



## Analyse de marqueurs audiovisuels de l'intelligibilité perçue

**Cadre :** Nous ne sommes pas passifs lorsque nous écoutons un interlocuteur, mais produisons de nombreux signes verbaux et non verbaux, en particulier des mouvements de tête et des mimiques faciales. De nombreuses études ont déjà été menées sur les indices audiovisuels produits par un interlocuteur pour marquer son accord avec ce qui est dit, son degré d'appréciation du partenaire, pour gérer les tours de parole, etc. En revanche, peu d'information a été apportée sur les marqueurs de compréhension et d'évaluation du niveau d'intelligibilité de la parole perçue.

**Question :** Quels sont précisément les indices produits pour marquer le degré de compréhension, pour demander une répétition ? Ces indices sont-ils très variables d'un individu à l'autre ou extrêmement codifiés et reproductibles ? Comment ces indices sont-ils perçus par le locuteur lorsqu'il parle et comment influencent-ils sa propre façon de parler, notamment le niveau de clarté de sa parole ?

**Projet :** Une expérience d'interaction face à face sera conduite avec une vingtaine de participants : un jeu interactif impliquant l'expérimentateur et le participant sera imaginé de façon à susciter entre eux des incompréhensions contrôlées, mais relativement naturelles. Nous enregistrerons les signaux audio, les mouvements de la tête et du visage (yeux, sourcils, lèvres, ...) (Vidéos, Système Optotrack) de façon à caractériser les indices audibles et visibles produits par les participants lors de ces incompréhensions. Les marqueurs audiovisuels seront dans un premier temps étiquetés manuellement sur une petite partie des données puis, suivant une approche de classification semi-supervisée, un modèle audiovisuel de ces marqueurs sera ensuite généré afin d'étiqueter le reste des données. Différents descripteurs de parole (mesures spectrales, prosodiques, qualité vocale) et du visage (intensité d'unités d'action faciale, position angulaire de la tête) seront extraits en utilisant des outils existants (openSMILE, openFACE, ...). Des analyses statistiques seront enfin conduites pour interpréter les stratégies réalisées.

**Perspectives :** Ces résultats serviront également de base à une seconde expérience pour explorer comment le locuteur perçoit les mimiques faciales de son interlocuteur lorsqu'il parle et comment il adapte sa façon de parler, en particulier son niveau d'intelligibilité, en réaction aux signes d'(in)compréhension que lui envoie l'interlocuteur. Cette deuxième étape pourra s'effectuer au cours d'une thèse, dans le prolongement du stage.

**Compétences requises :** Au minimum, des bases de programmation et de traitement du signal, notions en production de la parole et en psychologie comportementale. Une connaissance préalable des logiciels Matlab, R et Praat sera appréciée, ainsi que de bonnes capacités rédactionnelles.

**Compétences développées pendant le stage :** Connaissances théoriques sur l'interaction sociale et la multimodalité de la parole ; Familiarisation avec diverses méthodes d'investigation de capture de mouvement ; Gestion de A à Z d'une expérience sur l'humain ; Compétences en traitement de signaux acoustiques et cinématiques, en programmation Matlab et en analyses statistiques.

**Encadrement :** Le projet se déroulera sur une période de 4-5 mois, encadré par Maëva Garnier et Fabien Ringeval.

**Indemnisation :** Indemnités mensuelles de 400 € sur 6 mois.