

# Project

## Anticipatory cues in intonation processing: effects of cross-linguistic and speaker-specific differences

Caterina Petrone<sup>(1)</sup>, Alessandra Lonobile<sup>(2)</sup>, Kiwako Ito<sup>(3)</sup>

(1) LPL, (2) University of Naples Federico II (Italy), (3) The Ohio State University (USA)

### Abstract

-

### Publications

Petrone C, Lo Nobile A, Zielinski C, Ito K. Effects of the prenuclear contour in the processing speaker commitment in French, Proceedings of the 8th international conference on Speech Prosody, 2016, 4 pages

### Fiche-résumé contribution CReX



## Emoz

**Effet de la prosodie sur la reconnaissance des émotions**

**Investigateurs :** C. Petrone (LPL), A. Lonobile (USN, Italy), K. Ito (OSU, US)

**Durée :** 5 semaines

**Contribution :** aide à la mise au point du design expérimental (préparation des stimuli, configuration de l'expérience), aide à la passation (6 sujets), traitement des données, mise en forme des figures,

**Objectif :** détecter les indices prosodiques émotionnels précoces à l'aide d'une expérience en oculométrie (système Tobii)

■ **Paradigme** – L'expérience consiste à enregistrer les mouvements oculaires pendant une tâche de reconnaissance de l'expression du visage supposé du locuteur, en fonction de l'énoncé entendu dit par ce dernier. Le participant voit une image apparaître à l'écran, et entend une phrase en parallèle (ex. « Hugo a vendu la guitare »). L'image montre deux photographies du visage du locuteur, présentant une expression faciale différente. Le participant doit cliquer à l'aide de la souris sur la photographie qui lui semble correspondre avec l'état émotionnel du locuteur au moment où il prononce la phrase. 126 images sont ainsi jouées. Les paires d'expressions correspondent avec les énoncés de type affirmation/doute, colère/peur ou tristesse/joie. Une première contribution à ce projet a consisté à homogénéiser les photographies des visages à l'aide de Gimp (cadrage / couleurs / dimensions) et à les assembler par paire (exemple fig. 1). Les stimuli ont ensuite été assemblés en vidéo avec un logiciel de montage pour lier la piste audio et l'image.

■ **Passation** – Aide à la passation de l'expérience pour 6 sujets sur les 22 au total.

■ **Prétraitement** – Des programmes ont été mis au point pour lire l'ensemble des données brutes et les trier par condition expérimentale. Des paramètres supplémentaires ont été extraits de tableaux de données et intégrés aux données d'eye-tracking : (1) les données avec les temps correspondants à la fin de la prononciation des noms, verbes et complément d'objet des 126 phrases ; (2) les données issues d'une tâche de notation sur une échelle de 1 à 5 pour juger de l'émotion véhiculée par l'énoncé de chaque phrase, réalisée auprès de 15 sujets indépendants de l'expérience d'eye-tracking ; (3) des données de valeurs des fréquences fondamentales et pentes des différents contours fréquentiels des phrases. Les temps critiques par phrase (nom, verbe, objet) ont été repérés manuellement à l'aide de PhonEdit après avoir extrait les pistes audio des stimuli assemblés en vidéo à l'aide du package FFmpeg (pour prévenir tout décalage pouvant apparaître lors du montage des stimuli en vidéo). A l'issue du prétraitement, les données de fixation sont rassemblées par zone d'intérêt (visage en partie droite ou gauche de l'écran – fig. 2), dans des matrices [stimuli x sujet x échantillon temporel]. Ces matrices ont été déclinées en plusieurs versions selon le temps critique utilisé pour le réaligement temporel par stimulus ( $t_{\text{beg}}$  : début de la phrase,  $t_{\text{noun}}$  : fin du nom,  $t_{\text{verb}}$  : fin du verbe,  $t_{\text{click}}$  : réponse par clic souris). Des versions ont également été définies en fonction des scores des stimuli obtenus lors de l'étude de normalisation (ex. fixations pour les stimuli très peu ambigus pour la condition « doute » vs fixation pour les stimuli ayant obtenus un mauvais score dans la tâche de notation pour cette même condition).

■ **Analyse** – L'analyse s'est basée sur le calcul des courbes de pourcentages de fixation au cours du temps par zone d'intérêt (Area Of Interest), moyennées à travers les stimuli et les sujets. Pour se faire, les données de fixation ont donc été réalignées en fonction des différents temps critiques de

prononciation par phrase (fin du nom, fin du verbe, fin de la phrase). Ce réalignement a permis d'effectuer les calculs de moyenne de fixation entre les différents stimuli, les durées de prononciation n'étant pas homogènes entre les différentes phrases. Des courbes « composites » des moyennes de proportions de fixation par AOI ont ainsi pu être créées, les temps critiques moyens définissant les limites pour assembler chaque portion de courbe (temps moyens  $t_{\text{beg}}$  à  $t_{\text{noun}}$ ,  $t_{\text{noun}}$  à  $t_{\text{verb}}$  et  $t_{\text{verb}}$  à  $t_{\text{click}}$ ). Les figures des courbes composites ont été mises en forme (exemple fig. 3), l'ensemble des figures représentant les différentes conditions (doute/affirmation, colère/peur, tristesse/joie), pour les différents types de stimuli (peu ou très ambigus, locuteur homme ou femme), ont été générées automatiquement à l'aide du programme créé.

■ **Diffusion** – La méthodologie a été rédigée et intégrée dans l'acte de conférence :

Petrone C, Lo Nobile A, Zielinski C, Ito K. Effects of the prenuclear contour in the processing speaker commitment in French, Proceedings of the 8th international conference on Speech Prosody, 2016, 4 pages

Les programmes ont été déposés sur GitHub pour être partagés avec la communauté. Bien que le traitement soit très spécifique à l'étude, certaines fonctions et sous-fonctions pourraient aider au traitement de données expérimentales similaires acquises avec Tobii (stimulation visuelle avec réalignement des fixations par rapport à la stimulation auditive).

[https://github.com/blri/eye\\_tracking\\_tobii/tobii\\_emoz](https://github.com/blri/eye_tracking_tobii/tobii_emoz)

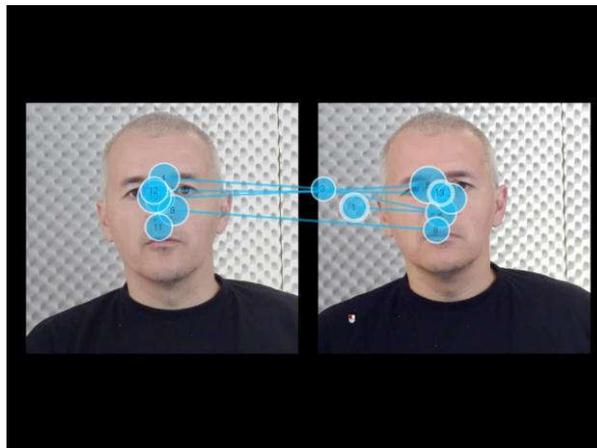


Figure 1: exemple de stimulus visuel avec les fixations du participant superposées sur l'image (ronds bleus)



Figure 2 : définition des régions d'intérêts (AOI) pour l'analyse des données d'eye-tracking

Mean fixation proportions across stimuli and subjects - assertion vs incredulity

*Including only well differentiated stimuli at the rating task (scores < 1.5 and > 4.5) and subjects that clicked on the consistent AOI after the end of the sentence*

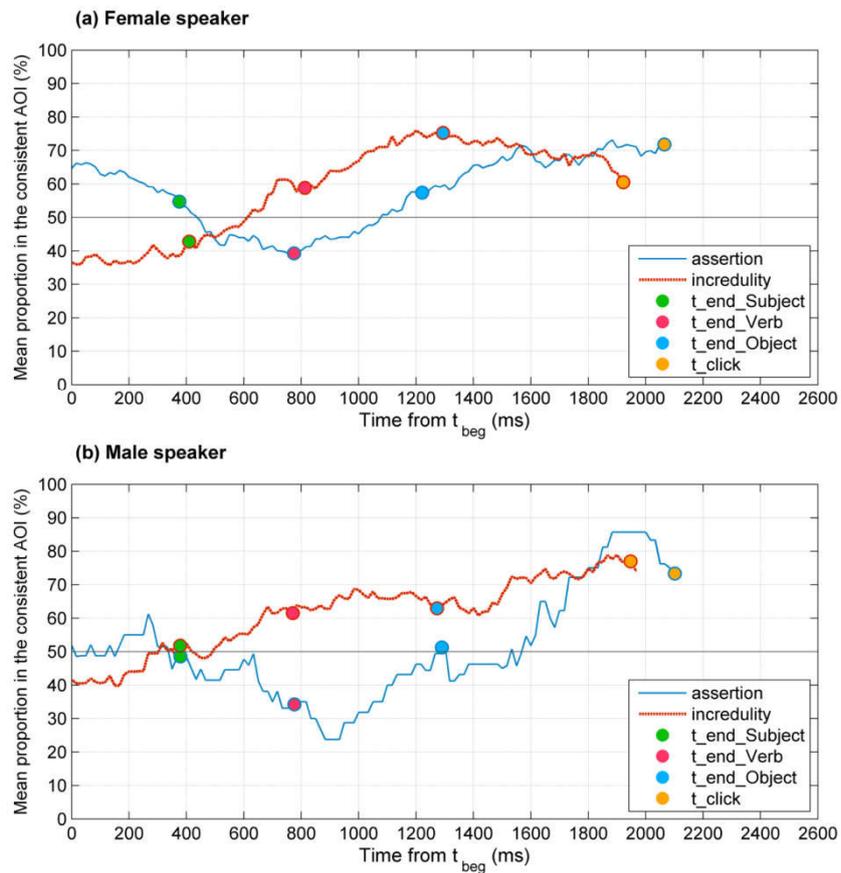


Figure 3 : exemple de figure de résultats montrant les pourcentages de fixation moyennés sur l'ensemble des sujets et stimuli, et réalignés par temps critiques moyens. Un effet d'anticipation pour la condition liée à l'incrédulité se lit sur les courbes, avec augmentation du pourcentage de fixation dans la zone liée à la condition d'incrédulité dès la fin de la prononciation du nom. Pour la condition liée à la certitude (ton neutre), cette augmentation des fixations dans la zone d'intérêt correspondante (courbe bleue), n'est pas aussi précoce et ne débute qu'après la prononciation du verbe.