vélocité (a) et de la durée des gestes (b) selon la rigidité des gestes, d'après Cho (2001).

D'après les travaux de Goldstein et Lancia – dont les mesures cepstrales cherchaient avant tout à émuler des résultats articulatoires –, et les résultats que nous avons obtenus, nous voyons que la modulation cepstrale est capable de rendre compte de phénomènes liés à la coarticulation des mots et permet de mesurer à quel point deux phonèmes sous forme de diphones sont articulatoirement proches l'un de l'autre. Cette facilité d'usage permet de l'étendre à de nombreuses études sans effort : par exemple, on pourrait imaginer l'intégration de cette mesure en complément de mesures

EMA dans le cadre d'études s'intéressant aux mouvements plus ou moins rapides des articulateurs comme Svensson Lundmark (2023). L'étude de Svensson Lundmark constate que les mouvements d'accélération et de ralentissement répétés des articulateurs coïncident coïncidant avec les frontières acoustiques des segments ; la modulation cepstrale étant une mesure acoustique rendant compte de l'articulation des phonèmes, elle pourrait directement mettre en relation les résultats articulatoires et acoustiques.

6.2 Résultats selon les données syntaxiques

6.2.1 Parties du discours

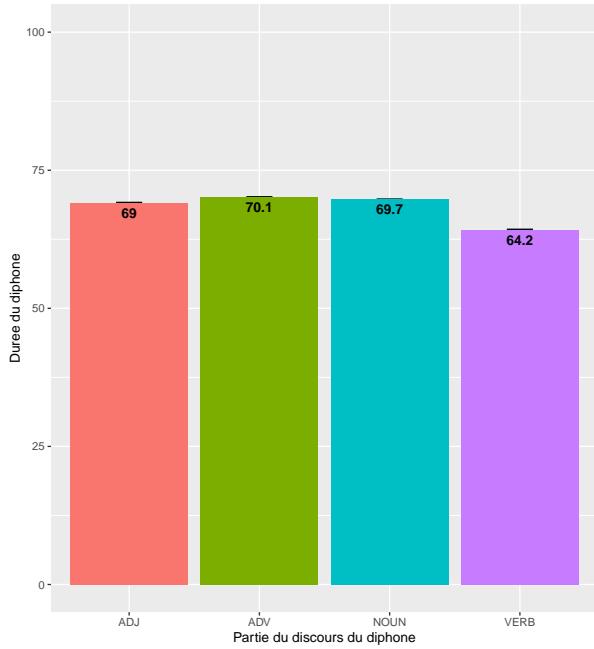
Nous avions fait l'hypothèse que l'hyperarticulation finale et le renforcement initial seraient visibles peu importe la catégorie grammaticale du mot : c'est effectivement le cas pour toutes les parties du discours. Cependant, les mots lexicaux ont des valeurs de durée et de modulation cepstrale plus importantes que les mots grammaticaux. Nous retrouvons aussi des différences entre les deux corpus du fait de leur vocabulaire largement différent : Ester comporte peu de pronoms « tu » par exemple, alors que c'est l'un des pronoms dominants dans NCCFr.

Certaines parties du discours ont des résultats sortant du cadre *hyperticulation finale, renforcement initial* : les noms propres ont une durée moyenne, mais une modulation cepstrale plus forte – nous interprétons ceci comme étant dû au caractère complexe et parfois peu commun des noms propres, comme l'occurrence « CROUS » à 20 unités ou « Tolbiac » (17 unités) – ils demandent une articulation plus ample. Au contraire, les interjections ont une modulation cepstrale moyenne mais une durée plus longue : elles ne demandent pas d'articulation complexe et celle-ci est donc moyenne, mais elles sont allongées comme on s'y attend pour des interjections telles que « euh » où le locuteur réfléchit pendant qu'il parle. Enfin, les auxiliaires sont réduits par rapport aux autres catégories : ceux-ci sont courts et hypoarticulés, étant très couramment utilisés et simples à produire.

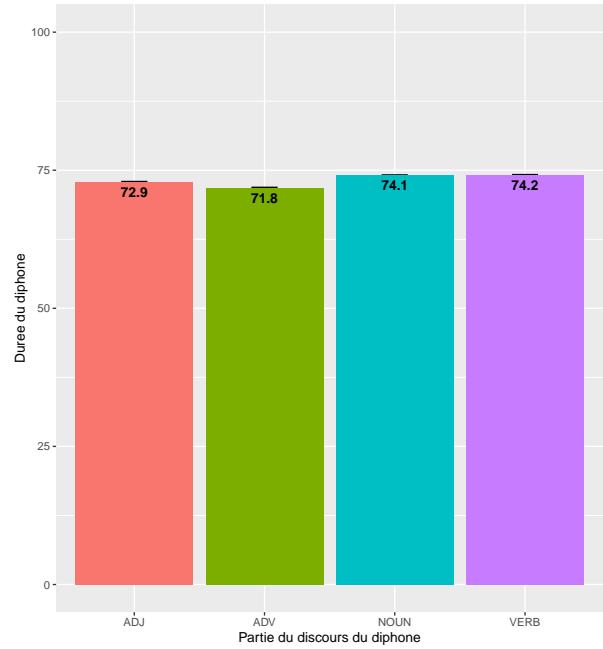
Les résultats obtenus pour les parties du discours sont cohérents avec ceux trouvés dans la littérature : Goodman et al. (1990) avaient en effet trouvé que les mots de classe ouverte (c'est-à-dire les mots lexicaux) étaient mieux articulés que les mots de classe fermée (les mots grammaticaux). Les mots lexicaux sont effectivement moins prévisibles pour l'auditeur que les mots grammaticaux et lui donnent plus d'informations, ils sont donc mieux articulés par les locuteurs.

Nous pouvons comparer les résultats de Gahl et al. (2012) au sujet des parties du discours : ils

ont en effet étudié la longueur du mot selon sa partie du discours en prenant en compte les adjectifs, les adverbes, les noms et les verbes visibles sur la Figure 4. Nous filtrons notre corpus NCCFr pour obtenir les durées des mêmes parties du discours (Figure 39a), d'abord toute position confondue puis dans les différentes positions d'une séquence. Comme pour Gahl et al., la durée des verbes est moins longue que les autres parties du discours. Cependant, nous n'observons pas de durée moins importante pour la catégorie adverbiale. Une ANOVA révèle que seule l'interaction entre l'adverbe et le nom n'est pas significative ($p = 0,009$). Nous avons comparé les résultats de l'étude avec ceux de notre corpus spontané car Gahl et al. avaient également utilisé un corpus spontané : la différence de résultats peut être due à de nombreux facteurs, que ce soit la langue ou encore l'acquisition et l'annotation du corpus, dont nous reparlerons plus tard. Une analyse avec Ester n'apporte pas de résultats significativement différents entre les parties du discours (Figure 39b).



(a) Durée des diphones selon leur appartenance à des parties du discours, corpus NCCFr.



(b) Durée des diphones selon leur appartenance à des parties du discours, corpus Ester.

FIGURE 39 – Durée des diphones selon leur appartenance aux parties du discours adjectif, adverbe, nom et verbe.

6.2.2 Dépendances syntaxiques

Nous avons effectué une courte analyse de l'impact des dépendances syntaxiques sur les données. Cette étude devra être élargie et complétée, en voici les premiers résultats : nous avons comparé deux dépendances syntaxiques rattachées directement à la racine – *sujet* et *complément d'objet* –

en termes de durée et de modulation cepstrale dans le niveau syntaxique du mot. Ces dernières ont des tendances similaires pour les deux dépendances, mais la dépendance des compléments d'objets possède des valeurs plus élevées pour la durée et la modulation cepstrale : on peut interpréter ce résultat comme étant dû au fait que le complément d'objet est souvent situé en fin de groupe accentuel alors que c'est moins probable pour les sujets.

En étudiant les parties du discours *déterminant*, *adjectif* et *nom* à l'intérieur de ces dépendances, nous voyons que le corpus NCCFr a des durées plus longues et une modulation cepstrale plus importante pour les positions qui ne sont pas finales, notamment pour les déterminants. On peut en conclure que les déterminants sont renforcés en position sujet, de par la propension des sujets à être en début d'énoncé. Ce renforcement est surtout présent en début de mot, ainsi qu'au milieu de celui-ci.

Enfin, nous avons comparé les deux parties du discours utilisées dans la dépendance *déterminant* : les déterminants et les déterminants numéraux. Notre but était de voir s'ils sont produits de façon similaire car ils sont dans la même dépendance et sont deux classes de déterminants. Outre les durées plus longues dans le corpus Ester, nous avons vu un phénomène intéressant : la durée des numéraux en fin de séquence se stabilise à 105ms pour les deux corpus, et celle des déterminants est de 85ms pour les deux corpus également. On observe donc qu'un numéral est plus affecté par l'allongement final qu'un déterminant. Cela va dans le sens de notre hypothèse indiquant que le taux d'informativité est plus important pour les déterminants numéraux, et que celui-ci se traduit par une meilleure articulation et une durée plus longue des diphones lui appartenant (Goodman et al., 1990). Toutefois, cela peut être aussi dû à la fréquence très élevée des déterminants (plus nombreux que les numéraux), qui les rend moins rares et donc moins sujets à l'allongement.

Une continuation de cette étude sur les dépendances syntaxiques pourrait par exemple s'intéresser à la différence entre compléments d'objet et modificateurs, ou les relations entre un mot et celui qui lui est syntaxiquement dépendant. Elle pourrait aussi analyser la différence de comportement d'une même partie du discours dans plusieurs dépendances syntaxiques différentes.

6.3 Phonèmes et types de syllabe

Jusqu'ici, nous avons analysé le comportement de diphones dans différentes conditions prosodiques et syntaxiques. Cependant, nous avons également étudié les phonèmes produits dans différents types de syllabes – seulement en termes de durée car notre mesure de la modulation cepstrale

s'effectue sur un diphone entier et pas des phonèmes. Nos observations montrent que le type de syllabe (ouverte de forme (C)V ou fermée de forme V(C)C) influence la façon dont l'allongement final se propage dans la syllabe : les deux types de syllabes subissent un allongement de la même longueur, mais les syllabes ouvertes ne voient que leurs voyelles allongées alors que le phénomène est plus divers pour les syllabes fermées. En effet, lorsque la syllabe est fermée par une occlusive sourde, une fricative ou une consonne nasale, c'est la consonne qui porte l'allongement et la voyelle est moins allongée qu'eux. Au contraire, lorsque la syllabe est fermée par un autre type de consonne, l'allongement se répartit de façon égale dans toute la rime de la syllabe.

Un autre phénomène intéressant concerne la longueur des syllabes selon la position dans l'énoncé et le mot : il y a un croisement entre les durées des deux types de syllabes – les syllabes fermées sont plus longues en position initiale, et baissent énormément en position intermédiaire de sorte à ce que les syllabes ouvertes soient alors plus longues. Cet écart se creuse davantage en position finale où la syllabe ouverte est bien plus longue. Il semble donc que le type de syllabe ait aussi un impact sur le début de la séquence : les syllabes fermées subissent un allongement qui n'est pas présent pour les syllabes ouvertes.

Ce résultat a été trouvé au niveau prosodique de la séquence, et s'applique également au niveau du mot dans le corpus Ester, mais pas NCCFr (voir Annexe, Figure 44).

Finalement, une analyse a été effectuée sur le nombre de syllabes d'un mot, et a révélé que les mots de plus de sept syllabes ont des diphones plus courts que des mots moins longs : il y a un effet de compensation dans la parole poussant à une régularisation de la durée des mots par l'allongement ou la réduction de la longueur de ses diphones.

La longueur plus importante des consonnes en position postpausale/initiale plutôt qu'en position prépausale s'inscrit dans un schéma d'allongement initial de la consonne couramment décrit (White et al., 2020) : alors que la voyelle est traditionnellement décrite comme allongée en fin de mot ou d'énoncé, la consonne est allongée en début de niveau prosodique. White et al. étudient l'universalité de ces critères sur trois populations : des locuteurs anglais, italiens et hongrois. Leurs résultats indiquent que l'allongement initial des consonnes est plus exploité par les auditeurs que l'allongement final de la voyelle ; seuls les locuteurs anglais l'utilisent pour segmenter les énoncés. Nous expliquons ces résultats par le fait que les autres langues peuvent utiliser d'autres indices en frontière finale de niveau prosodique pour aider à la segmentation : il faudrait observer l'impact de l'articulation des voyelles en position finale sur la segmentation.

Nos résultats sont également en accord avec Klatt (1975) : les voyelles sont plus longues en fin de frontière prosodique que pour les autres positions. De plus, dans des configurations où la voyelle est suivie d'une consonne en position finale, celle-ci peut affecter sa durée : les résultats de Klatt montrent que les occlusives sourdes raccourcissent la voyelle alors que des fricatives, nasales ou occlusives voisées l'allongent au contraire. Cela n'est cependant pas semblable à nos résultats : les occlusives sourdes ainsi que les nasales et les fricatives tendent à allonger la voyelle contrairement aux occlusives voisées ou aux liquides.

Le fait que les occlusives voisées ne subissent pas d'allongement final contrairement aux occlusives sourdes, aux nasales ou aux fricatives est de notre point de vue expliqué articulatoirement : les occlusives voisées sont moins faciles à tenir que leur contrepartie sourde, les nasales ou les fricatives, car le voisement doit être maintenu tout en fermant totalement la cavité buccale. Afin de faciliter leur production, elles sont donc moins longues que les autres.

6.4 Autres facteurs de variation

6.4.1 Fréquence fondamentale

L'impact de la fréquence fondamentale f_0 et du locuteur a été étudié sur les corpus. La f_0 est mesurée sur chaque phonème voisé puis, si le diphone entier est voisé, on calcule la moyenne des deux valeurs obtenues. On recueille les valeurs de f_0 selon la position des diphones dans la séquence et le mot, en retirant comme dans la partie Résultats les hésitations des corpus. Cependant, la f_0 ne présente pas de variations liées à des positions prosodiques. Nous obtenons deux constats à partir des résultats : la f_0 est toujours plus haute dans le corpus NCCFr que dans Ester. De plus, la f_0 est plus haute de 20Hz dans la dépendance *sujet* que la dépendant *complément d'objet* : nous pensons que cela est lié au fait que les sujets sont souvent situés en début d'énoncé comparés aux compléments d'objet, et qu'ils subissent donc moins la ligne de déclinaison de f_0 que ces derniers.

Des études sur la f_0 à certaines frontières prosodiques ont montré que celle-ci peut avoir un impact sur les données : Tseng and Su (2008) étudient le niveau prosodique du paragraphe, c'est-à-dire un niveau constitué de plusieurs énoncés successifs ; leurs résultats montrent que la f_0 en position initiale de paragraphe est plus haute que celle en milieu de paragraphe, elle-même plus haute que la fin de paragraphe. Au niveau de la perception, Mitterer et al. (2016) observent qu'en anglais la f_0 ne contribue pas à la perception des frontières prosodiques par les auditeurs, contrairement à la durée : ce résultat a des conséquences intéressantes pour notre analyse – si la f_0 n'est pas vecteur de

distinction de frontières prosodiques pour les auditeurs, alors dans une perceptive où les locuteurs s'adaptent aux besoins des auditeurs, les locuteurs pourraient se concentrer sur des indicateurs de frontière prosodique plus saillants comme la durée ou l'articulation des phonèmes et ne pas produire de différences de f_0 importantes aux frontières.

6.4.2 Mots monosyllabiques ou entre deux pauses

Nous avons pris en compte les mots monosyllabiques dans nos analyses, ainsi que ceux étant entre deux pauses (c'est-à-dire postpausals et prépausals en même temps). En effet, ceux-ci ne pouvaient pas être catégorisés autrement dans nos données car appartenant à plusieurs positions en même temps – mais sont-ils influencés par une position en particulier ?

En observant les données globales de durée et de modulation cepstrale opposant mots monosyllabiques et entre deux pauses (Tableau 2), nous constatons une différence fondamentale entre les deux : les diphones de mots entre deux pauses s'alignent sur la durée des diphones prépausals, alors que ceux d'un mot monosyllabique s'alignent sur celle des diphones de début ou de milieu de mot – autrement dit, au niveau de la séquence, un mot entre deux pauses s'aligne sur une durée en finale de séquence alors qu'au niveau du mot, le mot monosyllabique s'aligne sur une durée de début ou milieu de mot. Concernant la modulation cepstrale, celle-ci a des résultats plus ambigus et est souvent située entre celle d'un diphone postpausal/de début de mot et celle d'un diphone prépausal/en fin de mot.

Nous avons filtré ces diphones selon leur appartenance à un mot grammatical ou lexical (Tableaux 6 et 7) : un résultat notable est qu'un diphone dans un mot monosyllabique lexical a une durée proche de celle d'une fin de mot, alors que s'il est dans un mot grammatical, la durée est bien plus courte et ressemble à un début ou milieu de mot. Nous pouvons interpréter cette durée plus courte par la fréquence très élevée de mots monosyllabiques grammaticaux, attendu car nombre de mots grammaticaux sont monosyllabiques.

Enfin, le type de syllabe constituant les diphones étudiés a également été pris en compte (Tableaux 9 et 10) : pour une syllabe ouverte, le diphone d'un mot entre deux pauses émule un allongement de la voyelle retrouvé en fin de séquence alors que celui d'un mot monosyllabique est plus proche des mesures retrouvées en début de mot. Pour une syllabe fermée, le constat est le même.

Nous pouvons en conclure que le comportement d'un diphone d'un mot entre deux pauses est le même que celui d'un mot prépausal : on y retrouve l'hyperarticulation déjà observée. Mais le diphone d'un mot monosyllabique se comporte comme celui d'un début de mot, à part pour un

mot lexical – le caractère souvent grammatical du mot monosyllabique joue donc un rôle dans ces résultats. Nous expliquons le comportement des diphones dans un mot entre deux pauses par l'influence prépausale de sa position : en effet, un mot entre deux pauses doit être soit considéré et produit comme un postpausal, soit comme un prépausal. Ici, c'est la position prépausale qui impose son influence : selon le modèle du π -gesture (Byrd and Saltzman, 2003), la frontière prosodique finale de séquence a une activation du π -gesture plus forte que la position initiale de séquence, et cause ainsi un allongement de la production du mot avec des gestes articulatoires bien distincts sans chevauchement, menant à un phénomène d'hyperarticulation.

Les mots monosyllabiques ne sont pas souvent étudiés aux côtés de mots multisyllabiques : les expériences contrôlées utilisent soit uniquement des mots monosyllabiques, soit uniquement des mots multisyllabiques. Mo et al. (2010) étudient les mots monosyllabiques selon leur position dans les frontières prosodiques d'un énoncé : ceux-ci subissent les mêmes phénomènes d'hyperarticulation finale et de renforcement initial que les autres mots, le noyau étant le principal élément phonologique subissant ces phénomènes. Une analyse des mots monosyllabiques dans les différentes positions de la séquence pourrait être envisagée dans une continuation de notre étude.

6.4.3 Modulation cepstrale et locuteurs

Une des réserves à propos de la modulation cepstrale est que son fonctionnement par diphone utilisé dans notre étude est plus adapté à une comparaison diphone-par-diphone qu'à des mesures générales sur un corpus entier. Après tout, la modulation cepstrale mesure à quel point les deux parties du diphones sont éloignées articulatoirement parlant, et il n'y aura pas les mesures standard pour [biɪ̯] et [nɔ̯n] car le premier a des diphones plus éloignés articulatoirement que le second. Cependant, nous faisons l'hypothèse que la taille des corpus égalise ces possibles différences entre diphones afin d'obtenir des mesures globales, et les résultats vont dans le sens de cette hypothèse. Toutefois, si le corpus est plus petit les effets de chaque diphone deviendront plus saillants, et nous en avons un exemple. Dans la [Méthode](#), nous avons montré le graphique de la moyenne de la modulation cepstrale de tous les locuteurs de NCCFr (Figure 15b) et nous voyons qu'un locuteur se démarque en particulier : le sixième en partant de la droite. Or, il s'avère que ce locuteur est issu du plus petit fichier de parole du corpus NCCFr – 29_11_07_nb1_1_16 – : il a dix fois moins de diphones qu'un fichier moyen de NCCFr. Ainsi, sa moyenne de modulation cepstrale est calculée sur moins de diphones que les autres fichiers et cela a donné lieu à une mesure plus importante.

Au contraire, le douzième locuteur en partant de la droite (27_11_07_nb2_1_16) de la Figure

15a basée sur la dispersion de la modulation cepstrale est repérée car sa mesure est plus basse que la moyenne : en analysant ses données, nous constatons qu'il possède deux fois plus de diphones produits que les autres fichiers en moyenne, ce qui a dû biaiser la mesure. Il est intéressant de voir que la dispersion et la moyenne de modulation cepstrale ne détectent pas les mêmes fichiers à part : la moyenne est sensible à des corpus trop petits alors que la dispersion est sensible à des corpus plus grands que les autres. En termes de durée, aucun locuteur ne se distingue particulièrement dans nos données. Toutefois, les locuteurs ont régulièrement été utilisés en tant que variable aléatoire au sein des modèles mixtes utilisés au long de l'analyse des données, et sont l'un des facteurs de variation les plus importants après les diphones d'après le modèle.

6.5 Critiques

Certaines critiques peuvent être formulées à l'égard de notre étude, et nous souhaitons les aborder ici. La première concerne la taille des corpus, qui est une force mais aussi une faiblesse. En effet, ceux-ci possèdent des données très variées qui peuvent vite rendre un résultat significatif alors qu'il ne l'est pas vraiment. Nous avons pris en compte ce facteur d'influence en effectuant des tests de Tukey afin d'avoir le détail des interactions testées, mais une analyse de la taille d'effet sur les données pourrait être bienvenue.

La comparaison entre les parties du discours *déterminant* et *numéral* lorsqu'elles sont en dépendance *déterminant* est visée par cette taille d'effet. En effet, nous avons noté que les numéraux étaient significativement plus longs que les déterminants et avaient une modulation cepstrale plus élevée. Cependant, il y a presque dix fois moins de numéraux que de déterminants dans les corpus : cela peut entraîner des valeurs plus élevées dans les résultats pour les numéraux, et être à l'origine de la différence significative que nous avons observée.

Nous avons départagé les différentes mesures de modulation cepstrale au fil de notre analyse : les mesures normalisées et non normalisées n'apportaient pas de résultats différents, et la dispersion de la modulation cepstrale a été conservée car elle fonctionne mieux avec de "vraies" pauses de plus de 100ms. Cependant, ce constat n'a pu être effectué que parce que les corpus (en particulier NCCFr) comportaient des emplacement annotés comme des pauses alors qu'ils faisaient moins de 100ms. Un exemple est donné Figure 40.

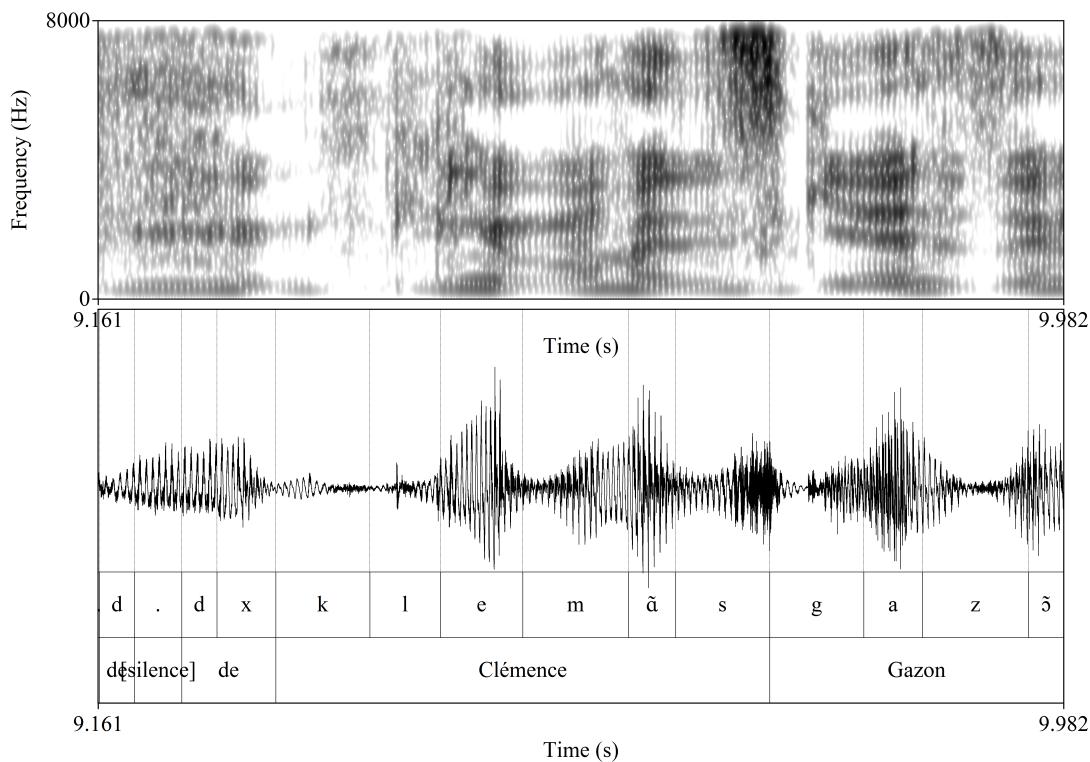


FIGURE 40 – Extrait de 29_11_07_nb1_1_16 montrant une partie très courte du fichier annotée comme une pause.

Nous voyons que sur le deuxième champ de la TextGrid est annoté une pause alors que celle-ci ne dure que 40ms et elle contient de la parole voisée ; la case n'est même pas assez grande pour contenir le texte « [silence] ». C'est ce type de résultats que nous avons filtré par la suite pour nos analyses.

Il y a également un autre problème d'annotation du corpus NCCFr auquel nous avons été confronté : celle des occlusives sourdes. L'annotation de ces consonnes est souvent source de débats, car il n'est pas toujours facile de sélectionner le début de la consonne quand celui-ci n'est pas différenciable de la pause le précédent, par exemple. C'est le cas de la Figure 41 : il est difficile de déterminer quand commence la consonne donc un choix arbitraire a dû être fait – peut-être qu'il empiétait sur la durée de la consonne.

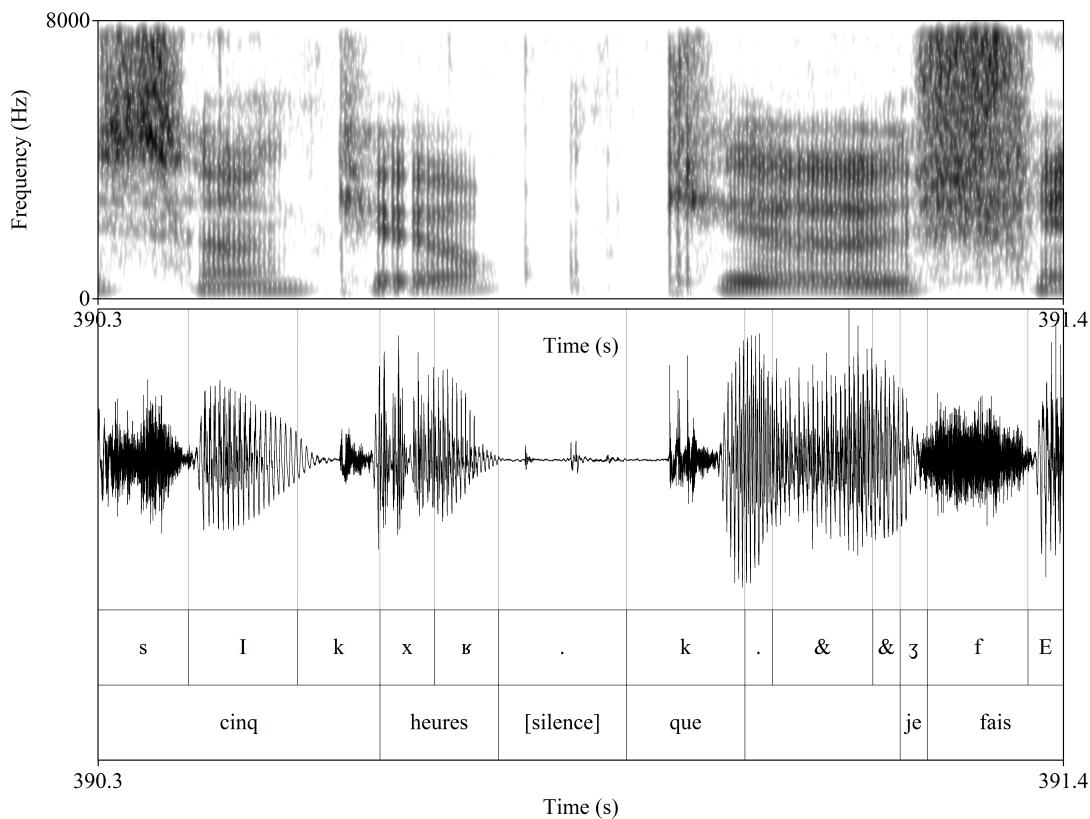


FIGURE 41 – Extrait de 29_11_07_nb1_1_16 montrant une occlusive sourde annotée après une pause.

Une dernière observation concerne l'annotation de la fin des occlusives sourdes lorsque celles-ci présentent une marque d'expiration : il arrive que cette expiration soit considérée comme une pause alors qu'elle est très rapide (30ms) et inaudible, comme sur la Figure 42.

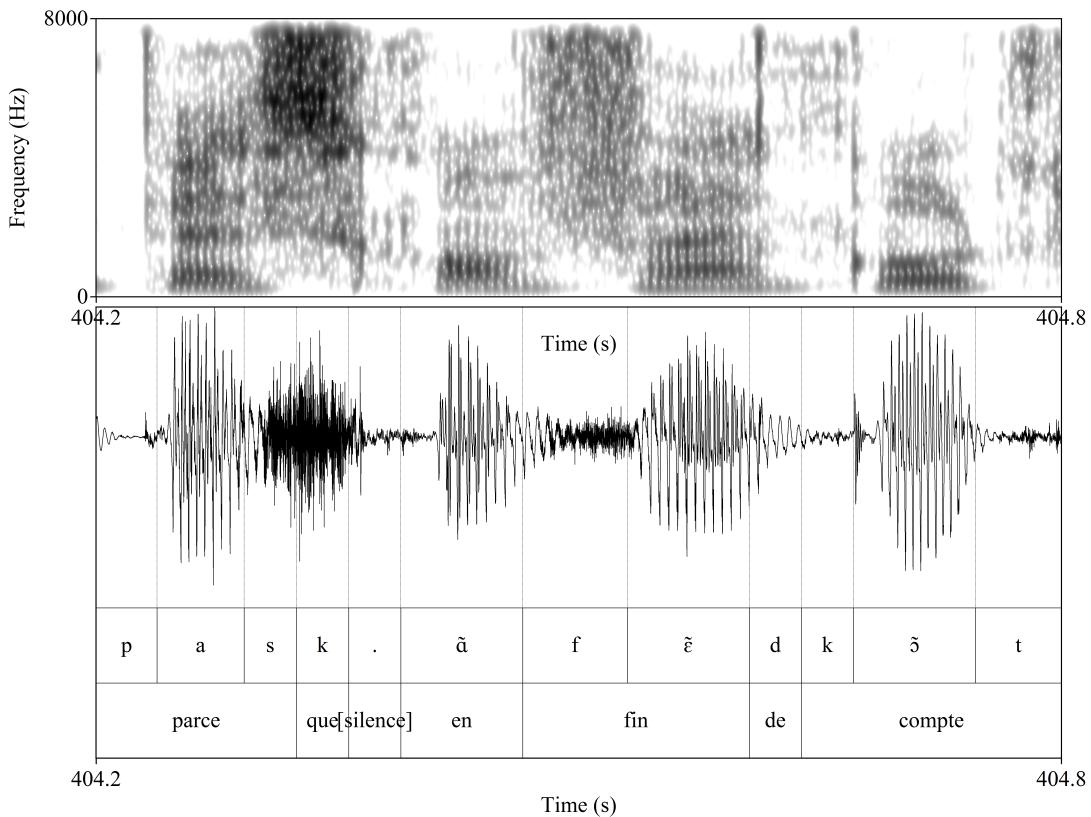


FIGURE 42 – Extrait de 29_11_07_nb1_1_16 montrant une pause annotée sur une expiration d’occlusive sourde.

Tous ces problèmes sont liés à l’aligneur utilisé pour annoter le corpus ; il existe d’autres exemples de problèmes liés à l’alignement comme celui du schwa français, qui permet grâce à sa difficulté d’annotation de comparer des aligneurs entre eux (Bürki et al., 2008).

Il reste encore de nombreuses analyses que nous pourrions mener sur ces corpus : nous avons cité les dépendances syntaxiques plus haut, mais d’autres mesures peuvent être prises en compte comme les bigrammes dont nous avons parlé dans l’État de l’art, que nous n’avons utilisés que superficiellement : nous avons cherché à répliquer un des résultats de Gahl et al. (2012) visible Figure 4 où la durée des mots diminuait à mesure que le bigramme suivant était fréquent. Afin de tester cela, nous avons construit un modèle mixte prenant en prédicteurs la durée du diphone et la fréquence d’apparition du mot suivant (bigramme), en prenant en compte les effets aléatoires des diphones et des locuteurs. Le résultat est visible Figure 43, il correspond au résultat trouvé par Gahl et al. : plus un mot est fréquent, moins il est long. L’utilisation du même modèle mixte pour la dispersion de modulation cepstrale nous apprend que cela est aussi le cas pour l’articulation du

diphone : plus il est fréquent, plus il est coarticulé ($p < 2.2e - 16$).

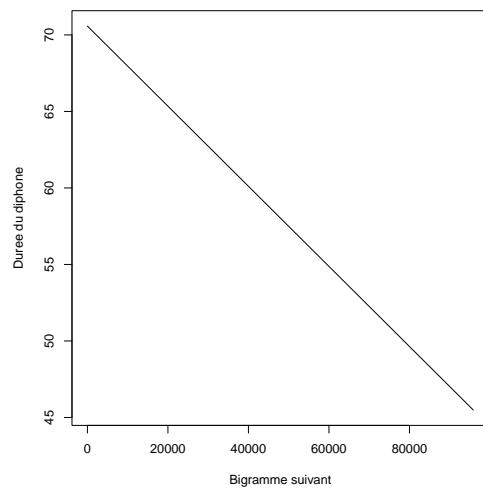


FIGURE 43 – Modèle mixte pour la durée en fonction du bigramme suivant, corpus NCCFr.

7 Conclusion

Notre étude était centrée sur l'analyse de la durée et la modulation cepstrale dans deux corpus, de parole spontanée et journalistique respectivement. Elle s'articulait autour de plusieurs hypothèses :

Notre étude s'articulait autour des deux hypothèses suivantes :

- 1) Les phénomènes prosodiques d'hyperarticulation finale et de renforcement initial sont retrouvés dans les deux corpus, mais le corpus journalistique possède des valeurs plus élevées.
- 2) Les parties du discours sont également régies par ces phénomènes, mots lexicaux comme grammaticaux.
- 3) La forme de la syllabe (ouverte ou fermée) a un impact sur l'agissement de ses phénomènes selon le phone final de la syllabe.
- 4) Les dépendances syntaxiques peuvent également faire varier les résultats.

D'après nos résultats, la première hypothèse est vérifiée : non seulement l'hyperarticulation finale et le renforcement initial sont retrouvés dans les résultats, mais le corpus journalistique Ester possède bien des valeurs plus élevées. Une analyse poussée de quelques diphones nous informe que la durée des pauses entre séquences est un facteur de variation dans la modulation cepstrale : la dispersion de la modulation cepstrale est plus adaptée à des pauses longues de plus de 100ms.

La deuxième hypothèse est également vérifiée car les mots lexicaux comme grammaticaux sont régis par ces phénomènes. Cependant, leur impact est moins important sur les mots grammaticaux, et certaines parties du discours sont un facteur de variation dans les données – par exemple, le nom propre possède une modulation cepstrale plus élevée que les autres catégories grammaticales sans avoir de variation particulièrement importante de durée.

La troisième hypothèse est confirmée : alors que la syllabe ouverte porte l'allongement final sur sa voyelle, la syllabe fermée possède des règles plus précises – selon la consonne finale, l'allongement est porté soit sur la consonne, soit sur la rime entière.

Enfin, concernant la quatrième hypothèse, celle-ci confirmée pour les dépendances syntaxiques *sujet* et *complément d'objet* : les compléments d'objet ont une hyperarticulation finale et un renforcement initial plus importants que les sujets. Nous expliquons ceci par une informativité plus importante de la part du complément d'objet, car il est situé plus loin dans un énoncé que le sujet. Les deux types de déterminants présents dans la dépendance *déterminant* ont également des valeurs différentes : les déterminants numéraux ont des valeurs plus élevées que les autres ; cela peut aussi s'expliquer par une plus grande informativité de la part des déterminants numéraux.

Ce mémoire a observé de nombreux paramètres de variation, et certains, comme les dépendances syntaxiques, méritent encore d'être étudiés : nous n'avons analysé que trois dépendances syntaxiques, et une hypothèse telle que « les dépendances *compléments d'objet* et les dépendances *modifieurs* d'un même élément réagissent différemment aux phénomènes d'hyperarticulation/renforcement car elles ne se comportent pas pareil dans l'énoncé (les premières ne peuvent pas être supprimées sans que l'énoncé ne perde son sens) » pourrait être approfondie dans une étude ultérieure.

Nous avons également montré comme Goldstein (2019) et Lancia et al. (2020) que la modulation cepstrale était un outil adapté pour la recherche en phonétique acoustique, ici sur des phénomènes prosodiques ; sa facilité d'utilisation et sa robustesse la rendent attractive pour de futurs travaux dans le domaine de la phonétique.

8 Bibliographie

- Astésano, C., Bard, E. G., and Turk, A. (2003). Structural and rhythmic influences on the occurrence of the initial accent in french. In *Proceedings of the 15th International Congress of Phonetic Sciences*, pages 503–506.
- Audibert, N., Fougeron, C., Gendrot, C., and Adda-Decker, M. (2015). Duration-vs. style-dependent vowel variation : A multiparametric investigation. In *18th International Congress of Phonetic Sciences (ICPhS'15)*, page 5.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. *Speech perception and linguistic experience*, pages 171–204.
- Beysson, C. (2017). Déterminants et quantificateurs généralisés dynamiques. In *TALN 2017-Traitement Automatique des Langues Naturelles*, pages 81–93.
- Browman, C. P. and Goldstein, L. (1992). Articulatory phonology : An overview. *Phonetica*, 49(3-4) :155–180.
- Bürki, A., Gendrot, C., Gravier, G., Linarès, G., and Fougeron, C. (2008). Alignement automatique et analyse phonétique : comparaison de différents systèmes pour l’analyse du schwa. *Traitement Automatique des Langues*, 49(3) :165–197.
- Byrd, D. and Saltzman, E. (2003). The elastic phrase : Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31(2) :149–180.
- Chen, M. Y. (1997). Acoustic correlates of english and french nasalized vowels. *The Journal of the Acoustical Society of America*, 102(4) :2360–2370.
- Cho, T. (2001). *Effects of prosody on articulation in English*. University of California, Los Angeles.
- Cohn, A. C. (1993). Nasalisation in english : phonology or phonetics. *Phonology*, 10(1) :43–81.
- Corblin, F. and De Swart, H. (2004). *Handbook of French semantics*. CSLI publications Stanford.
- Delattre, P. (1963). Comparing the prosodic features of english, german, spanish and french. *IRAL*, 1 :193–210.
- Encrevé, P. (1988). *La liaison avec et sans enchaînement : phonologie tridimensionnelle et usages du français*, volume 15. Seuil.

- Fink, A. and Goldrick, M. (2015). The influence of word retrieval and planning on phonetic variation : Implications for exemplar models. *Linguistics Vanguard*, 1(1) :215–225.
- Fónagy, I. (1980). L’accent français : accent probabilitaire (dynamique d’un changement prosodique). *Studia Phonetica Montréal*, 15 :123–233.
- Fujimura, O. (1990). Methods and goals of speech production research. *Language and Speech*, 33(3) :195–258.
- Gahl, S., Yao, Y., and Johnson, K. (2012). Why reduce ? phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of memory and language*, 66(4) :789–806.
- Galantucci, B., Fowler, C. A., and Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic bulletin & review*, 13(3) :361–377.
- Galliano, S., Geoffrois, E., Gravier, G., Bonastre, J.-F., Mostefa, D., and Choukri, K. (2006). Corpus description of the ester evaluation campaign for the rich transcription of french broadcast news. In *LREC*, pages 139–142. Citeseer.
- Galliano, S., Gravier, G., and Chaubard, L. (2009). The ester 2 evaluation campaign for the rich transcription of french radio broadcasts. In *Tenth Annual Conference of the International Speech Communication Association*.
- Gendrot, C. and Audibert, N. (2019). La distinction/e/vs/ /en français standard est-elle maintenue en finale de mot ? étude sur des corpus de parole journalistique et de parole spontanée. *Langue française*, pages 53–66.
- Gendrot, C., Gerdes, K., and Adda-Decker, M. (2016). Détection automatique d’une hiérarchie prosodique dans un corpus de parole journalistique. *Langue française*, pages 123–149.
- Gnamian, B. E. A. (2018). Syntaxe et semantique : Complementarite, dependance et autonomie de deux notions grammaticales. *Revue malienne de Langues et de Littératures*, pages 80–89.
- Goldstein, L. (2019). The role of temporal modulation in sensorimotor interaction. *Frontiers in Psychology*, 10 :2608.
- Goodman, J. C., Nusbaum, H. C., Lee, L., and Broihier, K. (1990). The effects of syntactic and discourse variables on the segmental intelligibility of speech. In *ICSLP*.

- Guibon, G., Courtin, M., Gerdes, K., and Guillaume, B. (2020). When collaborative treebank curation meets graph grammars. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 5293–5302, Marseille, France. European Language Resources Association.
- Han, W., Chan, C.-F., Choy, C.-S., and Pun, K.-P. (2006). An efficient mfcc extraction method in speech recognition. In *2006 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 4–pp. IEEE.
- Keating, P., Cho, T., Fougeron, C., and Hsu, C.-S. (2004). Domain-initial articulatory strengthening in four languages. *Phonetic interpretation : Papers in laboratory phonology VI*, pages 143–161.
- Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of phonetics*, 3(3) :129–140.
- Lancia, L., Li, J., and Goldstein, L. (2020). Complexity patterns underlying speech production activity. In *ISSP 2020*.
- Lévêque, N., Slis, A., Lancia, L., Bruneteau, G., and Fougeron, C. (2022). Acoustic change over time in spastic and/or flaccid dysarthria in motor neuron diseases. *Journal of Speech, Language, and Hearing Research*, 65(5) :1767–1783.
- Li, Q., Soong, F. K., and Siohan, O. (2000). A high-performance auditory feature for robust speech recognition. In *Interspeech*, pages 51–54.
- Lindblom, B. (1990). Explaining phonetic variation : A sketch of the h&h theory. *Speech production and speech modelling*, pages 403–439.
- Marandin, J.-M., Beyssade, C., Delais-Roussarie, E., and Rialland, A. (2002). Discourse marking in french : C accents and discourse moves. In *Speech Prosody 2002, International Conference*.
- Marengo, S. (2014). Dépendances sémantiques et syntaxiques : quatre cas de figure pour l'adjectif en français. *Travaux de linguistique*, pages 103–120.
- Mitterer, H., Cho, T., and Kim, S. (2016). How does prosody influence speech categorization ? *Journal of Phonetics*, 54 :68–79.
- Mo, Y., Cole, J., and Hasegawa-Johnson, M. (2010). Prosodic effects on temporal structure of monosyllabic cvc words in american english. In *5th International Conference on Speech Prosody : Every Language, Every Style, SP 2010*. International Speech Communications Association.

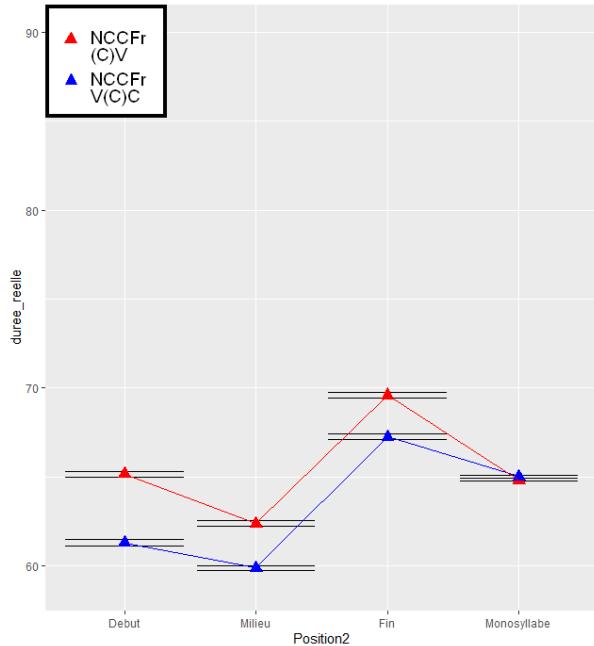
- Muda, L., Begam, M., and Elamvazuthi, I. (2010). Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques. *arXiv preprint arXiv :1003.4083*.
- Munson, B. and Solomon, N. P. (2016). The influence of lexical factors on vowel distinctiveness : Effects of jaw positioning. *The International journal of orofacial myology : official publication of the International Association of Orofacial Myology*, 42 :25.
- Nespor, M. and Vogel, I. (1983). Prosodic structure above the word. *Prosody : Models and measurements*, pages 123–140.
- Rebernik, T., Jacobi, J., Jonkers, R., Noiray, A., and Wieling, M. (2021). A review of data collection practices using electromagnetic articulography. *Laboratory Phonology*, 12(1).
- Scarborough, R. A. (2003). Lexical confusability and degree of coarticulation. In *Annual Meeting of the Berkeley Linguistics Society*, volume 29, pages 367–378.
- Slis, A., Lévêque, N., Fougeron, C., Pernon, M., Assal, F., and Lancia, L. (2021). Analysing spectral changes over time to identify articulatory impairments in dysarthria. *The Journal of the Acoustical Society of America*, 149(2) :758–769.
- Svensson Lundmark, M. (2023). Rapid movements at segment boundaries. *The Journal of the Acoustical Society of America*, 153(3) :1452–1467.
- Tabain, M. (2003). Effects of prosodic boundary on/ac/sequences : articulatory results. *The Journal of the Acoustical Society of America*, 113(5) :2834–2849.
- Tabain, M. and Perrier, P. (2005). Articulation and acoustics of/i/in preboundary position in french. *Journal of Phonetics*, 33(1) :77–100.
- Tiwari, V. (2010). Mfcc and its applications in speaker recognition. *International journal on emerging technologies*, 1(1) :19–22.
- Torreira, F., Adda-Decker, M., and Ernestus, M. (2010). The nijmegen corpus of casual french. *Speech Communication*, 52(3) :201–212.
- Tseng, C.-y. and Su, Z.-y. (2008). Discourse prosody context-global f0 and tempo modulations. In *Ninth Annual Conference of the International Speech Communication Association*.

- Vaissière, J. and Michaud, A. (2006). Prosodic constituents in french : a data-driven approach. *Prosody and syntax*, pages 47–64.
- Welby, P. S. (2003). *The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation*. The Ohio State University.
- White, L., Benavides-Varela, S., and Mády, K. (2020). Are initial-consonant lengthening and final-vowel lengthening both universal word segmentation cues ? *Journal of Phonetics*, 81 :100982.
- Wu, Y. and Adda-Decker, M. (2021). Réduction des segments en français spontané : apports des grands corpus et du traitement automatique de la parole. *Corpus*.
- Zellou, G. (2017). Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics*, 61 :13–29.
- Zellou, G. (2022). Coarticulation in phonology. *Elements in Phonology*.
- Zheng, F., Zhang, G., and Song, Z. (2001). Comparison of different implementations of mfcc. *Journal of Computer science and Technology*, 16 :582–589.

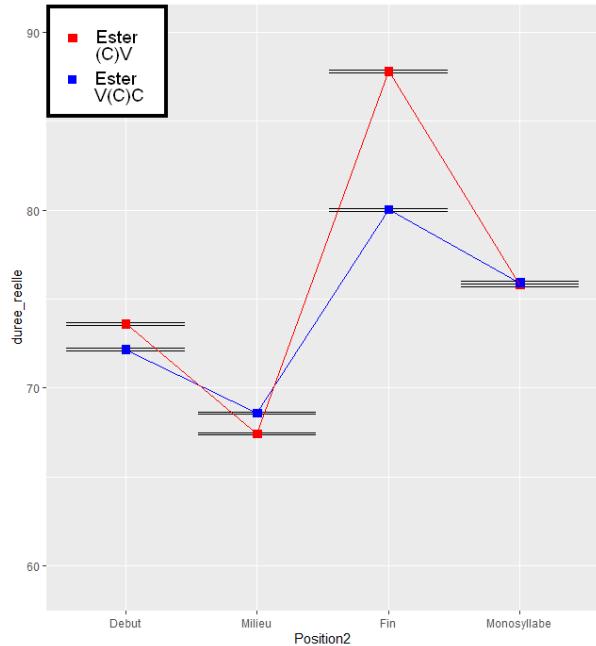
9 Annexe

Modalité	Position	Occurrences NCCFr	NCCFr	Occurrences Ester	Ester
<i>Durée</i>					
Niveau Séquence	Début	72948	88	81181	143
	Inter	392009	73	628674	79
	Fin	54496	104	61309	259
	Mono	7092	112	1256	225
Niveau Mot	Début	4509	60	3157	64
	Inter	139971	86	269679	107
	Fin	0	NA	0	NA
	Mono	414445	77	517869	98
<i>Modulation cepstrale</i>					
Niveau Séquence	Début		11.4		14.4
	Inter		9.3		9.4
	Fin		11.2		11.8
	Mono		10.6		16.1
Niveau Mot	Début		8.2		8.2
	Inter		10		9.9
	Fin		NA		NA
	Mono		9.7		10.3

TABLE 13 – Tableau contenant les durées et modulations cepstrales de dipones appartenant à des mots différents dans NCCFr et Ester, dans différentes structures prosodiques.



(a) Durée du diphone dans une syllabe ouverte ou fermée, selon sa position dans un mot du corpus NCCFr.



(b) Durée du diphone dans une syllabe ouverte ou fermée, selon sa position dans un mot du corpus Ester.

FIGURE 44 – Durée du diphone selon sa position dans un mot.

PoS :position	<i>p-value</i>
Mot_lexical :intermediaire-Mot_grammatical :intermediaire	0.0e+00
Mot_grammatical :postpausale-Mot_grammatical :intermediaire	0.0e+00
Mot_lexical :postpausale-Mot_grammatical :intermediaire	0.0e+00
Mot_grammatical :prepausale-Mot_grammatical :intermediaire	0.0e+00
Mot_lexical :prepausale-Mot_grammatical :intermediaire	0.0e+00
Mot_grammatical :postpausale-Mot_lexical :intermediaire	0.0e+00
Mot_lexical :postpausale-Mot_lexical :intermediaire	0.0e+00
Mot_grammatical :prepausale-Mot_lexical :intermediaire	0.0e+00
Mot_lexical :prepausale-Mot_lexical :intermediaire	0.0e+00
Mot_lexical :postpausale-Mot_grammatical :postpausale	0.0e+00
Mot_grammatical :prepausale-Mot_grammatical :postpausale	0.0e+00
Mot_lexical :prepausale-Mot_grammatical :postpausale	0.0e+00
Mot_grammatical :prepausale-Mot_lexical :postpausale	0.0e+00
Mot_lexical :prepausale-Mot_lexical :postpausale	0.0e+00
Mot_lexical :prepausale-Mot_grammatical :prepausale	5.3e-06

TABLE 14 – Test de Tukey sur le corpus NCCFr en fonction de la durée selon de la partie du discours (PoS) et la position du diphone.

Résumé du mémoire

Ce mémoire s'intéresse à l'hyperarticulation finale et le renforcement initial retrouvés aux frontières prosodiques en français : aucune étude n'a été effectuée sur un corpus de parole spontanée, ainsi nous souhaitons voir si ces phénomènes y sont identifiés.

Nous utilisons la durée et la modulation cepstrale – une mesure encore peu usitée permettant de mesurer le taux de changement articulatoire entre deux segments – sur des diphones afin de voir si ces phénomènes prosodiques sont bien présents. Nous utilisons un corpus de parole spontanée, ainsi qu'un corpus de parole journalistique en tant que témoin. Nous mesurons aussi l'impact de la syllabe, les parties du discours et les dépendances syntaxiques sur nos mesures en frontières prosodiques.

Nos résultats montrent que l'hyperarticulation finale et le renforcement initial sont bien retrouvés dans les deux corpus, et que les parties du discours, les dépendances syntaxiques et la répartition des phonèmes dans la syllabe ont une influence sur ces phénomènes prosodiques.