



Rôle des indices acoustiques dans la compréhension de la parole chez les individus normoentendants et malentendants

Léo VARNET Chercheur postdoctoral - École Normale Supérieure

Laboratoire des Systèmes Perceptifs (LSP) CNRS UMR 8248 - Département d'Études Cognitives (DEC)

29, rue d'Ulm - 75230 Paris Cedex 05

1

Indices acoustiques et compréhension de la parole

Pour comprendre un son de parole, par exemple lors d'une conversation, le cerveau humain doit tout d'abord convertir le signal acoustique continu en une série d'unités linguistiques discrètes : les phonèmes, les syllabes, puis les mots. Il s'appuie pour cela sur certains éléments transitoires du son, appelés « indices acoustiques », qui sont caractéristiques du phonème auquel ils appartiennent et permettent d'identifier ce dernier de manière univoque. Une étape première et essentielle de la reconnaissance d'un mot prononcé par notre interlocutrice ou notre interlocuteur consiste donc à extraire ces indices acoustiques du son de parole et à les combiner en un percept phonémique ¹. Autrement dit, il existe un mécanisme de traduction entre les réalisations acoustiques des phonèmes et leurs représentations mentales abstraites dans le cerveau.

Ainsi, depuis les années 50, un large corpus d'études en psycholinguistique a démontré que l'identification des consonnes plosives telles que /d/ ou /g/ repose sur la détection d'indices multiples et redondants. Ces indices se situent notamment au niveau des attaques des premier, deuxième et troisième formants (F1, F2 et F3, cf. Figure 1), qui correspondent à des concentrations

d'énergie localisées en temps et en fréquence ²⁻⁴. C'est l'extraction et le traitement par le système auditif de ces détails extrêmement fins du son de parole qui permet au cerveau de reconnaître la syllabe effectivement prononcée.

La parole apparaît alors comme une forme de « code », reliant le signal physique et son contenu linguistique. L'intelligibilité du message repose donc en premier lieu sur deux éléments : 1) une transmission correcte des indices acoustiques et 2) leur bonne extraction par le système auditif de l'auditrice ou de l'auditeur. À l'inverse, une altération de ces indices au sein du signal (par exemple lors d'une communication dans un environnement bruité) ou des déficiences lors de leur traitement (comme c'est le cas chez les personnes souffrant de troubles auditifs) peuvent compromettre gravement la compréhension.

Néanmoins, le son de parole est un signal complexe, constitué d'une multitude d'informations redondantes, certaines étant effectivement utilisées par le cerveau pour différencier les phonèmes, tandis que d'autres n'ont aucune incidence sur la compréhension. Dans l'exemple de la catégorisation des phonèmes /d/ et /g/ mentionné ci-dessus (Figure 1), par exemple, l'attaque des second et troisième formants (F2 et F3) constitue un indice acoustique primaire, jouant un rôle essentiel dans la compréhension, tandis que l'attaque du premier formant (F1) est un indice secondaire, de moindre importance dans la décision ⁵. Cette redondance des indices permet au système auditif d'assurer l'intelligibilité du message avec une grande robustesse. Il est ainsi capable d'adapter sa stratégie pour s'appuyer sur les indices les plus fiables dans un contexte d'écoute donné ⁶.

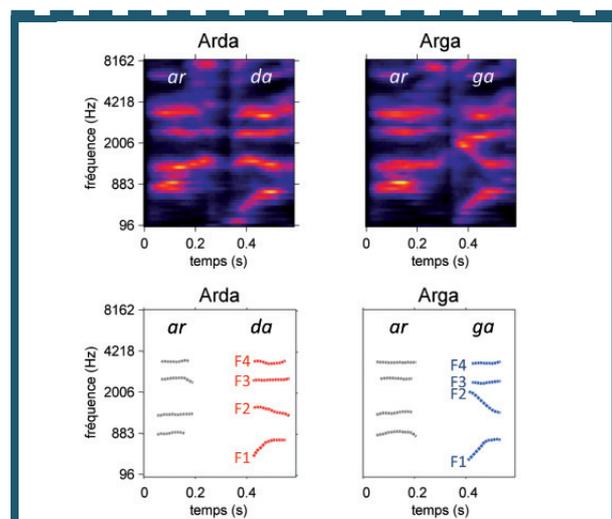


Figure 1. En haut : représentation temps-fréquence de deux productions de « Arda » et « Arga ». Les régions présentant la plus forte concentration d'énergie (formants) apparaissent en couleurs chaudes. En bas : trajectoires des quatre premiers formants F1, F2, F3 et F4 (les formants correspondant aux syllabes « da » et « ga » sont représentés en rouge et en bleu, respectivement). La configuration exacte des attaques de F1, F2 et F3 détermine la perception de la deuxième syllabe comme un /da/ ou un /ga/.

2

Adaptations de la stratégie d'écoute lors de la compréhension de la parole dans le bruit

C'est le cas, notamment, pour des situations d'écoute réelles, dans lesquelles la communication est le plus souvent altérée par la présence d'un bruit de fond provenant de l'environnement sonore (discussions alentour, bruit ambiant...) ou par des problèmes de transmission (« friture » lors d'une communication téléphonique). Bien que ces conditions soient particulièrement défavorables pour la compréhension, la capacité du système auditif à utiliser un indice ou l'autre selon le contexte d'écoute lui assure une fiabilité jusqu'à présent inégalée par les systèmes de reconnaissance vocale. Ce mécanisme de « re-pondération » est difficile à mettre en évidence expérimentalement. En manipulant artificiellement les différents indices impliqués dans une tâche de catégorisation /d/-/t/ (trait de voisement), Serniclaes et Arrouas ⁷ ont démontré que les auditrices et auditeurs s'appuyaient exclusivement sur l'indice principal lorsque les phonèmes étaient présentés dans le silence.

Au contraire, en présence d'un bruit rendant cet indice primaire ambigu, leurs stratégies perceptives se reportaient sur l'usage d'indices secondaires, plus résistants aux dégradations.

L'absence d'indice acoustique absolument nécessaire à l'intelligibilité et la capacité d'adaptation extrêmement rapide du système auditif en réponse à la présence de bruit concurrent sont deux concepts pris en compte dans le modèle des canaux indépendants⁸⁻¹⁰. Ce modèle général de compréhension de la parole fournit la base théorique du calcul de l'Indice d'Articulation, une métrique d'intelligibilité de parole dans le bruit produisant des prédictions extrêmement fiables⁹. Ici, la plasticité du traitement dépend non seulement de la simple présence de bruit mais également de la composition spectrale de ce dernier : lorsque le spectre de parole est partiellement masqué par le bruit, le système auditif est en mesure de se reporter sur des bandes de fréquence pour lesquelles le niveau de bruit est localement plus favorable. Ce point est illustré par un cas extrême : les sons de parole restent intelligibles après suppression totale des fréquences supérieures à 1550 Hz, aussi bien qu'après suppression totale des fréquences inférieures à 1550 Hz^{10,11}. C'est donc que des indices hautes - ou basses - fréquences sont utilisés selon le type de dégradation (voir Figure 2).

3

Effets des pertes auditives sur les stratégies d'écoute

Un tel mécanisme de sélection est également utile lorsque les sources de dégradation du signal sont non plus externes à l'auditrice ou à l'auditeur, comme pour le bruit de fond, mais internes, comme dans le cas de la malentendance. En effet, les pertes auditives peuvent affecter différemment l'information contenue dans plusieurs régions fréquentielles. Par exemple les surdités « en pente de ski », qui affectent les fréquences les plus aiguës du son en laissant intactes les fréquences graves, constituent l'un des profils audiométriques les plus répandus chez les personnes souffrant de pertes neurosensorielles. La situation des individus souffrant de ce type de pertes est alors relativement similaire, en première approximation, à celle des participantes et participants confrontés aux stimuli filtrés passe-bas dans l'expérience décrite Figure 2. Ainsi, même dans le cas où leurs pertes auditives les priveraient de l'usage des fréquences hautes, ceux-ci sont en mesure de se reporter sur l'usage d'indices basse-fréquences et, ainsi, de comprendre, bien qu'imparfaitement, les sons auxquels ils sont confrontés¹².

Dans ce contexte, le rôle de l'appareillage auditif est d'amplifier les fréquences correspondant aux pertes de sensibilité de l'utilisatrice ou de l'utilisateur, afin de restaurer l'audibilité des indices acoustiques

localisés dans ces régions. Malheureusement, en pratique, les bénéfices de l'amplification en termes d'intelligibilité apparaissent souvent limités, particulièrement dans les cas de surdités en pente de ski^{13,14}. Ainsi, rétablir une sensibilité auditive normale chez les individus malentendants ne garantit pas un rétablissement de l'intelligibilité. De plus, les personnes appareillées commettent certaines erreurs systématiques d'identification de phonèmes plus fréquemment que les personnes normoentendantes (p.ex. confusions entre consonnes plosives)^{15,16}. Ceci suggère l'existence de pertes supralimaires qui se surajoutent aux pertes de sensibilité mesurées par l'audiogramme, et ne sont pas compensées par l'amplification fournie par une audioprothèse.

En croisant des mesures d'audibilité, de détection et de reconnaissance de phonèmes dans le bruit, Turner et coll.^{17,18} ont démontré que, contrairement au cas des personnes normoentendantes, les performances des personnes atteintes de pertes neurosensorielles dans une tâche d'identification de phonèmes ne sont pas directement prédites par leur capacité à détecter les indices acoustiques cruciaux. Autrement dit, dans le système auditif malentendant, les indices acoustiques peuvent être correctement détectés sans pour autant être utilisés efficacement pour la compréhension.

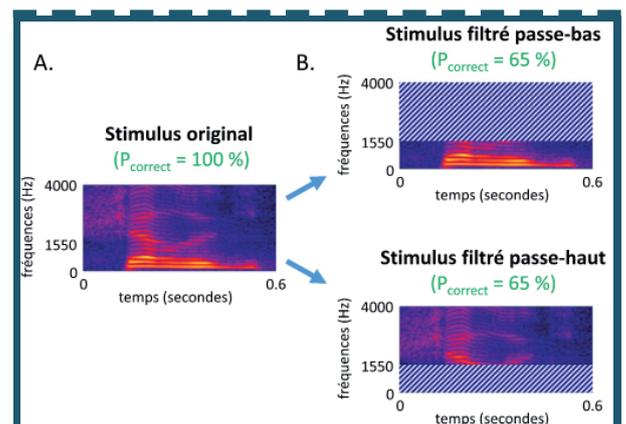


Figure 2. Illustration schématique de l'expérience de Harvey Fletcher^{10,11}. Pcorrect correspond au pourcentage de bonnes réponses dans une tâche d'identification de syllabes (c.-à-d. à l'intelligibilité du son de parole). A. Représentation temps-fréquence d'un son de parole parfaitement intelligible. B. Deux versions filtrées de ce stimulus sont obtenues en supprimant les fréquences supérieures (respectivement inférieures) à 1550 Hz. En moyenne, les sons de parole filtrés de cette manière sont toujours globalement intelligibles (Pcorrect > 0 %). Ceci démontre que des indices multiples et redondants (situés au-dessous et au-dessus de 1550 Hz) permettent de comprendre la parole, et que le système auditif est capable d'utiliser les uns ou les autres en fonction de la situation d'écoute.

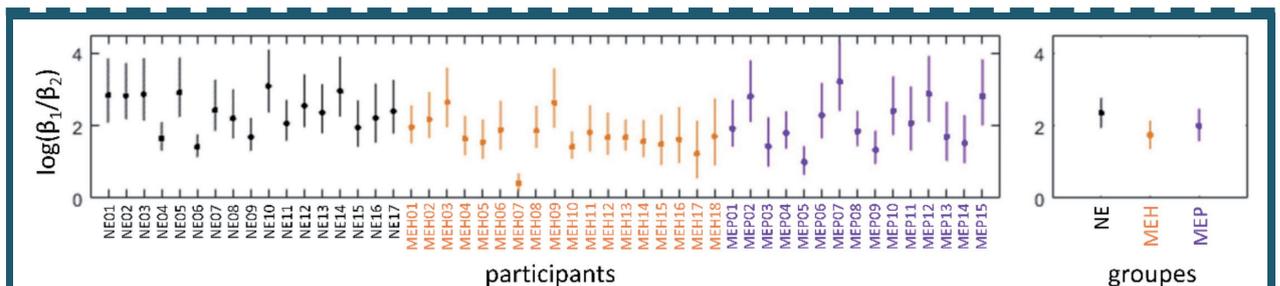


Figure 3. Log-ratios de sensibilité, mesure de l'importance accordée à l'indice haute-fréquence relativement à l'indice basse-fréquence dans la catégorisation /d/-/g/, par participant (panneau de droite) et par groupe (panneau de gauche). Les log-ratio positifs indiquent que les participants attribuent un poids perceptuel plus important à l'indice haute-fréquence. Les barres d'erreurs indiquent les intervalles de crédibilité à 95% des estimations.²⁰



Dans une étude récente ^{5, 19, 20} nous avons mesuré l'importance des indices acoustiques primaires et secondaires dans la catégorisation /da/-/ga/ (cf. Figure 1) sur des individus normoentendants (groupe NE) ou malentendants (groupe ME). Ces derniers avaient un profil de perte soit plat (pertes équivalentes dans les hautes et les basses fréquences, groupe MEP) soit en pente de ski (pertes sévères dans les hautes fréquences et modérées dans les basses fréquences, groupe MEH) et bénéficiaient tout au long de l'expérience d'une correction auditive ajustée individuellement. Les résultats indiquent que les membres du groupe NE partageaient la même stratégie d'écoute générale : ils attribuaient un poids 9 fois plus important à l'indice situé dans les basses fréquences (attaques de F2 et F3) qu'à l'indice situé dans les hautes fréquences (attaque de F1). Ce rapport de poids pour le groupe NE est représenté en noir sur la figure 3, avec une échelle logarithmique (une valeur positive indiquant un poids plus important pour l'indice haute-fréquence).

Globalement, les deux groupes ME appareillés avaient une stratégie relativement similaire à celle du groupe NE : ils accordaient plus d'importance à l'indice haute-fréquence qu'à l'indice basse-fréquence (log-ratios positifs, cf. figure 3). Néanmoins, une différence substantielle a pu être observée entre les log-ratios des groupes NE (en noir sur la figure 3) et MEH (orange) : les membres de ce dernier groupe attribuaient relativement moins de poids à l'indice haute-fréquence, situé dans la région de leurs pertes cochléaires. Ainsi, les individus souffrant de pertes haute-fréquence ne parviennent pas à utiliser l'indice haute-fréquence aussi efficacement que leurs homologues normoentendants et, ce, même lorsqu'ils sont équipés d'un système d'amplification restaurant l'audibilité de cet indice. Autrement dit, même une fois appareillées, les personnes malentendantes ne sont pas en mesure d'adapter leurs stratégies d'écoute pour rejoindre celle des personnes normoentendantes.

Ces observations convergent avec les résultats préliminaires obtenus par d'autres équipes de recherche ²¹. Là aussi, les participantes et participants souffrant de pertes hautes-fréquences appareillés ont montré des difficultés à utiliser les fréquences hautes de la parole aussi efficacement que les individus normoentendants. Ces expériences mettent en évidence l'impact des troubles supraliminaire sur le traitement des indices acoustiques et, in fine, sur la compréhension de la parole.

4

Conclusion

La multiplicité et la redondance des indices acoustiques dans le son de parole permettent au système auditif d'assurer un décodage robuste de ce dernier. Il lui est en effet possible, dans le cas où un indice serait corrompu ou absent, de se reporter sur d'autres indices. Une telle adaptation des stratégies d'écoute est particulièrement profitable lorsque le signal est corrompu par l'ajout de bruit - comme c'est le cas dans la plupart des situations d'écoute réelles - ou lorsque des déficiences altèrent son traitement (par exemple chez les individus souffrant de troubles auditifs). Dans les deux cas, l'auditrice ou l'auditeur est en mesure d'utiliser préférentiellement les bandes de fréquence les plus fiables, limitant ainsi considérablement les effets délétères de la dégradation sur l'intelligibilité.

Dans ce contexte, le rôle de l'audioprothèse est de restaurer l'audibilité des indices situés dans la région des pertes cochléaires. Cependant, rétablir une sensibilité auditive normale chez les individus malentendants ne garantit pas un rétablissement de l'intelligibilité. Ainsi, des études récentes ont montré que les personnes souffrant de pertes hautes-fréquences appareillées demeurent incapables

d'utiliser efficacement les indices amplifiés et, en conséquence, ne tirent qu'un bénéfice limité de leur prothèse.

Merci à Christian Lorenzi, Christophe Micheyl, Diane Lazard et Chloé Langlet pour leurs contributions, et à Lisa Labbouz pour la relecture. Merci également à Starkey Hearing Technologies et à l'ENS Paris qui ont financé ces travaux de recherche.

Références

1. L. Varnet, "Identification des indices acoustiques utilisés lors de la compréhension de la parole dégradée," Theses, Université Claude Bernard - Lyon I, Lyon, France, 2015.
2. A. M. Liberman, P. C. Delattre, F. S. Cooper, and L. J. Gerstman, "The role of consonant-vowel transitions in the perception of the stop and nasal consonants," *Psychol. Monogr. Gen. Appl.*, vol. 68, no. 8, pp. 1-13, 1954.
3. L. Varnet, K. Knoblauch, W. Serniclaes, F. Meunier, and M. Hoen, "A Psychophysical Imaging Method Evidencing Auditory Cue Extraction during Speech Perception: A Group Analysis of Auditory Classification Images," *PLoS ONE*, vol. 10, no. 3, p. e0118009, Mar. 2015.
4. P. Delattre, "From Acoustic Cues to Distinctive Features," *Phonetica*, vol. 18, no. 4, pp. 198-230, 1968.
5. L. Varnet, C. Micheyl, and C. Lorenzi, "Caractérisation des stratégies individuelles d'écoute de la parole au moyen d'un blob noise," presented at the Congrès Français d'Acoustique CFA '18, Le Havre, France, 2018.
6. R. V. Shannon, F.-G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, "Speech Recognition with Primarily Temporal Cues," *Science*, vol. 270, no. 5234, pp. 303-304, Oct. 1995.
7. W. Serniclaes and Y. Arrouas, "Perception des traits phonétiques dans le bruit," *Verbum*, no. 2, pp. 131-144, 1995.
8. J. B. Allen, "How do humans process and recognize speech?," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 4, pp. 567-577, 1994.
9. J. B. Allen, "Consonant recognition and the articulation index," *J. Acoust. Soc. Am.*, vol. 117, no. 4 Pt 1, pp. 2212-2223, Apr. 2005.
10. F. Li and J. B. Allen, "Multiband product rule and consonant identification," *J. Acoust. Soc. Am.*, vol. 126, no. 1, pp. 347-353, Jul. 2009.
11. H. Fletcher, "The nature of speech and its interpretation," *J. Frankl. Inst.*, vol. 193, no. 6, pp. 729-747, Jun. 1922.
12. G. Gilbert, C. Micheyl, C. Berger Vachon, and L. Collet, "Frequency-importance functions for speech in young and older listeners," in *Forum Acousticum*, Seville, 2002.
13. N. A. Lesica and B. Grothe, "Efficient Temporal Processing of Naturalistic Sounds," *PLoS ONE*, vol. 3, no. 2, p. e1655, Feb. 2008.
14. R. Plomp, "Auditory handicap of hearing impairment and the limited benefit of hearing aids," *J. Acoust. Soc. Am.*, vol. 63, no. 2, pp. 533-549, Feb. 1978.
15. A. Abavisani and J. B. Allen, "Evaluating hearing aid amplification using idiosyncratic consonant errors," *J. Acoust. Soc. Am.*, vol. 142, no. 6, p. 3736, Dec. 2017.
16. C. Scheidiger, J. B. Allen, and T. Dau, "Assessing the efficacy of hearing-aid amplification using a phoneme test," *J. Acoust. Soc. Am.*, vol. 141, no. 3, p. 1739, Mar. 2017.
17. C. W. Turner and M. P. Robb, "Audibility and recognition of stop consonants in normal and hearing-impaired subjects," *J. Acoust. Soc. Am.*, vol. 81, no. 5, pp. 1566-1573, May 1987.
18. C. W. Turner, D. A. Fabry, S. Barrett, and A. R. Horwitz, "Detection and recognition of stop consonants by normal-hearing and hearing-impaired listeners," *J. Speech Hear. Res.*, vol. 35, no. 4, pp. 942-949, Aug. 1992.
19. L. Varnet, C. Langlet, C. Lorenzi, and C. Micheyl, "Perceptual strategies for consonant discrimination in individuals with and without hearing loss," presented at the IHCON International Hearing Aid Research Conference, Tahoe City, California., 2018.
20. L. Varnet, C. Langlet, C. Lorenzi, D. S. Lazard, and C. Micheyl, "Normal-hearing and hearing-impaired individuals show distinct listening strategies for consonant discrimination in noise," *Trends Hear.*, no. X, submitted.
21. S. E. Yoho and A. K. Bosen, "Band importance functions of listeners with sensorineural hearing impairment," *J. Acoust. Soc. Am.*, vol. 143, no. 3, pp. 1943-1943, Mar. 2018.