



Apport du geste dans l'acquisition de la prononciation en L2 : quand la réalité ne correspond pas aux attentes

Charlotte Alazard-Guiu¹, Leonardo Contreras Roa², Marie Philippart De Foy^{1,2}, Lionel Fontan³, Nadia Yassine-Diab⁴, Julien Pinquier², Isabelle Ferrané²

(1) Laboratoire de NeuroPsychoLinguistique, EA4156, Toulouse, France

(2) IRIT, Université de Toulouse, CNRS, Toulouse INP, UT3, Toulouse, France

(3) Archean Technologie, Montauban France

(4) Laboratoire Lairdil, Toulouse France

charlotte.alazard@univ-tlse2.fr, l.contreras.roa@u-picardie.fr, marie.philippart-de-foy@univ-tlse2.fr, lfontan@archean.tech, nadia.yassine-diab@univ-tlse3.fr, julien.pinquier@irit.fr, isabelle.ferrane@irit.fr

RÉSUMÉ

Nous présentons une réflexion méthodologique à la suite d'une expérimentation réalisée dans le cadre d'un projet interdisciplinaire où nous avons voulu tester l'impact de la gestualité corrective sur l'acquisition du segmental et du suprasegmental. Pour cela, nous avons mis en place un protocole pré-test/post-test dans trois conditions expérimentales : un groupe test qui a reçu des gestes correctifs appropriés aux erreurs réalisées, un groupe test qui a reçu des gestes correctifs inadéquats et un groupe contrôle qui n'a pas reçu de correction multimodale. Nous avons enregistré 45 étudiants japonophones de FLE, équitablement répartis entre les trois groupes. Les analyses des 7200 productions orales (160*45) n'ont pas permis de différencier les groupes ni de valider nos hypothèses de recherche. En revanche, les limites observées soulèvent non seulement des réflexions méthodologiques liées au protocole expérimental et à sa passation, mais elles nous amènent également vers une nouvelle proposition de design expérimental.

ABSTRACT

Contribution of gesture in the acquisition of L2 pronunciation: when reality does not match expectations

We present a methodological reflection following an experiment carried out in an interdisciplinary project. Within the framework of this project, we wanted to experimentally test the impact of corrective gestures on the acquisition of segmental and suprasegmental pronunciation skills. A pretest/posttest/control group design was set up with three experimental conditions: a test group that received corrective gestures adapted to the anticipated errors, a test group that received corrective gestures inadéquats to the anticipated error, and a control group that received audio-only correction. We recorded 45 Japanese-speaking students of French as a Foreign Language, equally distributed among the three groups. The acoustic and perceptual analyses of the 7200 oral productions (160 *45) did not allow us to differentiate between the three groups. However, the observed limitations raised methodological reflections and also lead us to a new proposal of experimental design.

MOTS-CLÉS : gestes correctifs, acquisition de la prononciation en L2, didactique expérimentale

KEYWORDS : corrective gestures, acquisition of L2 pronunciation, experimental didactics

1 Introduction

Si les dernières décennies ont vu naître un intérêt croissant pour l'apport de la multimodalité dans l'acquisition d'une langue seconde (L2) (Gullberg 2006 ; Tellier 2008), peu de recherches se sont intéressées à la question de l'utilisation du geste dans l'apprentissage de la phonétique en L2. Certaines études basées sur des observations de classe ont montré que les apprenants semblaient sensibles à cette modalité. En effet, après une phase d'entraînement les apprenants avaient tendance à naturellement imiter la gestuelle de l'enseignant et paraissaient mieux maîtriser la dimension prosodique de la L2 étudiée (McCafferty 2006 ; Smotrova 2015). Cependant, malgré ces observations prometteuses, l'absence de méthodologie bien définie (présentation de la méthode de correction phonétique utilisée par les enseignants, justification des gestes utilisés, etc.), n'a pas permis une évaluation et une diffusion plus large de ces approches, basée sur les pratiques de classe. Plus récemment, d'autres études ont ouvert la voie à une approche plus contrôlée du rôle des gestes dans l'acquisition de la prononciation en L2 pour deux types de gestes: les gestes rythmiques (comme les *beating gestures*, Gluhareva & Prieto 2017 ou les *clapping gestures*, Zhang *et al.*, 2020) et les gestes iconiques (Connell *et al.*, 2013 ; Morett & Chang, 2015 ; Yuan *et al.*, 2019). En outre, dans un récent article, Yuan *et al.* (2019) ont montré que l'apprentissage de contrastes prosodiques (variation de la fréquence fondamentale, F0 ou *pitch*) était significativement facilité par l'intégration des gestes co-verbaux iconiques dans la phase d'entraînement des apprenants. Même après un temps relativement court, les apprenants sinophones débutants étaient capables de reconnaître des intonations différentes dans la langue cible, ici l'espagnol. Nos propres observations et les premiers résultats d'études récentes (Alazard, 2013 ; Alazard *et al.*, 2018) vont dans le même sens. Nous pensons que la visualisation des gestes, et en particulier celles de gestes co-verbaux métaphoriques ou iconiques et de battements (McNeill, 1992) pourraient constituer un outil facilitant l'acquisition du système phonético-phonologique en L2, et par là le développement des compétences de prononciation.

2 Protocole initial

Le projet INGPRO a pour objectif l'étude de la relation entre les gestes co-verbaux correctifs et l'acquisition du système phonético-phonologique en langue étrangère, dans le cadre d'un enseignement de la prononciation assisté par ordinateur où l'apprenant doit fonctionner en autonomie ou autonomie guidée. Le projet reposait sur 5 points essentiels : (1) le choix d'un ensemble de gestes à étudier, (2) le choix d'un ensemble de stimuli et leur combinaison ou non avec des gestes correctifs de façon à obtenir des données contrastives, (3) la définition d'un protocole expérimental visant à organiser les mises en situation des apprenants de L2, (4) la collecte des productions des apprenants soumis à notre protocole au travers de la réalisation d'une application dédiée et (5) l'analyse de ces productions pour évaluer et discuter de l'apport du geste en situation de correction phonétique.

2.1 Élaboration du protocole

A partir de la littérature et de l'analyse de corpus vidéos de correction phonétique¹, nous avons sélectionné six gestes co-verbaux différents, parmi les gestes les plus utilisés par les enseignants. Certains présentaient un usage iconique (gestes illustratifs d'un concept concret, cf. Tellier *et al.*,

¹ Comme peu de méthodes de correction phonétique ont recours à la gestualité co-verbale, nous nous sommes basés sur les principes correctifs de la méthode verbo-tonale (MVT) d'intégration phonétique.

2011) et d'autres un usage métaphorique (gestes illustratifs d'un concept abstrait, *ibid*). Dans la première catégorie, nous trouvons, par exemple, les gestes utilisés pour la correction du suprasegmental : les gestes mimant les mouvements intonatifs (intonation montante ou montante-descendante) ou un geste d'allongement pour illustrer l'accentuation primaire en français. Dans la seconde catégorie, nous trouvons un geste vers le haut indiquant la tension musculaire, utilisé en cas de dévoisement par exemple, un geste vers le bas indiquant le relâchement et un geste d'ouverture pour faciliter la nasalisation (les gestes sont visibles en suivant le lien du projet²).

Les gestes retenus ont été qualifiés de congruents lorsqu'ils sont appropriés à la correction de l'erreur attendue et d'incongruents dans le cas inverse (pour les gestes incongruents, il s'agissait alors des mêmes gestes appliqués à d'autres erreurs).

Afin de contrôler la variabilité de la réalisation des gestes par les enseignants, nous avons choisi de passer par le biais du numérique en présentant aux apprenants un avatar et des animations 3D *via* une plateforme existante d'apprentissage des langues assisté par ordinateur (Fontan & Le Coz, 2017). Dans la version actuelle de la plateforme, l'apprenant peut pratiquer sa prononciation en autonomie car le programme propose un retour correctif (*feedforward*³) qui consiste en la répétition du stimulus audio original en mettant l'accent sur le ou les sons mal prononcés par l'apprenant. L'objectif de cette première étude était de déterminer si l'ajout d'un *feedforward* gestuel est associé à une amélioration significative des compétences de production des apprenants. En nous basant sur les méthodologies des études précédentes (Gluhareva & Prieto, 2017 ; Yuan *et al.*, 2019) et sur les contraintes statistiques que cela engendre (nombre de sujets, nombre de stimuli etc.), nous avons élaboré le protocole suivant basé sur 3 groupes de sujets :

- 1) Un **premier groupe test** qui recevra des *feedforwards* audio et gestuels adaptés à l'erreur supposée (groupe 1).
- 2) Un **second groupe test** qui recevra des *feedforwards* audio et gestuels non adaptés à l'erreur supposée (groupe 2) ;
- 3) Un **groupe contrôle** qui recevra des *feedforwards* audio uniquement (audio avec emphase sans geste correctif) (groupe 3).

La phase de test proprement dite proposait 40 stimuli (25 mots et 15 phrases simples) répétés deux fois et ce quel que soit le groupe d'apprenants concerné. Un jeu de 120 vidéos (avatar mimant ou non les gestes) a été généré, 3 par stimulus. La figure 1 illustre le design expérimental pour le test d'un stimulus : « *je m'appelle Marie* ».

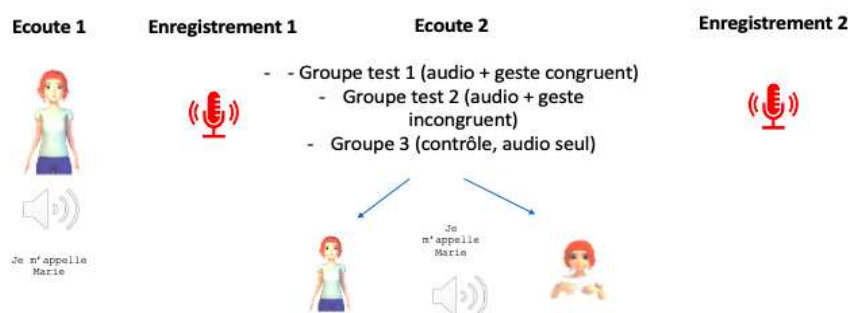


FIGURE 1: Design expérimental

² <https://www.irit.fr/SAMOVA/site/projects/current/ingpro/>

³ Contrairement au *feedback* qui est un retour sur une erreur produite, le *feedforward* est un nouvel énoncé qui va intégrer le son mal prononcé dans un contexte plus facilitateur.

En résumé, chaque sujet a écouté et répété 80 stimuli (40 x 2). Il y avait en tout quatre productions pour un même mot : répétition initiale (R1) puis répétition après correction (R2) pour chaque stimulus. Celui-ci, présenté deux fois à chaque apprenant au cours de sa session, donnait lieu à deux répétitions supplémentaires (R3) et (R4). La durée totale du test était de 30 minutes par sujet. Nous avons formulé l'hypothèse selon laquelle nous observerons des performances supérieures pour les sujets du groupe 1 après un *feedforward* correctif congruent par rapport aux étudiants du groupe contrôle. Nous nous attendions également à observer des performances inférieures pour les sujets du groupe 2 après un *feedforward* correctif incongruent.

2.1.1 Choix des stimuli

En fonction des erreurs fréquemment rencontrées chez les apprenants japonais de FLE (Kamiyama *et al.*, 2016), un ensemble de 7 erreurs cibles a été défini. Les tableaux 1 et 2 décrivent les segments visés et les erreurs attendues pour le suprasegmental et le segmental respectivement. Nous avons associé à chaque erreur potentielle, le geste correctif adapté (i.e. groupe congruent) ainsi que plusieurs choix de gestes inadapés (i.e. groupe incongruent).

Cible prosodique	Erreur attendue	Geste correctif adapté	Geste correctif inadapés (exemple)
Contour montant	C. descendant	geste montant (iconique)	geste d'allongement
Acc. de durée	Acc. d'intensité	geste d'allongement (iconique)	geste de relâchement

TABLE 1 : Illustration DES erreurs attendues pour chaque cible prosodique

Nous avons utilisé la base de données « LexPro⁴ » pour sélectionner les mots utilisés dans les stimuli que ce soit pour la recherche d'erreur au niveau segmental ou au niveau suprasegmental. Dans le premier cas, nous avons sélectionné des mots adaptés au niveau débutant, en contraignant le phonème cible, sa position dans le mot et la fréquence du mot. Nous avons ensuite intégré 15 mots dans de courtes phrases (4 à 5 syllabes) avec la construction suivante : *C'est + déterminant + nom* (“C'est la ville”, “c'est la fête”)

Segment visé	Erreur attendue	Geste correctif adapté	Geste correctif inadapés (exemple)
/b/	[v] / [β]	geste de tension (métaphorique)	geste d'allongement
/ʒ/	[dʒ]	geste de relâchement (métaphorique)	geste d'ouverture
/y/	[u]	geste montant (métaphorique)	geste d'allongement
/ã/	Voyelle orale	geste d'ouverture (iconique)	geste de relâchement
/ʁ/	suppression	geste d'allongement (iconique)	geste montant

TABLE 2 : Erreurs attendues pour chaque segment cible.

⁴ LexPro est une base de données créée par Archean (Fontan *et al.* 2018)

Pour la partie segmentale, nous avons élaboré une liste de stimuli à partir des erreurs identifiées (cf. table 2) et en appliquant les contraintes suivantes : mots de 1 à 3 syllabes et pas d'autre erreur dans la syllabe cible. Nous avons éliminé tous les mots internationaux (comme *taxi*) ou transparents (c'est-à-dire des mots qui se ressemblent beaucoup à travers les langues, *menu* ou *musique* par exemple).

2.1.2 Recrutement des sujets

Le protocole prévoyait de recruter des apprenants de niveau débutant en déplacement à Toulouse dans le cadre d'un échange entre universités. En raison de la pandémie, les 45 sujets (dont 38 femmes) ont été recrutés parmi les étudiants de deuxième et troisième années du cursus de français « Langue et Littérature française » dans l'université Tokyo University of Foreign Studies (Tokyo, Japon), selon les critères d'inclusion suivants : japonophone natif, pas de problème d'audition auto-déclaré ; pas de problème visuel non corrigé ; pas de trouble de la parole auto-déclaré. Une indemnisation a été attribuée aux participants.

2.2 Analyses des résultats

Suite aux passations, qui ont eu lieu au Japon et non en France comme cela été initialement prévu, 7200 énoncés ont été recueillis. Nous avons proposé une analyse acoustique et une analyse perceptive des données. Nous avons d'abord annoté ces productions de manière semi-automatique sous Webmaus (Kisler *et al.*, 2017) et Praat (version 6.1, Boersma & Weenink), puis corrigé manuellement et traité l'ensemble des données. Suite à des test de Kruskal-Wallis (alternative non-paramétrique au test ANOVA à un facteur), nous n'avons observé aucun effet statistique des deux facteurs d'intérêt (effet de groupe et de répétition) pour aucun des sons ou segments ciblés probablement en partie à cause du faible nombre d'occurrences des erreurs que nous attendions. En effet, alors que nous nous étions focalisés sur une réalisation possible pour chaque erreur attendue, nous avons observé énormément de variabilité dans les productions des apprenants, comme illustré sur la Table 3.

Segment visé	Erreur attendue	Erreurs observées
/b/	Fricatisation du /b/ réalisé [β] ou [v]	Erreurs de dévoisement : /b/ réalisé [p]
/ʒ/	Affrication en [dʒ]	Absence d'erreur ou dévoisement: /ʒ/ réalisé [ʒ~ʃ]
/y/	Posteriorisation et désarrondissement de la voyelle réalisée [u]	Diph/triphthongaison de la voyelle produite [iy] / [iu] ou [iyi] / [iui] OU Assourdissement de la voyelle pour les items <i>étudiant</i> et <i>assurance</i> dû à la consonne précédente /t/ - /s/
/ã/	Dénasalisation de la voyelle réalisée [a]	Épenthèse consonantique en coda [m], [n], [ŋ], [N]
/ʁ/	suppression	Fricative glottale sourde [h]

TABLE 3 : Erreurs attendues vs. erreurs observées

En raison de la variabilité des réponses, le groupe 1 n'a pas toujours reçu les *feedforwards* gestuels adaptés à l'erreur initialement ciblée. Selon les productions, ils ont même parfois reçu des corrections gestuelles incongruentes. La comparaison entre les trois groupes n'est de fait plus pertinente.

En parallèle des analyses acoustiques, plusieurs analyses automatiques ont été réalisées via la plateforme Paty⁵. Cette plateforme propose des jeux de modèles acoustiques différents permettant de réaliser un décodage phonétique ou un décodage au niveau mots. Les modèles utilisés ont été entraînés sur 340 heures de données issues de différents corpus audio français, complétés par une phase d'augmentation des données (Heba, 2021). Les analyses réalisées ont montré un taux de reconnaissance très faible sur l'ensemble des mots indiquant que même à un niveau intermédiaire les apprenants font encore beaucoup d'erreurs de prononciation.

3 Discussion

Autrement dit, le principal problème n'était pas l'absence d'erreurs mais la grande variabilité observée dans les réalisations ce qui peut être dû à deux causes que nous avons identifiées :

- Tout d'abord, en raison de contraintes phonotactiques, certaines erreurs attendues apparaissaient moins dans certaines positions : par exemple, le /b/ en position initiale est beaucoup plus sujet à des erreurs de dévoisement qu'à des erreurs de fricatisation comme cela était attendu, sauf pour le mot "boîte" pour lequel nous avons davantage observé l'erreur attendue. Pour l'erreur ciblée sur le phonème /y/, certains items choisis (comme *étudiant* ou *assurance*) sont problématiques car la consonne précédente (/t/ ou /s/ respectivement) est plus susceptible de provoquer un assourdissement de la voyelle et/ou une réduction de la durée de celle-ci ainsi que d'avoir un impact sur les valeurs formantiques. Dans la suite du protocole, il sera donc nécessaire de proposer plusieurs items par position (initiale, médiane et finale) pour chaque erreur ciblée.
- D'autre part, en raison de la différence entre le niveau ciblé et le niveau réel des participants puisqu'à cause de la pandémie, nous avons dû réaliser les passations au Japon et non en France, auprès d'apprenants de niveau intermédiaire (B1+, 17 sujets avaient un niveau supérieur ou égal au B2) et non pas des débutants complets. La majorité des sujets enregistrés avait de plus effectué des séjours de plusieurs mois à l'étranger, principalement dans des pays anglophones. Dans le futur, il serait intéressant de se questionner sur l'influence des autres langues connues par l'apprenant sur l'acquisition de la prononciation du français. (cf. pour l'anglais L2 Wen & Jiha (2016) et de nous appuyer sur des études quantitatives issues de corpus d'apprenants japonophones pour déterminer précisément la fréquence d'apparition de chaque réalisation erronée, en fonction du niveau des apprenants (via, par exemple, l'analyse du corpus CLIJAF, Detey *et al.*, soumis).

D'autant plus qu'au-delà des contextes phonétiques, il faut aussi rappeler la dimension très individuelle du système perceptif de l'apprenant (Billières, 2012) et la difficulté de constituer un groupe homogène quand on fait passer ce type d'expérimentation.

Dans le contexte classe, la méthode verbo-tonale offre une pédagogie individualisée basée sur le diagnostic en temps réel de l'erreur de l'apprenant. Dans notre précédent design expérimental, il

⁵ <https://paty.irit.fr/demo/>: Plateforme de Parole Atypique développée en partenariat entre l'IRIT et le LPL

est dommage que les apprenants aient reçu le même *feedforward*, qu'ils aient produit l'erreur attendue ou pas, ce qui ne leur permettait pas de comprendre quand ils avaient produit une erreur ou simplement d'effectuer quelque chose de différent. Même si cela n'a pas été possible à mettre en place dans l'étude précédente, un design de type magicien d'OZ serait à privilégier⁶. Nous nous questionnons néanmoins sur la faisabilité expérimentale de ce type de protocole, évidemment plus proche de la réalité de terrain mais qui soulève un certain nombre de contraintes méthodologiques, comme par exemples, le facteur de risque d'erreurs de diagnostic de la part du praticien de correction phonétique qui devra jouer le rôle du magicien d'OZ en temps réel ou encore le fait d'assurer l'équilibre entre les différentes conditions (nombre de répétitions par sujet).

4 Vers un nouveau protocole

Dans la partie précédente, nous avons souligné la grande variabilité dans les productions des apprenants. Nous avons expliqué la nécessité d'augmenter le nombre de stimuli (en prenant en compte plusieurs positions pour chaque son ciblé), la difficulté de créer des groupes homogènes et l'importance du recours à un magicien d'Oz pour pouvoir proposer des corrections adaptées aux erreurs produites, en temps réel.

À partir de ces réflexions, nous allons modifier certains items afin de mieux contrôler le contexte phonétique du segment et d'éviter des effets de coarticulation. Le segment cible apparaîtra dans différentes positions dans le mot (sauf pour le /ʁ/ final). Dès lors, pour pouvoir inclure la variable "position" dans le modèle statistique et augmenter la puissance statistique du modèle, nous augmenterons le nombre d'items afin d'avoir au moins 3 mots par position.

Segment visé	initiale	médiane	finale
/b/	Bagage, bonheur, bilan	Robot, ibis, abonnement	Robe, kébab, club
/ʒ/	Japonais, jolie, génial	Age, déjeuner, aujourd'hui	Page, cage, horloge
/y/	Unique, utile, utopie	Numéro, fumée, furieux	Salut, manu, début
/ã/	Encre, embrasser, antique	Attendez, l'Angleterre, mensonge	Moment, vivant, Laurent

TABLE 4 : Proposition d'items pour le nouveau protocole

Si la situation actuelle nous en donne la possibilité, nous prévoyons de réaliser de nouvelles passations sur Toulouse, avec un groupe de Japonais débutants, dans cadre d'un programme d'échange avec l'université de Toulouse 2, en février 2023.

Dans le protocole précédent, nous avons cherché à comparer des gestes co-verbaux pouvant avoir deux usages différents: un usage iconique et un usage métaphorique dans l'acquisition du suprasegmental et du segmental. La comparaison de deux usages gestuels différents constitue véritablement l'un des points forts et novateurs de notre projet. En outre, si l'importance des gestes iconiques a déjà été démontrée scientifiquement (Gluhareva & Prieto 2017; Zhang *et al.*, 2020 ; Connell *et al.*, 2013 ; Morett & Chang, 2015 ; Yuan *et al.*, 2019) dans le cas de

⁶ Cela avait été initialement envisagé pour cette première expérimentation mais cela s'est avéré incompatible avec les conditions de réalisations des passations puisqu'il aurait fallu trouver un expérimentateur spécialisé dans le diagnostic en correction phonétique sur place, au Japon.

l'acquisition du suprasegmental, très peu d'études se sont intéressées à l'impact de la gestuelle dans l'acquisition du segmental (nous citerons en exemple Hardisson, 2003 sur l'impact des gestes articulatoires dans l'acquisition de contrastes phonémiques). Grâce à notre protocole nous espérons pouvoir apporter des éléments nouveaux concernant l'incidence de la gestualité co-verbale iconique et métaphorique dans l'acquisition du segmental en L2.

Enfin, nous proposerons un design expérimental de type magicien d'OZ, mais seulement après avoir fait passer un test de prononciation aux apprenants et de nous être assurés de l'homogénéité du groupe. En effet, le recrutement sur un critère de niveau n'est pas suffisant étant donné la précision requise pour l'évaluation de la prononciation. En outre, les niveaux du CECRL ne sont pas déterminés sur la base de critères de prononciation précis et ils ne permettent pas de garantir l'homogénéité des participants quant à leur maîtrise de la prononciation. L'idéal serait de demander à un panel de spécialistes d'évaluer la prononciation des participants avant les passations pour équilibrer les trois groupes. Tout l'intérêt et la difficulté de la pratique verbo-tonale et de notre étude à venir, sera de s'adapter à la réalité du système de l'apprenant, en temps réel. Pour pouvoir dépasser une approche de type « étude de cas » qui rendrait très difficile des analyses groupées habituellement utilisées en statistiques, nous devons préalablement contrôler les niveaux réels des sujets enregistrés, sur la base de critères phonétiques, aussi bien en perception qu'en production de la parole.

Remerciements

L'étude présentée ici a été réalisée dans le cadre du projet INGPRO (Incidence des Gestes sur la PRONonciation, financement R&S, région Occitanie 2019-2021). Nous remercions la région Occitanie pour le soutien financier et l'ensemble des partenaires du projet, en particulier Amélie Le Chevanton, Maxime Le Coz, Corentin Barcat et Thomas Pellegrini.

Références

ALAZARD, C. (2013). *Rôle de la prosodie dans la fluence en lecture oralisée chez des apprenants de Français Langue Etrangère*. Thèse de doctorat, Université de Toulouse 2.

ALAZARD-GUIU, C., SANTIAGO, F., MAIRANO, P. (2018). L'incidence de la correction phonétique sur l'acquisition des voyelles en langue étrangère : étude de cas d'anglophones apprenant le français. Actes de *Journées d'Etudes sur la Parole JEP 2018*, 116-124.

BILLIERES, M. (2012). La perception des sonorités parolières entre cognition et émotion, In: CALVO MEDINA M.;-V.; MURILLO PUYAL, J. (Éds.) « Perception phonique et Parole », CIPA, Mons, 23-38.

CONNELL, L., CAI, Z.G., & HOLLER, J. (2013). Do you see what I'm singing? Visuospatial movement biases pitch perception. *Brain and Cognition* 81, 124–130.

DETEY, S., FONTAN, L., LE COZ, M., BARCAT, C., & KAWAGUISHI, Y. (soumis). L3 French segmental substitutions in spontaneous speech of Japanese learners and corpus data for computer-assisted pronunciation training research: An illustrated overview. Soumis dans le *Journal of the Phonetic Society of Japan*.

FONTAN, L., & LE COZ, M. (2017). Correction automatique d'erreurs de prononciation en L2 : démonstration d'outils logiciels pour les apprenants japonais de FLE. Actes de *Congres CAP*.

FONTAN, L., LE COZ, M., DETEY, S., DOMIN, C., et HAPPEL, F. (2018). Présentation du logiciel d'entraînement à la prononciation CAPT-L2 : aspects phonétiques et lexicaux. Workshop « Traitement automatique de la parole et ressources pour la didactique de l'oral en L2 : variation, corpus, techniques », IRIT, Toulouse.

GLUAREVA, D., & PRIETO, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research* 21 (5), 609-631.

GULLBERG, M. (2006). Some reasons for studying gesture and second language acquisition (Hommage à Adam Kendon). *International Review of Applied Linguistics in Language Teaching* 44 (2), 103-124.

HARDISON, D.M. (2003). Acquisition of second-language speech: effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, 24(4), 495-522.

HEBA, A. (2021). *Reconnaissance automatique de la parole à large vocabulaire : des approches hybrides aux approches End-to-End*. Thèse de doctorat, Université Toulouse 3 Paul Sabatier.

KAMIYAMA, T, DETEY, S, KAWAGUCHI, Y. (2016). Les japonophones. In S. Detey, I. Racine, Y. Kawaguchi & J. Eychenne (éds), *La prononciation du français dans le monde: du natif à l'apprenant*. Paris : CLE International 155-162.

KISLER, T., REICHEL, U. D., SCHIEL, F. (2017). Multilingual processing of speech via web services, *Computer Speech & Language* (45), 326-347.

MCCAFFERTY, S. (2006). Gesture and the materialization of second language prosody. *International Review of Applied Linguistics in Language Teaching* 44(2), 197-209.

MCNEILL, DAVID (1992). *Hand and mind: what gestures reveal about thought*. Chicago/London : University of Chicago Press.

MORETT, L. M., & CHANG, L. Y. (2015). Emphasizing sound and meaning : Pitch gestures enhance Mandarin lexical tone acquisition. *Language. Cognition and Neuroscience* 30, 347-353.

SMOTROVA, T. (2017). Making pronunciation visible : Gesture in teaching pronunciation. *Tesol Quarterly* 51(1), 59-89.

TELLIER, M. (2008). The effect of gestures on second language memorization by young children. *Gesture* 8.2, 219-235.

TELLIER, M., GUARDIOLA, M., & BIGI, B. (2011). Types de gestes et utilisation de l'espace gestuel dans une description spatiale: méthodologie de l'annotation. Actes de *Atelier DEGELS, 18èmes conférence annuelle Traitement Automatique des Langues Naturelles (TALN)*, 45-56.

YUAN, C., GONZALES-FUENTES, S., BAILLS, F., & PRIETO, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*. 41(1), 5-32.

WEN, X., & JIA, Y. (2016). Joint Effect of Dialect and Mandarin on English Vowel Production: A Case Study in Changsha EFL Learners. *Interspeech*, 185-189.