Actes de JEP-TALN-RECITAL 2024
Toulouse, France, 8 au 12 juillet 2024

Analyse Factorielle de signaux sonores : développement d'une méthode automatique de détermination des frontières optimales entre canaux de fréquence.

Agnieszka Duniec[1] Elisabeth Delais-Roussarie[1] Olivier Crouzet[1]
(1) Laboratoire de Linguistique de Nantes, LLING – UMR6310, Université de Nantes / CNRS
chemin de la Censive du Tertre, 44312 Nantes Cedex, France
agnieszka.duniec@etu.univ-nantes.fr, olivier.crouzet@univ-nantes.fr

RÉSUMÉ

Des études récentes supportent l'hypothèse d'une relation entre les propriétés statistiques des signaux de parole et les mécanismes perceptifs : les gammes de fréquence présentant une corrélation dans leurs modulations d'amplitude pourraient être associées à des frontières spectrales relativement stables envisagées comme optimales sur le plan perceptif. Cependant, des limites afférentes à ces études antérieures ressortent : (1) elles se fondent pour la plupart sur des critères subjectifs à travers l'observation visuelle des courbes de résultats statistiques, et (2) elles n'envisagent pas que les résultats puissent varier en fonction des échantillons de données sélectionnés, de la nature des signaux utilisés, ou de la taille des échantillons. Même si cette position peut être argumentée en lien avec l'approche du *codage efficace*, cet aspect afférent au degré de variation potentiel nécessite d'être évalué. Nous avons mis en place une méthode de détermination automatique des frontières qui permet de répliquer les travaux antérieurs en introduisant une évaluation expérimentale de ces limites et discutons de quelques résultats préliminaires en comparaison avec les études précédentes.

ABSTRACT

Factor Analysis of acoustic signals : development of an automatic method for the determination of optimal frequency boundaries.

Recent studies have led to support the hypothesis of a relationship between statistical properties of spectro-temporal modulations in speech and mechanisms of perceptual analysis : frequency channels exhibiting co-modulated amplitude would be associated with frequency boundaries considered as *optimal* in perceptual terms. However, limits pertaining to some of these studies have to be taken into accout : (1) previous results associated with frequency boundary determination were for some part based on visual inspection of statistical curves, (2) studies have generally assumed that these results would hold for any sample of speech signals, and for any sample size. Even though such assumption may hold in relation to the *efficient coding* hypothesis, the degree of variation relating to the observed results requires to be investigated. A method for the automatic determination of frequency boundaries associated with these approaches has been developped and applied on a speech database. This method provides a way to replicate previous data while experimentally investigating variation. Some preliminary results are discussed and compared with previous studies.

MOTS-CLÉS : perception, statistiques des signaux naturels, hypothèse du codage efficace, implants cochléaires.

KEYWORDS: perception, natural signal statistics, efficient coding hypothesis, cochlear implants.

Actes des 35ème Journées d'Études sur la Parole, pages 252–260.

1 Introduction

L'*hypothèse du codage efficace* pour la perception sonore (Smith & Lewicki, 2006) tire ses origines des travaux sur les *statistiques des signaux naturels* (Simoncelli & Olshausen, 2001; McDermott & Simoncelli, 2011). Du point de vue de cette approche, les signaux acoustiques de communication sont caractérisés par des propriétés statistiques régulières, lesquelles seraient au fondement des mécanismes d'analyse perceptive malgré la diversité apparente des réalisations sonores. Si certaines de ces études (Ming & Holt, 2009; Kluender *et al.*, 2013) se fondent sur une modélisation des signaux à travers des trains d'impulsions (*spike trains*) en lien avec la théorie de l'information et les approches contemporaines du codage neuronal (Smith & Lewicki, 2005), d'autres ont adopté une approche statistique plus traditionnelle fondée sur l'Analyse Factorielle des modulations d'amplitude (Ueda & Nakajima, 2017; Grange & Culling, 2018).

Plusieurs de ces études se sont penchées sur cette hypothèse dans le but d'identifier des propriétés statistiques des signaux de parole dans différentes langues (Ueda & Nakajima, 2017) ainsi qu'en lien avec le codage de la parole dans des implants cochléaires (Grange & Culling, 2018; Ming & Holt, 2009).

1.1 Hypothèse du codage efficace et implant cochléaire

Ming & Holt (2009) ont mesuré les performances de reconnaissance de parole vocodée, souvent décrite comme simulant les informations diffusées par les implants cochléaires, chez des auditeurs normo-entendants. Ils ont montré que, sans changer le nombre de canaux spectraux (6 en l'occurrence) les changements de localisation des frontières spectrales en parole vocodée ont des effets sur les taux de reconnaissance de mots et de segments phonétiques. Les performances sont meilleures si les localisations de ces frontières concordent avec des positions spectrales dérivées de modélisations issues de la théorie de l'information et correspondent donc à une « perspective efficace ». Ces frontières seraient en outre nettement plus basses que celles qui découlent d'une organisation tonotopique, lesquelles sont généralement la référence pour la décomposition spectrale dans les implants cochléaires.

Dans une toute autre perspective, Ueda & Nakajima (2017), ont développé une méthode d'analyse inspirée des travaux de Plomp *et al.* (1967) sur les voyelles : ils étendent cette approche à l'étude d'un corpus de phrases et procèdent, sur la base de signaux acoustiques de parole codés sur environ 20 canaux de représentation spectrale à des Analyses en Composantes Principales (ACP) portant sur les enveloppes d'énergie de ces canaux. Il font varier le nombre de facteurs associés à la sortie de l'ACP (2, 3, 4, 5, 6). Leur travail aboutit à la conclusion que 4 facteurs suffiraient à représenter optimalement des signaux de parole, et ce pour chacune des 8 langues de leur échantillon. Ils constatent par ailleurs que les 3 frontières fréquentielles découlant de chacune des ACP à 4 facteurs réalisées sur ces 8 langues seraient parfaitement appariées (env. 540, 1720, 3300 Hz), ce qui les amène à conclure que les langues seraient de manière générale fondées sur des indices qui seraient parfaitement adaptés à un traitement perceptif « parcimonieux » (ou efficace) de la parole.

Grange & Culling (2018) ont répliqué l'étude de Ueda & Nakajima (2017) en modifiant légèrement l'algorithme d'analyse statistique (accroissement du nombre de canaux spectraux entrés dans l'ACP à environ 100 canaux notamment, estimation de la contribution de chacune des 20 premières Composantes Principales issues de l'ACP à travers les valeurs propres –*eigenvalues*). Ils ont ensuite évalué ces données à la lueur des performances observées en perception de parole vocodée (simulations

d'implants cochléaires) et ont abouti à des conclusions assez similaires aux travaux précédents. Leurs résultats suggèrent néanmoins que, pour rendre compte de manière appropriée des propriétés acoustiques de la parole vocodée, il faudrait 6 à 7 canaux spectraux pour représenter optimalement ces signaux. Cette limite correspond, dans le graphique des valeurs propres (*scree plot*) qu'ils présentent, à un point d'inflexion au-delà duquel les valeurs propres semblent augmenter plus lentement. Cette limite est aussi associée dans les courbes de performance en fonction du nombre de canaux vocodés, à une amélioration moins marquée des performances observées à partir de 8 canaux spectraux. Ces deux mesures (l'une statistique issue de signaux naturels, l'autre comportementale issue de signaux vocodés) seraient donc cohérentes et suggèreraient que cette limite de 7/8 canaux pourrait refléter une version optimale de la représentation perceptive des signaux de parole appropriée à la tâche expérimentale utilisée (reconnaissance de mots dans des phrases simples présentées dans le silence).

1.2 Limites des études antérieures

Toutes les études évoquées ont dérivé, de l'analyse acoustique et statistique d'un corpus de parole, des estimations de localisation de frontières entre canaux de fréquence qui sont considérées comme optimales : elles différencieraient des canaux de fréquence maximalement comodulés entre eux et ces frontières sépareraient les canaux qui sont les moins corrélés et qui, par conséquent, seraient maximalement informatifs. Il ressort de ces approches qu'aucun des travaux précédents n'a évalué le degré de variation des estimations réalisées en fonction de la composition du corpus, de sa taille (environ 1h de parole dans les études précécentes) ni de la durée acoustique des items concaténés pour la constitution du corpus. Or il semble essentiel, avant d'envisager un caractère stable de ces *frontières optimales*, de pouvoir évaluer leur variabilité potentielle.

Une limite forte à l'étude de cette variation est liée au fait que dans les études précédentes, les valeurs de ces *fréquences optimales* séparant les canaux de fréquence ont été estimées par le biais d'inspections purement visuelles des graphiques de résultats. Il va de soi que pour être en mesure d'évaluer le degré de variation d'une telle mesure, il convient de mettre en place une méthode de détermination automatique qui ouvrirait la voie au calcul d'un grand nombre d'estimations différentes en étudiant aussi bien les effets liés à la taille globale du corpus que ceux afférents à la dispersion des résultats lorsqu'on sélectionne aléatoirement des sous-ensembles distincts du corpus d'origine. Nous présentons dans cet article la méthode que nous avons développée dans cette perspective et donnons une illustration préliminaire des résultats à partir de la comparaison avec les études antérieures (Ueda & Nakajima, 2017; Grange & Culling, 2018; Ming & Holt, 2009)

2 Méthode

Les analyses acoustiques et statistiques ont été réalisées dans l'environnement Matlab. Les scripts d'analyse sont disponibles sur un dépôt OSF [1].

1. Lien vers le dépôt OSF en lecture : https://page.hn/9ij6kk

2.1 Corpus

La base de données utilisée est la *Clarity Speech Database* (Graetzer *et al.*, 2021) qui contient des enregistrements de parole en anglais en accès libre (échantillonnés à 44.1 kHz sur 32 bits au format WAV). La base de données complète est composée d'environ 10000 phrases issues du *British National Corpus (BNC)*, lues par 40 locuteurs et locutrices de l'anglais Britannique. Dans le cadre de l'étude présentée ici, un échantillon aléatoire de 1600 phrases en a été extrait dans le but de cibler une durée totale équivalente à celle utilisée dans les précédentes études (environ 1 h de parole).

2.2 Paramètrage acoustique des signaux

Préalablement à l'analyse statistique des signaux, nous procédons à une paramétrisation acoustique comparable à celles qui ont été utilisées dans les travaux précédents (Ueda & Nakajima, 2017; Grange & Culling, 2018). Les signaux extraits de chaque fichier sont concaténés les uns aux autres. Le signal concaténé est soumis à un filtrage passe-bas (fréquence de coupure 8 kHz) et sous-échantillonné à 16 kHz. Les enveloppes de modulation temporelle des signaux sont extraites à partir d'un banc de filtres dont la largeur croît de manière logarithmique avec la fréquence centrale (canaux de largeur $\frac{1}{4}$ d'ERB, Moore & Glasberg, 1983, ce qui correspond à 116 canaux spectraux allant jusqu'à la fréquence supérieure maximale de 8 kHz). Ces enveloppes subissent une rectification demi-onde puis un filtrage passe-bas avec une fréquence de coupure de 50 Hz (fréquence d'échantillonnage 100 Hz). Les signaux d'enveloppe résultants sont ensuite élevés au carré et convertis en notes centrées réduites (*z-scores*). Cette chaîne de paramétrage permet de procéder à une analyse des co-modulations d'énergie entre les bandes de fréquence (corrélation entre les enveloppes d'énergie des canaux).

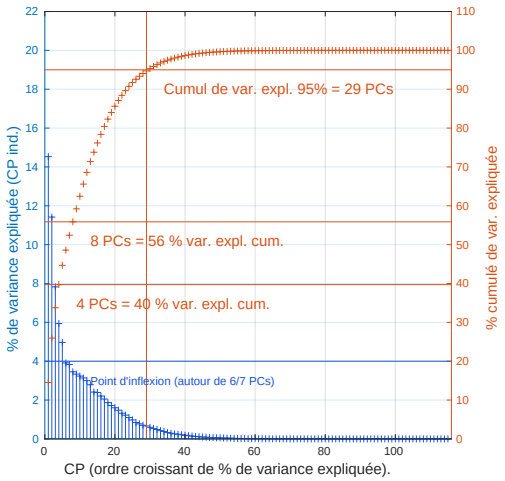


FIGURE 1 – Graphe des valeurs propres (% de variance expliquée) issu de l'Analyse Factorielle (en bleu : % associés à chaque CP individuelle — en rouge : % cumulés). La CP pour laquelle on atteint un cumul égal à 95% est indiquée, ainsi que le % de variance cumulée expliquée pour resp. 4 et 8 Composantes Principales.

La matrice de données résultante, composée de 116 canaux de fréquence, correspond aux modulations temporelles de l'enveloppe de chaque canal spectral au cours du temps. Elle est transférée vers un outil statistique d'Analyse en Composantes Principales afin de procéder à une Analyse Factorielle.

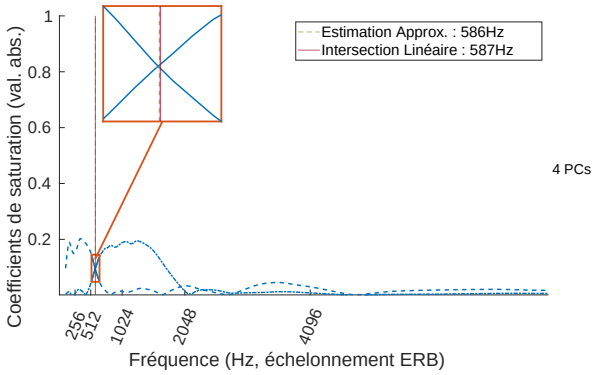
2.3 Analyse Factorielle

L'*Analyse Factorielle* est une méthode descriptive d'analyse de données qui repose sur la technique d'*Analyse en Composantes Principales* (ACP). Elle permet une étude simultanée de plus de 2 dimensions (analyse multivariée). L'objectif est de représenter l'essentiel de l'information contenue dans un tableau de données quantitatif en réduisant le nombre de facteurs explicatifs. Le principe

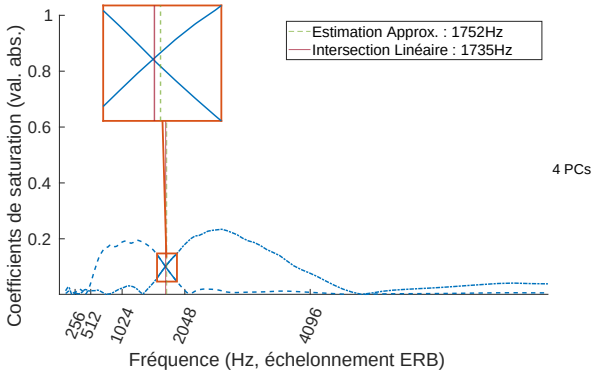
est de transformer des variables liées (statistiquement corrélées entre elles) en nouvelles variables *synthétiques*. Les Composantes Principales (CP) sont donc regroupées en facteurs abstraits.

Concrètement, les variables initiales sont représentées dans un nouvel espace de facteurs définis par les vecteurs propres de la matrice de corrélations. L'hypothèse sous-jacente à l'application de cette méthode sur des signaux sonores est que certains canaux spectraux contiendraient des informations redondantes et qu'il serait alors économique de restreindre l'analyse perceptive à une séparation en zones de fréquences étant maximalement informatives (donc minimalement redondantes). En celà, l'Analyse Factorielle permettrait d'identifier les canaux de fréquence optimaux pour différencier de manière parcimonieuse les propriétés sonores distinctives d'un corpus. Une description plus précise de la procédure d'ACP mise en oeuvre est disponible dans Duniec *et al.* (2022). Pour information, le graphe des valeurs propres indiquant le pourcentage cumulé de variance expliquée en fonction du nombre de CP est représenté dans la Fig. 1.

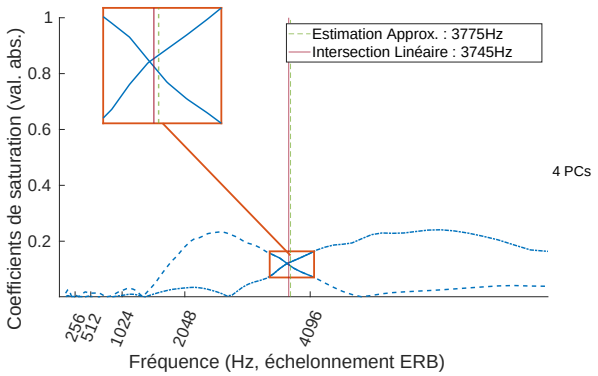
Tout comme dans les travaux antérieurs (Ueda & Nakajima, 2017; Grange & Culling, 2018; Duniec *et al.*, 2022), ces valeurs sont représentées sous forme de courbes de coefficients de saturation dont la valeur absolue fait ressortir quelles sont les gammes de fréquences qui sont maximalement corrélées entre elles (cf. Duniec *et al.*, 2022, pour une illustration). Ce regroupement indique quelles fréquences sont associées / co-modulées. Les gammes de fréquences correspondantes sont, dans le cadre de l'Analyse Factorielle, associées à un facteur synthétique / une Composante Principale. C'est sur la base de ces ensembles de courbes de coefficients de saturation que nous procédons ensuite à une estimation automatique de ces frontières pour chaque condition associée au nombre de Composantes Principales retenues.



(a) 1ère frontière



(b) 2e frontière



(c) 3e frontière

FIGURE 2 – Illustration de la procédure de détermination des intersections entre les courbes adjacentes de coefficients de saturation par prédiction linéaire.

Les intervalles des courbes de coefficients de saturation pour lesquels la valeur absolue est relativement élevée sont interprétées comme représentant des zones spectrales co-modulées en amplitude, lesquelles sont considérées dans cette perspective comme ne fournissant que peu d'in-

formations perceptives distinctes. Déterminer la localisation de leurs frontières fournirait donc une information sur la zone optimale de découpage spectral qui pourrait par exemple être utilisée dans un implant cochléaire dans le but de dissocier de manière optimale les gammes de fréquence maximalement porteuses d'information pour un nombre de canaux déterminé.

2.4 Estimation des frontières entre canaux

La procédure de détermination des frontières entre chaque paire de courbes de coefficients de saturation adjacentes est ensuite appliquée.

On procède d'abord à une première phase de traitements préparatoires. Une interpolation linéaire des courbes de coefficients de saturation dans un rapport de $\frac{10}{1}$ vise à accroître le nombre de points de mesure sur l'échelle des fréquences. On estime ensuite la localisation en fréquence du pic de la courbe de coefficients de saturation associée à chaque Composante Principale. Pour chacune de ces courbes, on repère approximativement deux frontières (respectivement inférieure et supérieure au pic) à partir d'un critère de valeur du coefficient qui descende à 25% de la valeur maximale observée au pic en se déplaçant du pic vers chacun des bords de la courbe. On classe enfin les vecteurs associés aux courbes de coefficient de saturation par valeur croissante de fréquence associée au pic, ce qui permet de déterminer quelles sont les courbes adjacentes en termes de canaux de fréquence.

On procède enfin, pour chaque paire de courbes adjacentes, à la détermination des coordonnées de leur point d'intersection par modélisation linéaire [2] limitée à la zone probable du croisement. La Fig. 2 illustre cette dernière partie de la procédure pour une Analyse Factorielle retenant 4 Composantes Principales (3 frontières). Ceci repose sur la modélisation par une fonction linéaire de chacune des deux portions de courbes. Pour terminer, on estime les coordonnées de leur point d'intersection dans un espace *Valeur du coefficient de saturation en dB* $\sim$ *Fréquence en Hz* (après échelonnement ERB).

3 Résultats préliminaires

Nous proposons ici une première comparaison avec les données de la littérature. Il est important de noter que les divergences constatées pourraient trouver leur source dans différents paramètres. Le contenu du corpus de parole utilisé peut impacter les résultats. Ainsi, ces études reposent sur des corpus de parole comparables en termes de type de contenu (phrases, multi-locuteurs, langue anglaise) et de conditions de collecte mais qui sont néanmoins différents. Par ailleurs, des aspects précis de la méthode appliquée pour procéder à la paramétrisation acoustique puis à la mise en œuvre de l'Analyse Factorielle peuvent également jouer (par ex. Ueda & Nakajima, 2017 procédent à une décomposition spectrale initiale en seulement 20 canaux alors que nous adoptons, conformément à Grange & Culling, 2018, une décomposition spectrale en plus de 100 canaux). Enfin, le recours à une procédure automatisée vs. visuelle pourrait affecter la précision des estimations. Concernant ce dernier point cependant, les algorithmes que nous avons implémentés semblent produire des résultats conformes aux estimations visuelles.

Notre objectif principal dans cet article est de chercher à mettre en évidence deux points importants : (1) la méthode automatique mise en œuvre fournit des résultats cohérents en termes d'estimation de la

2. Des tests ont également été effectués avec des modélisations polynomiales d'ordre 2 et 3 mais les résultats de la modélisation linéaire étaient généralement plus conformes aux observations visuelles que les fonctions d'ordre supérieur.

fréquence à laquelle le croisement des courbes adjacentes se fait et (2) les données recueillies varient parfois fortement entre les études considérées, ce qui justifie pleinement la nécessité de procéder à une exploration de la variation associée à ces mesures.

Nous procédons à une comparaison de nos résultats avec les données de la littérature qui sont disponibles et qui ont eu recours à la même méthode de traitement (Analyse Factorielle), donc respectivement pour 4 et 8 canaux. Les données de Ming & Holt (2009) portent sur un découpage en 6 canaux mais sont fondées sur une méthode de traitement profondément différente, nous ne les considérons pas directement dans cet article pour des raisons d'espace disponible.

3.1 Résultats pour 4 canaux de fréquence

Les résultats pour 4 canaux de fréquence sont disponibles dans les principales études auxquelles nous nous sommes référés (Ueda & Nakajima, 2017; Grange & Culling, 2018), lesquelles sont toutes les deux fondées sur des ACP réalisées sur les enveloppes d'énergie malgré des choix différents en termes de nombre de canaux initiaux pour la décomposition spectrale (20 pour Ueda & Nakajima, 2017 vs. plus de 100 pour l'étude de Grange & Culling, 2018). Les valeurs numériques de localisation des frontières pour chacun des articles ont été déterminées en utilisant un logiciel d'extraction de données quantitatives à partir de graphes (`g3data` [3]) sur la base des figures publiées. Nous avons utilisé cet outil pour estimer / extraire les coordonnées de fréquence des points d'intersection des courbes adjacentes. Les données de comparaison sont présentées dans la Table 1. On peut constater que les résultats sont cohérents les uns par rapport aux autres mais que des divergences émergent et peuvent pour certaines atteindre 1 à 2 demi-tons d'écart.

Ces degrés de divergence peuvent paraître parfaitement acceptables et pourraient ne relever que de variations attendues dans tout travail quantitatif même s'il conviendra de documenter cette variation. La comparaison avec les résultats publiés pour 8 canaux (cf. *infra.* et Table 2) fait par contre ressortir des différences nettement plus marquées.

TABLE 1 – Estimations (en Hertz) de la localisation des frontières optimales entre canaux de fréquence et écarts mesurés par rapport aux données des travaux antérieurs (en demi-tons) pour 4 Composantes Principales (respectivement : Données de Ueda & Nakajima (2017) ; Données de Grange & Culling (2018) ; Notre estimation par modélisation linéaire ; Différences mesurées entre les estimations issues de la littérature et nos données).

	1/2	2/3	3/4
Ueda & Nakajima (2017)	540	1720	3300
Grange & Culling (2018)	573	1570	3827
Nos observations	587	1735	3745
Écart / Ueda & Nakajima (2017, demi-tons)	1.44	0.15	2.19
Écart / Grange & Culling (2018, demi-tons)	0.41	1.73	-0.38

3.2 Résultats pour 8 canaux de fréquence

Les résultats pour 8 canaux de fréquence (cf. Table 2) ne sont disponibles que pour l'étude de Grange & Culling (2018). Ils ont été déterminés en utilisant le même logiciel d'extraction de

3. https://github.com/pn2200/g3data/

données quantitatives à partir de graphes (`g3data`). On peut constater que certaines estimations sont considérablement plus divergentes, notamment dans les basses et moyennes fréquences (avec des écarts de l'ordre de 10 à 15 demi-tons –autour d'une octave–, ainsi que dans les hautes fréquences –3 à 4 demi-tons), ce qui laisse entrevoir des potentialités de variation considérables de ces propriétés statistiques. Il est notable que la procédure de traitement acoustique et statistique mise en œuvre dans notre approche est en tous points comparable à celle implémentée par Grange & Culling (2018), ce qui laisse supposer un potentiel de variation considérable qu'il conviendra d'explorer.

TABLE 2 – Estimation de la localisation des frontières optimales entre canaux de fréquence pour 8 CP (en Hz). Comparaison avec les observations de Grange & Culling (2018).

	1/2	2/3	3/4	4/5	5/6	6/7	7/8
Données de Grange & Culling (2018)	442	652	1159	1518	1916	2749	4104
Nos estimations	197	332	678	1474	2099	3377	5085
Écart (demi-tons)	-14.00	-11.69	-9.29	-0.51	1.58	3.56	3.71

4 Discussion

Nos observations contrastent assez nettement avec les résultats antérieurs en termes de correspondance des frontières de fréquences telles qu'elles peuvent être déduites de l'Analyse Factorielle de signaux de parole. Si les résultats pour 4 canaux semblent acceptables en termes de dispersion naturelle des mesures empiriques, les résultats obtenus avec 8 canaux mettent en évidence un degré de variation qui peut atteindre plus d'une octave dans certaines gammes de fréquence, ce qui aurait de toute évidence des conséquences perceptives notables.

En outre, l'hypothèse de Ueda & Nakajima (2017) concernant la stabilité de ces mesures pour 8 langues distinctes semble assez spéculative si l'on considère le degré de variation que nous avons commencé à documenter ici. En effet, cette comparaison montre que, pour des données reposant sur une méthode d'analyse acoustique et statistique équivalente, l'utilisation d'une base de données différente dans la même langue et pour un corpus de phrases multi-locuteurs comparable en taille, débouche sur des divergences de valeurs des frontières dont l'empan est considérable.

À tout le moins, même si toute variation est prévisible et attendue dans le cadre de la mesure empirique de phénomènes naturels, ces données mettent en évidence un potentiel de variation important que nous pourrons explorer de manière systématique avec la procédure présentée dans cet article.

Remerciements

Ce travail a reçu le soutien du programme Recherche – Formation – Innovation « Ouest Industries Créatives » (RFI-OIC, Région Pays de la Loire) par une allocation doctorale attribuée à AD.

Références

DUNIEC A., CROUZET O. & DELAIS-ROUSSARIE E. (2022). Analyse factorielle de signaux musicaux : comparaison avec les données de parole dans la perspective de l'hypothèse du codage efficace et de l'application aux implants cochléaires. In O. CROUZET, E. DELAIS-ROUSSARIE, A. DUNIEC, L. LEPRIEUR, P. L. ROHRER, M. TAHON, J. WOTTAWA, M. BRABANT, H. RIGUIDEL, N. BARBOT, S. GIBET, D. LOLIVE & A. SINI, Éds., *Actes des 34e Journées d'Études sur la Parole – JEP2022*, Île de Noirmoutier, France : Nantes Université / Le Mans Université / IRISA / CNRS International Speech Communication Association – ISCA archive.

GRAETZER S., AKEROYD M. A., BARKER J., COX T. J., CULLING J. F., NAYLOR G., PORTER E. & MUÑOZ R. V. (2021). Dataset of british english speech recordings for psychoacoustics and speech processing research : The clarity speech corpus. *Data in Brief*, p. 107951.

GRANGE J. & CULLING J. (2018). The factor analysis of speech : Limitations and opportunities for cochlear implants. *Acta Acustica united with Acustica*, **104**, 835–838. DOI : 10.3813/AAA.919253.

KLUENDER K. R., STILP C. E. & KIEFTE M. (2013). Perception of Vowel Sounds Within a Biologically Realistic Model of Efficient Coding. In G. S. MORRISON & P. F. ASSMANN, Éds., *Vowel Inherent Spectral Change*, Modern Acoustics and Signal Processing, p. 117–151. Berlin, Heidelberg : Springer. DOI : 10.1007/978-3-642-14209-3_6.

MCDERMOTT J. H. & SIMONCELLI E. P. (2011). Sound Texture Perception via Statistics of the Auditory Periphery : Evidence from Sound Synthesis. *Neuron*, **71**(5), 926–940. DOI : 10.1016/j.neuron.2011.06.032.

MING V. L. & HOLT L. L. (2009). Efficient coding in human auditory perception. *The Journal of the Acoustical Society of America*, **126**(3), 1312–1320. DOI : 10.1121/1.3158939.

MOORE B. C. J. & GLASBERG B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*, **74**, 750–753.

PLOMP R., POLS L. C. W. & VAN DE GEER J. P. (1967). Dimensional Analysis of Vowel Spectra. *The Journal of the Acoustical Society of America*, **41**(3), 707–712. DOI : 10.1121/1.1910398.

SIMONCELLI E. P. & OLSHAUSEN B. A. (2001). Natural Image Statistics and Neural Representation. *Annual Review of Neuroscience*, **24**(1), 1193–1216. DOI : 10.1146/annurev.neuro.24.1.1193.

SMITH E. & LEWICKI M. S. (2005). Efficient Coding of Time-Relative Structure Using Spikes. *Neural Computation*, **17**(1), 19–45. DOI : 10.1162/0899766052530839.

SMITH E. C. & LEWICKI M. S. (2006). Efficient auditory coding. *Nature*, **439**(7079), 978–982. DOI : 10.1038/nature04485.

UEDA K. & NAKAJIMA Y. (2017). An acoustic key to eight languages/dialects : Factor analyses of critical-band-filtered speech. *Scientific Reports*, **7**, 42468. DOI : 10.1038/srep42468.