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

Duniec, Agnieszka Crouzet, Olivier Delais-Roussarie, Elisabeth

Laboratoire de Linguistique de Nantes – LLING / UMR6310, Nantes Université / CNRS, France

Keep informed / Download

9

Introduction

- L'hypothèse du codage efficace [1] postule que les systèmes perceptifs sont optimalement adaptés aux propriétés statistiques des signaux naturels.
- Selon cette approche, les systèmes sensoriels auraient évolué de manière à encoder des signaux de notre environnement d'une façon optimale conformément à la théorie de l'information ;
- Ceci dans le but de transmettre un maximum d'information en utilisant le minimum de resources ;
- Implications possibles pour la localisation des frontières optimales dans les implants cochléaires [2];

Travaux antérieurs : Ming and Holt (2009) [3]

- L'identification de parole vocodée avec 6 canaux spectraux est globalement meilleure avec des frontières spectrales determinées sur les bases du "codage efficace" qu'avec des frontières cochléotopiques organisées suivant l'échelle logarithmique ;
- En général, meilleure reconnaissance des mots à l'intérieur des phrases et meilleure identification des phonèmes dans des non-mots ;
- Pas de différence significative pour les voyelles (effet de plafonnement possible) et les fricatives ;

Travaux antérieurs : Ueda and Nakajima (2017) [4]

- Analyse factorielle de signaux de parole dans 8 langues différentes [4];
  - 20 canaux de fréquence ;
  - Extraction des enveloppes d'énergie de ces canaux ;
  - Analyse en Composantes Principales afin de déterminer le nombre de dimensions optimal pour représenter les signaux ;
  - Fusion des canaux qui covarient en une seule dimension ;
  - Les composantes principales sont soumises à une rotation varimax afin de maximiser l'indépendance des facteurs et de fournir en sortie des vecteurs de coefficients de saturation (factor loading) ;
  - Chaque vecteur de coefficients représente une zone de fréquence du spectre modulée de manière maximalement indépendante des autres zones ;
  - Identification des bandes de fréquences de coupure à partir de ces vecteurs.

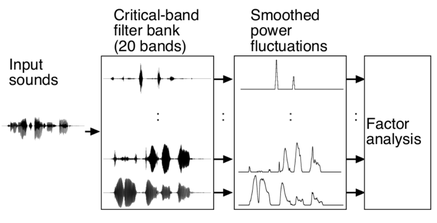


Figure 1: Diagramme schématique de traitement des données [4]

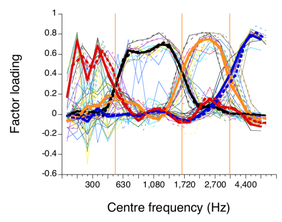


Figure 2: Vecteurs de coefficients de saturation pour la parole avec 4 composantes principales (Ueda and Nakajima [4]).

Travaux antérieurs : Grange and Culling (2018) [2]

- Environ 100 canaux de fréquence ;
  - Enregistrements uniquement en anglais ;
  - Procédure statistique similaire;
  - Comparaison avec des données perceptives (simulations d'implants cochléaires);

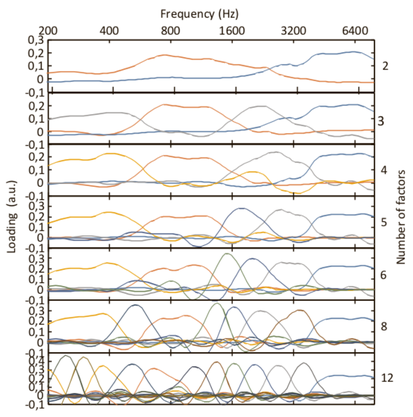


Figure 3: Vecteurs de coefficients de saturation en fonction du nombre de composantes prinipales pour la parole [2].

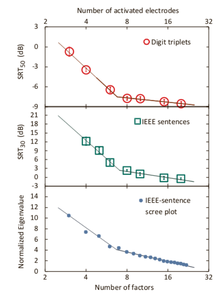


Figure 4: Graphique des valeurs propres ou Scree plot (en bas). Le point d'inflexion est comparé aux données de la performance en perception : seuil d'identification pour les triplets de chiffres (en haut) et pour les phrases simples (au milieu) [2].

Limites des études antérieures

1. **Variation ?** Aucune estimation du degré de variation des frontières "optimales" issues des analyses:
   - En fonction de la composition du corpus;
   - En fonction de la taille du corpus;
   - En fonction de la durée des items concaténés pour la constitution du corpus...
   - **Question liée:** Quel est le degré de variation de ces frontières sur la durée du corpus ?
2. **Méthode d'estimation des frontières ?** Détermination des frontières "optimales" par inspection visuelle des graphes;
   - Une méthode automatisée est nécessaire pour pouvoir caractériser la variation des résultats;
   - Ce qui permettrait de prélever un grand nombre d'estimations en faisant varier différents paramètres d'analyse (nombre de canaux, durée des items)...
   - ...et en sélectionnant un grand nombre de sous-échantillons aléatoires différents;

Méthode

- Nous présentons ici la méthode que nous avons développée;
- Et donnons une illustration préliminaire des résultats en comparant nos données avec les études antérieures [2, 3, 4];
- Analyses acoustiques et statistiques dans Matlab;
- Scripts disponibles sur un dépôt OSF : https://page.hn/9ij6kk;
- Réplicabilité des analyses rendue possible par la mise à disposition des scripts et l'utilisation d'une base de données libre;

Corpus

- Un corpus libre de parole lue [5] (*Clarity Speech*, https://doi.org/10.17866/rd.salford.16918180);
- Un échantillon aléatoire de 1,600 phrases (parmi un total de 10k phrases en tout (env. 4,500 s). Extraites du *British National Corpus* (BNC), produites par 40 locuteurs de l'anglais Britannique ;
- Les fichiers audio sont stockés en monophonique sur 32-bit (format WAV) et échantillonnés à 44.1 kHz;

Paramétrage acoustique des signaux

- Comparable avec Grange and Culling (2018) [2];
- Concaténation des fichiers sélectionnés aléatoirement;
- Filtrage passe-bas (8 kHz);
- Passage à travers un banc de filtres Gammatone — canaux de largeur $\frac{1}{4}$ d'ERB [6], ce qui correspond à 116 canaux spectraux allant jusqu'à la fréquence supérieure de 8 kHz;
- Extraction des modulations d'amplitude en-dessous de 50 Hz par rectification demi-onde, filtrage passe-bas puis élévation au carré et centration / réduction;
- Matrice résultante : modulations temporelles de l'enveloppe d'énergie (normalisée) de chacun des 116 canaux spectraux;
- Fournie en entrée de l'ACP (Temps = observations, Echantillon = individus);

Analyse Factorielle (Figs. 5 et 6)

- Repose sur une ACP (Analyse en Composantes Principales);
- Buts de l'AF : Regrouper des variables liées (statistiquement corrélées entre elles) en nouvelles variables *synthétiques*;
- Les Composantes Principales (CP) sont donc regroupées en *facteurs abstraits*. Par la suite, on parle de CP pour désigner ces facteurs;
- On trace les courbes de coefficients de saturation en faisant varier le nombre de CP sélectionnées;
- Les intersections entre ces courbes représentent potentiellement les *frontières optimales entre canaux de fréquence*;

Estimation initiale des frontières de fréquence (Fig. 7)

- Identification des courbes adjacentes par la mise en relation de la fréquence du pic et du rang de la CP;
- Ordonnancement des courbes en fonction de la fréquence du pic;
- Repérage du pic de chaque courbe de coefficients de saturation;
- Localisation de frontières initiales (resp. inférieure et supérieure) à partir d'un critère de diminution de l'amplitude par rapport au pic (25%);
- Calcul –pour chaque paire de courbes adjacentes– d'une valeur initiale de frontière à mi-chemin entre (1) la valeur de la frontière située en dessous du pic du canal supérieur et (2) la valeur la frontière située en dessus du pic du canal inférieur;

Détermination de l'intersection (Fig. 8)

- Approximation linéaire de chacun des deux segments de courbe ($y = ax + b$);
- Détermination de l'intersection entre les deux segments de droite par résolution du système d'équations linéaires;
- Des comparaisons ont été réalisées avec une modélisation par des polynômes d'ordre 2 mais elles donnent souvent des résultats moins conformes aux estimations visuelles;
- Extraction des données quantitatives des études antérieures avec `g3data` (https://github.com/pn2200/g3data/);

Comparaisons avec la littérature : 4 canaux

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

Comparaisons avec la littérature : 8 canaux

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

Résultats de l'ACP

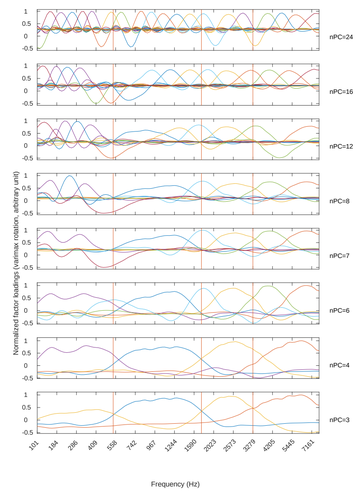


Figure 5: Coefficients de saturation (*factor loadings*) issus des ACP réalisées sur un échantillon des signaux de parole de la base de données *Clarity* [5], durée totale 1h, fréq. max. : 8k Hz. Les segments verticaux représentent les frontières identifiées par [4]

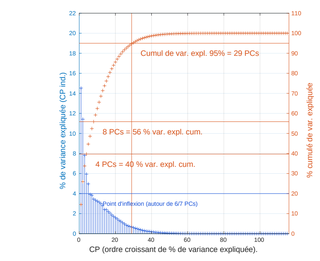
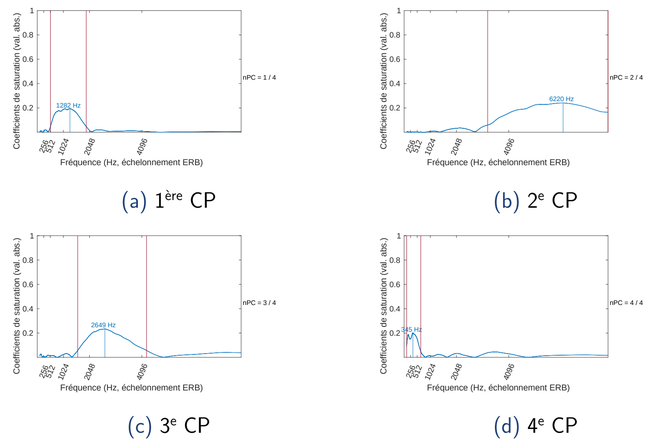


Figure 6: Graphe des valeurs propres (% de variance expliquée) issu de l'Analyse Factorielle (en bleu: % associés à chaque CP individuelle — en rouge: % cumulés). La CP pour laquelle on atteint un cumul égal à 95% est indiquée, ainsi que le % de variance cumulée expliquée pour resp. 4 et 8 Composantes Principales.

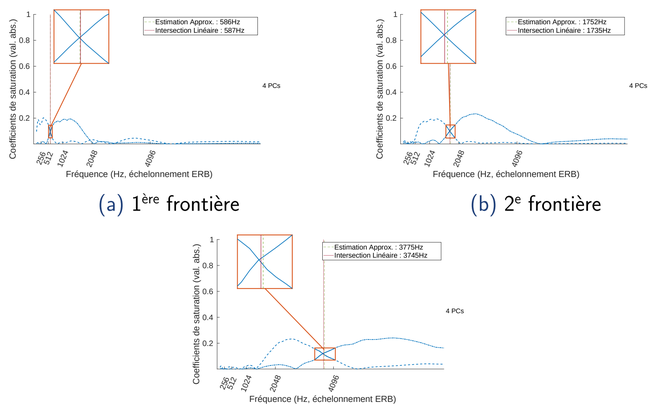
Phase préparatoire



(a) 1ère CP (b) 2e CP (c) 3e CP (d) 4e CP

Figure 7: Illustration de la procédure de localisation initiale des frontières inférieure et supérieure associées à chaque CP. Les courbes individuelles représentées correspondent à celles du panneau nPC = 4 de la Fig. 5.

Détermination des intersections



(a) 1ère frontière (b) 2e frontière (c) 3e frontière

Figure 8: Illustration de la procédure de détermination des intersections entre les courbes adjacentes de coefficients de saturation par prédiction linéaire dans une zone restreinte autour de leur fréquence d'intersection estimée. Le trait vertical rouge (plein) représente l'estimation finale par modélisation linéaire de l'intersection. Le trait vertical vert (pointillés) indique l'estimation initiale du centre de la zone de croisement probable.

Comparaisons avec la littérature : 6 canaux

Table 3: Estimation de la localisation des frontières optimales entre canaux de fréquence pour 6 CP (en Hz). Comparaison avec les observations de Ming & Holt (2009).

	1/2	2/3	3/4	4/5	5/6
Estimation cochléotopique (Ming & Holt, 2009)	475	768	1303	2258	4028
Estimation de Ming & Holt (2009)	461	677	904	1184	2019
Notre estimation	597	1508	2056	3275	5030
Différence / cochléotopique (ST)	3.95	11.68	7.90	6.44	3.85
Différence / Ming & Holt (ST)	4.47	13.87	14.22	17.62	15.80

Discussion

- Grande diversité de résultats en fonction des sources (corpus, méthodes...);
- Différences considérables de méthode entre les données de [2, 4] (méthode statistique ACP) et celles de [3] (implémentation de méthodes issues de la théorie de l'information);
- Variation attendue mais nos résultats justifient l'exploration approfondie de cette variation;

Remerciements

Agnieszka Duniec a bénéficié d'une allocation doctorale (2019–2023) du RFI Ouest Industries Créatives (RFI–OIC, Région Pays de la Loire) & Nantes Université.

Références

[1] E. C. Smith and M. S. Lewicki. "Efficient auditory coding". In: *Nature* 439.7079 (2006), pp. 978–982. — **[2]** J. Grange and J. Culling. "The Factor Analysis of Speech: Limitations and Opportunities for Cochlear Implants". In: *Acta Acustica united with Acustica* 104 (Sept. 2018), pp. 835–838. — **[3]** V. L. Ming and L. L. Holt. "Efficient coding in human auditory perception". In: *The Journal of the Acoustical Society of America* 126.3 (Sept. 2009), pp. 1312–1320. — **[4]** K. Ueda and Y. Nakajima. "An acoustic key to eight languages/dialects: Factor analyses of critical-band-filtered speech". In: *Scientific Reports* 7 (Feb. 2017), p. 42468. — **[5]** S. Graetzer et al. "Dataset of British English speech recordings for psychoacoustics and speech processing research: The Clarity Speech Corpus". In: *Data in Brief* (). — **[6]** B. C. J. Moore and B. R. Glasberg. "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns". In: *The Journal of the Acoustical Society of America* 74 (1983), pp. 750–753. —

olivier.crouzet@univ-nantes.fr