HAL
open science

Vers un appareillage auditif personnalisé qui préserve les émotions

Jean-Julien Aucouturier

To cite this version:

Jean-Julien Aucouturier. Vers un appareillage auditif personnalisé qui préserve les émotions. Actes du Congrés du Syndicat National des Audioprothésistes Français (UNSAF), 2019. hal-02371704

HAL Id: hal-02371704

https://hal.science/hal-02371704v1

Submitted on 20 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Vers un appareillage auditif personnalisé qui préserve les émotions.

Jean-Julien Aucouturier

Les résultats scientifiques de la psychoacoustique et des sciences de l'audition décrivent principalement les mécanismes bas-niveau permettant à un auditeur d'extraire les dimensions fondamentales d'un son, comme sa hauteur ou son intensité, ainsi que les caractéristiques permettant de comprendre la parole. La préservation de ces caractéristiques sonores a été traditionnellement, et à juste titre, la cible privilégiée de l'appareillage auditif.

Cependant, notre utilisation quotidienne du son et de la parole ne se limite pas à sa seule détection et compréhension, mais englobe aussi toute une série de mécanismes haut-niveau permettant, par exemple, de percevoir une intonation bienveillante ou moqueuse chez notre interlocuteur, de reconnaître le timbre de voix particulier d'un proche, ou d'apprécier l'émotion joyeuse ou triste d'une mélodie musicale. Alors que ces mécanismes de perception haut-niveau sont aujourd'hui relativement bien étudiés dans le domaine visuel (Jack & Schyns, 2015), nous manquons encore de recul sur les caractéristiques sonores à préserver pour faciliter ce genre de traitements chez la personne appareillée auditivement.

Dans une série de travaux récents au laboratoire des Sciences et Technologies de la Musique et du Son (IRCAM/CNRS/Sorbonne Université), nous avons développé une nouvelle méthodologie expérimentale, dite de « corrélation inverse », qui permet de mettre en évidence les représentations mentales sur lesquelles se base une personne pour juger de l'intonation linguistique ou émotionnelle d'une voix. Nous avons tout d'abord créé un nouveau logiciel de traitement du signal, appelé CLEESE (Burred et al. 2019), qui permet de générer un grand nombre de prononciations aléatoires d'un même mot : on enregistre par exemple le mot « vraiment » (qui peut être prononcé de façon affirmative – vraiment ! - ou interrogative – vraiment ?), et le logiciel peut transformer l'enregistrement de façon à varier la façon dont il est prononcé (sa *prosodie*) : faire monter la hauteur au début du mot, à la fin, ralentir la deuxième syllabe, etc.

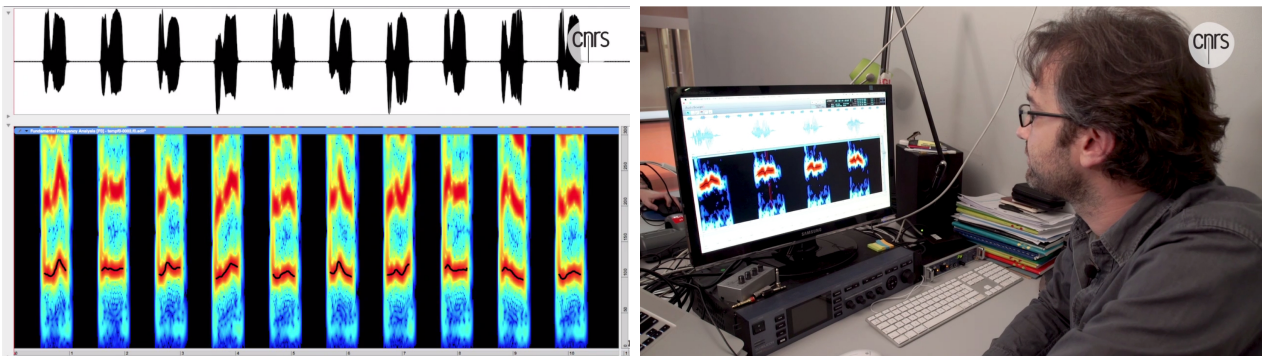
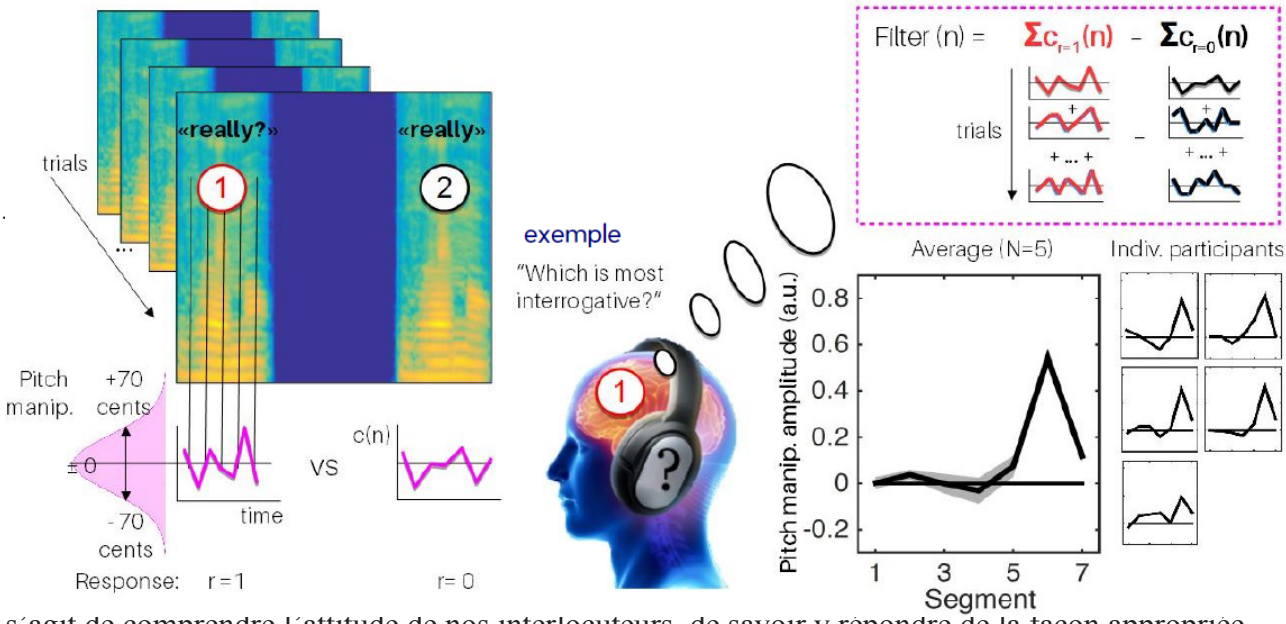


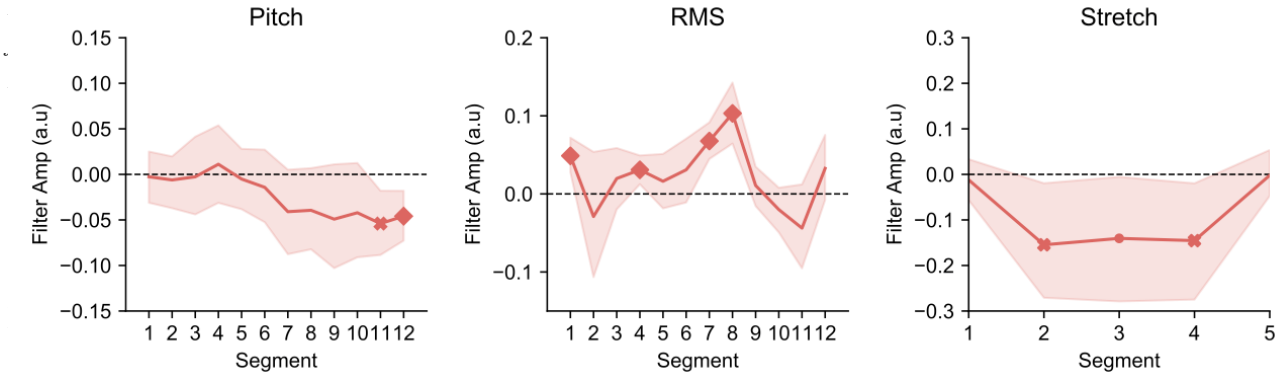
Fig.1. Le logiciel CLEESE permet de générer automatiquement des milliers de prononciations aléatoires d'un même mot. Pour un démonstration, voir par exemple ce film réalisé en 2018 par le CNRS : https://lejournal.cnrs.fr/videos/dis-moi-bonjour-et-je-te-dirai-qui-tu-es

On présente ensuite à des participants expérimentaux des paires de sons composées de 2 prononciations aléatoires générées par l'algorithme et on leur demande de choisir, pour chaque paire, lequel des 2 sons semble être prononcé de la façon la plus interrogative. Comme chacune des

prononciations proposées est purement aléatoire, aucune des options n'est réellement interrogative, mais celle-ci aura une intonation légèrement traînante à la fin qui la fera préférer à celle-là, dont la hauteur descend de façon plus assertive ; une autre aura un accent trop dynamique au début du mot, auquel on préférera une variation plus plate, etc. Les participants font de leur mieux, pendant des centaines de paires. A la fin de l'expérience, on regroupe toutes les exemples choisis comme interrogatifs par un participant, on en calcule la prononciation moyenne, et on soustrait la prononciation moyenne de toutes les exemples rejetées. On voit alors apparaître la prononciation idéale que le participant a cherché à juger dans tous les exemples : un profil de hauteur qui monte rapidement sur la deuxième syllabe du mot (Fig.2). Cette façon de prononcer le mot vraiment n'a jamais été proposée au participant (et pour cause, l'algorithme n'en ayant pas connaissance), mais correspond en fait à une estimation de la « représentation mentale » à laquelle le participant a comparé tous les autres exemples pour juger s'ils sont interrogatifs ou non.

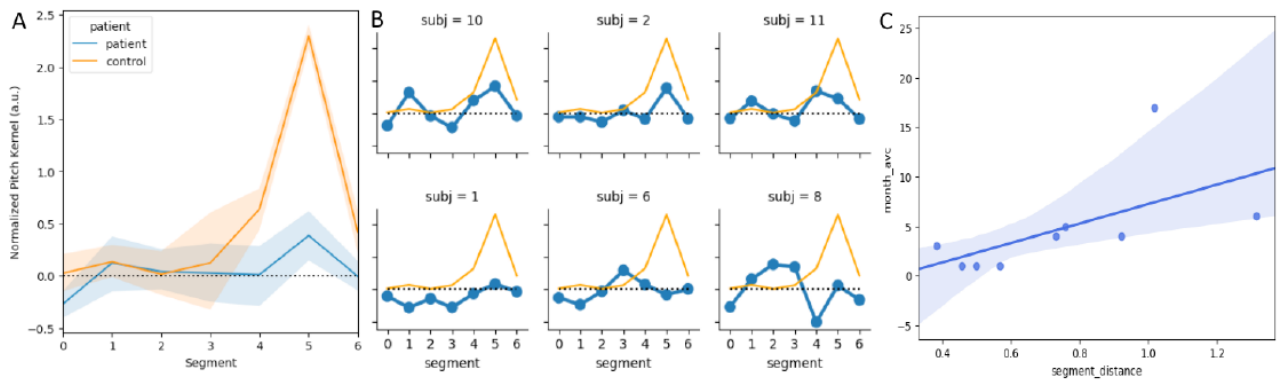


s'agit de comprendre l'attitude de nos interlocuteurs, de savoir y répondre de la façon appropriée, de reconnaître l'ironie, etc. Si le système auditif du participant n'encode pas bien ces caractéristiques, et en particulier leur caractère dynamique, certains de ces indices seront perdus, et les jugements correspondants rendus plus difficiles, ou impossibles.



pour les contrôles, à une hausse de la hauteur à la fin du mot, celle-ci (Fig. 4A – bleu) est beaucoup moins marquée que pour les contrôles (Fig. 4A - orange), qui témoigne d'un manque de sensibilité (il faut que l'intonation monte beaucoup plus pour que les patients détectent l'intention

interrogative). De plus, il existe une grande variété de profils au sein du groupe de patients, certains (comme P1, voir Fig. 4B) ayant des profils plats (cad qu'ils n'utilisent aucune information prosodique pour faire la tâche) et d'autres (comme P8, voir fig.4B) des profils carrément inversés (cad qu'ils jugent le mot comme plus interrogatif si son intonation *descend* à la fin du mot).



rééducation, à orienter vers un type de prise en charge ou un autre, voire même inspirer de nouvelles méthodes de traitement qui vont cibler sélectivement les déficits (apprendre à écouter la fin d'un mot, apprendre à différencier les intonations montantes et descendantes) ou de nouvelles modalités d'appareillage qui pourraient amplifier certains de ces indices pour pallier leur encodage auditif déficient.

Ces méthodologies modernes nous placent à l'orée d'une sorte d' «orthophonie computationnelle », qui permettra à terme de concevoir des appareillages personnalisés et facilitant le traitement d'un plus grand nombre de dimensions vocales et sonores que celles préservées aujourd'hui. Ces perspectives cliniques font l'objet d'une partie de nos actuels travaux de recherche.

Jack, R. E., & Schyns, P. G. (2015). The human face as a dynamic tool for social communication. Current Biology, 25(14), R621-R634

Burred, J. J., Ponsot, E., Goupil, L., Liuni, M., & Aucouturier, J. J. (2019). CLEESE: An open-source audio-transformation toolbox for data-driven experiments in speech and music cognition. PloS one, 14(4), e0205943 https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0205943

Goupil, L. & Aucouturier, JJ. (2019) A shared prosodic code automatically signals both confidence and honesty to human listeners. En préparation.

Jeulin, M. (2018) Etude de la prosodie réceptive interrogative après un accident vasculaire cérébral : utilisation du paradigme de reverse-corrélation. Mémoire Master de Sciences et Technologies – Santé, Université Pierre et Marie Curie.

Ponsot E., Burred JJ., Belin P., Aucouturier JJ., (2018) Cracking the social code of speech prosody using reverse correlation. PNAS 201716090. http://www.pnas.org/content/early/2018/03/20/1716090115

Joanette, Y., Ska, B., & Côté, H. (2004). Protocole Montréal d'évaluation de la communication (Protocole MEC). Isbergues, France: Ortho Édition.