Contrôle de l'anticipation vocalique d'arrondissement en Langage Parlé Complété

Virginie Attina, Marie-Agnès Cathiard, Denis Beautemps

Institut de la Communication Parlée (ICP) - UMR 5009 CNRS / INPG / Univ. Stendhal
46, av. Félix Viallet 38031 Grenoble Cedex 1, France
Mél : attina@icp.inpg.fr, cathiard@icp.inpg.fr, beautemps@icp.inpg.fr

ABSTRACT

"Langage Parlé Complété (LPC)" is the French manual system – corresponding to Cued Speech – used to complement lip reading and thus to enhance speech perception for hearing-impaired people. In an anticipatory rounding context, a French speaker was audiovisually recorded pronouncing and coding [i#yi] sequences with two different pause durations. The relative timing of the hand and lip movements and of the corresponding acoustic signal was quantified. The results showed that : (i) the manual cue follows the temporal organization of visible speech; (ii) the manual target position is always ahead of the corresponding lip target.

1. INTRODUCTION

Dans cet article, nous nous intéressons à l'étude des stratégies de coordination de la main et des lèvres, en relation avec le son produit, adoptées par le locuteur en situation de codage de Langage Parlé Complété (LPC) dans un contexte d'anticipation articulatoire. Le LPC, conçu pour faciliter la réception de la langue parlée pour les malentendants, désambiguïse les formes labiales identiques à l'aide de la main dont la position autour du visage vu de face code les voyelles et la forme (les clés digitales) code les consonnes [Cor67] (cf. sur la figure 1 quelques exemples). La coordination naturelle des gestes orofaciaux avec les mouvements de la main est la clé du système et constitue un enjeu majeur dans une perspective de synthèse audiovisuelle de la parole intégrant le codage LPC. Nous proposons d'étudier cette coordination à la lumière des travaux déjà menés sur l'anticipation articulatoire, en particulier celle d'arrondissement.

L'anticipation vocalique d'arrondissement a été largement étudiée sur le plan articulatoire, à partir du geste de la lèvre supérieure, pour des séquences V1CV2, où V1 est une voyelle étirée comme [i], V2 une voyelle arrondie comme [u] et C une ou plusieurs consonnes : le geste de protrusion/arrondissement du [u] pouvant démarrer dès la fin acoustique du [i] (modèle look-ahead [Hen67]) ou à date fixe avant le début acoustique du [u] (modèle time-locked [Bel81]) ; Perkell & Chiang [Per86] ayant finalement opté pour un modèle hybride à deux phases (cf. [Per90] pour une évaluation de ces trois modèles).

Un autre modèle, le « movement expansion model » ou M.E.M. a été établi pour le français [Abr95a]. Il prédit que l'anticipation ne peut commencer qu'à partir d'une constante d'exécution (qui est déterminée pour une transition sans consonne [iy]) et que la durée de l'anticipation augmentera avec la longueur de la séquence consonantique. Ainsi, l'anticipation ne serait déterminée ni à partir de la fin de V1 ni à partir du début de V2 mais serait fonction de la durée de l'intervalle consonantique. Ce modèle M.E.M. a été aussi appliqué à la constriction labiale pour la dimension d'arrondissement [Abr95b].

Ce phénomène d'anticipation vocalique qui repose sans doute à la base sur un principe d'économie articulatoire peut s'avérer efficace sur le plan perceptif. En particulier, il a été montré que l'information d'arrondissement, testée dans des transitions [i#y], est récupérable en perception visuelle jusqu'à 200 ms avant le début acoustique de la voyelle [y] [Cat96]. Il s'agit donc dans cette étude d'étudier l'organisation temporelle du codage manuel LPC en relation avec la structure coarticulée de la parole, c'est-à-dire de déterminer si le codage manuel LPC suit l'organisation articulatoire et perceptive naturelle de la parole.

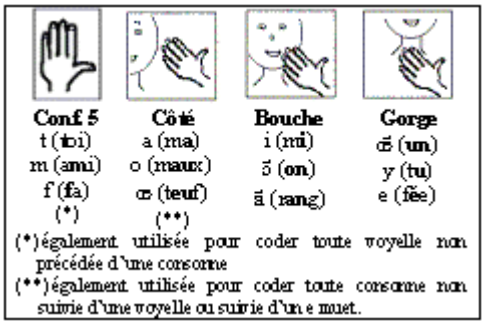

Figure 1. Positions de la main et configurations des doigts (clés digitales) utilisées en Langage Parlé Complété pour coder certaines voyelles et consonnes du Français.

2. PROTOCOLE EXPERIMENTAL

Cette étude s'appuie sur l'analyse de l'enregistrement vidéo d'un codeur LPC prononçant et codant un corpus sans sens lexical composé de voyelles du Français.

Corpus

Le corpus a été constitué afin d'étudier le geste d'arrondissement dans des séquences [i#yi], sans geste

consonantique, insérées dans la phrase porteuse : « T'as mis : UHI ise ? » [tami#yii:z], dans laquelle « UHI » représente un nom d'indien et « ise » un pseudo-verbe à la 3ème personne du singulier. Le codage LPC correspondant à cette phrase est illustré en figure 2. Pour cette phrase, la clé digitale reste la même, aussi bien pour coder [m, t] que pour les voyelles isolées (à l'exception de la clé utilisée pour la consonne finale [z]). Nous enregistrons 6 réalisations de cette phrase selon deux conditions de pause, courte [#] et longue [#:], afin de maximiser la variabilité de l'anticipation. Au total, nous obtenons 12 séquences.

Figure 2. Codage LPC de la phrase « T'as mis : UHI ise ? » [tami#yii:z].

Locuteur

La locutrice est une femme âgée de 36 ans qui pratique le LPC quotidiennement depuis plus de 8 ans, en raison de la surdité de son enfant. Elle est titulaire du diplôme de codeuse LPC depuis 1996 et code en classe régulièrement selon les besoins de codage des élèves. Elle nous a été recommandée par une orthophoniste enseignante de LPC, qui, pour son choix, s'est appuyée sur la qualité de codage de cette personne, c'est-à-dire la fluidité et la bonne visibilité des mouvements de sa main durant le codage ainsi que sa forme labiale bien claire.

Matériel et procédure

L'enregistrement a été effectué au moyen du poste Visage-Parole de l'ICP [Lal91] par deux caméras en vue de face, l'une filmant le visage du locuteur et sa main en vue d'ensemble, l'autre filmant ses lèvres en gros plan. La synchronisation des deux caméras était assurée par le repérage d'un pavé de diodes (LED) allumé sur une trame vidéo. Des pastilles colorées ont été placées sur le dos de la main droite du sujet pour récupérer les mouvements de sa main – la pastille la plus proche du poignet étant celle retenue pour notre étude – et ses lèvres ont été maquillées en bleu afin de récupérer avec précision les contours des lèvres. Enfin, le sujet portait des lunettes aveugles dont le centre était repéré par une pastille bleue. La tête était maintenue par un système de casque fixe afin de limiter toute mobilité.

Traitement des données

Les images correspondant aux séquences ont été numérisées à une fréquence d'échantillonnage de 25 Hz et ont été détramées pour le traitement (ce qui nous permet d'avoir une information toutes les 20 ms). Grâce à un système de traitement des images (système TACLE [Aud00]), nous avons pu mesurer pour chaque séquence le décours temporel de l'aire intérolabiale (S). N'ayant pas d'enregistrement en vue de profil, nous n'avons pas le décours de la lèvre supérieure. Notons cependant que le paramètre d'aire aux lèvres est pertinent à la fois sur le plan articulatoire et sur le plan acoustique pour l'arrondissement [Abr80]. Nous avons également mis au point un logiciel de suivi de pastilles colorées qui nous délivre les coordonnées en x et en y du barycentre de chaque pastille de la main. Le son a été échantillonné à une fréquence de 22050 Hz.

Le résultat du traitement fournit trois signaux synchrones pour les 6 séquences en longue pause et les 6 séquences en petite pause : (i) la trajectoire de l'aire intérolabiale S avec un point toutes les 20 ms, (ii) à la même cadence, la trajectoire des coordonnées x et y du centre de la pastille de la main mesurées en référence avec le centre de la lunette droite (iii) et le signal acoustique (figure 3). Sur chacune des trajectoires, pour la portion [i#yi], le début des transitions entre voyelles a été repéré par la position du pic d'accélération et la fin par le pic de décélération, la tenue de la voyelle étant déterminée par l'intervalle entre la fin de la transition précédente et le début de la suivante. Nous obtenons ainsi pour le signal de la main les repères suivants : M1 est le démarrage du geste LPC vers la voyelle [y], M2 l'atteinte de la position cible qui sera tenue jusqu'à M3, date à laquelle la main démarre le geste vers la voyelle suivante [i] ; enfin, M4 correspond à l'atteinte de la position manuelle du [i]. Pour les lèvres nous avons aussi quatre repères : L1 pour le démarrage du geste labial d'arrondissement, L2 pour l'atteinte de la cible du [y] ; les repères L3 et L4 correspondent respectivement au début du geste et à la cible du [i]. Pour le signal acoustique, nous avons A1 qui indique le début de la voyelle [y] et A2 celui de la voyelle [i] (à partir d'un critère sur l'énergie de la voyelle).

3. RESULTATS

Sur le plan articulatoire, pour l'ensemble des 12 réalisations de cette séquence (toute condition de pause confondue), la valeur moyenne de l'aire intérolabiale S des cibles [i] est de 2,4 cm^2, et celle des cibles [y] est de 0,4 cm^2, ce qui est classique pour ce type de voyelle. La valeur moyenne de la pause, mesurée à partir du signal acoustique, est de 387 ms pour les 6 réalisations avec petite pause et de 1075 ms pour les 6 réalisations avec longue pause. Remarquons que ces durées traduisent un rythme de parole relativement lent de la locutrice.

Il est à noter que, pour l'analyse temporelle de ces séquences, seule la trajectoire de la coordonnée y de la pastille a été prise en compte car le mouvement de la main se produit en majeure partie selon la direction verticale ; la position codant le [i] étant la bouche et celle codant le [y] étant la gorge (figures 1 et 2). La trajectoire de la main selon cette direction est appelée trajectoire LPC dans la suite.

Nous allons présenter séparément la coordination main-lèvres-son observée pour la transition [i#y] qui nous permettra de tester l'anticipation d'arrondissement à travers la pause et pour la transition [yi] qui nous donnera

le geste minimal incompressible pour passer d'une voyelle à l'autre.

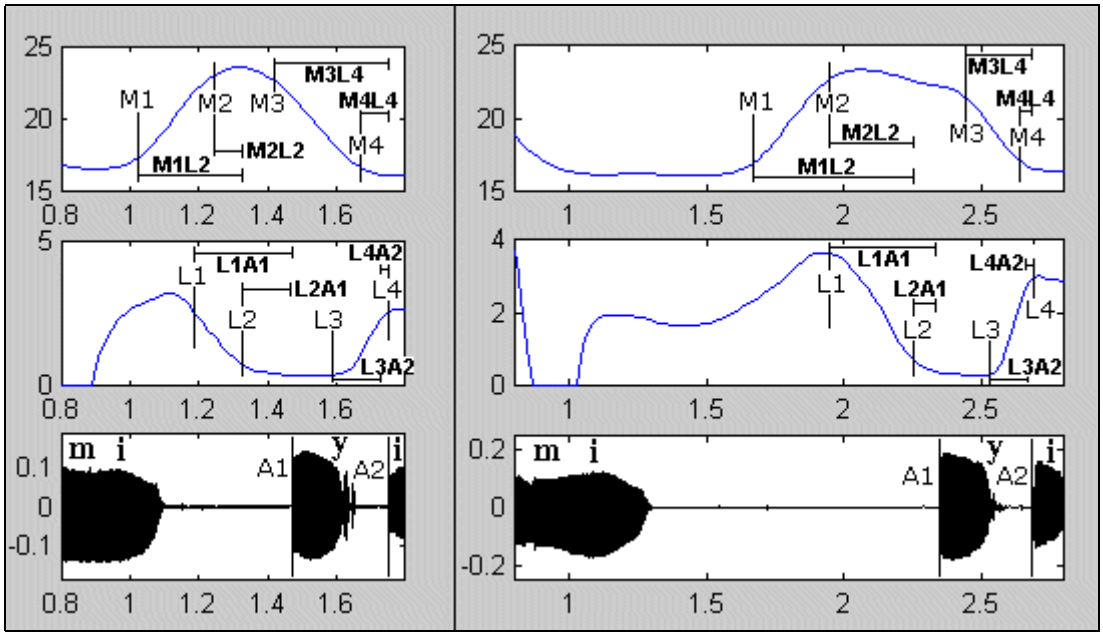

Figure 3: Exemple de séquence [mi # yi] avec petite pause (à gauche) et longue pause (à droite) : de haut en bas, déplacement en y de la main (quand y augmente, la main descend), évolution temporelle de l'aire aux lèvres (S) et signal acoustique correspondant. Sur chaque signal sont indiqués les instants repérés et les différents écarts temporels calculés.

Etude de la transition de [i] vers [y]

A partir des instants repérés sur les différents signaux, les écarts temporels suivants ont été calculés :

- entre le début de la transition labiale vers [y] et l'instant de son début acoustique (L1A1) ;
- entre l'atteinte de la cible labiale du [y] et l'instant de son début acoustique (L2A1) ;
- entre le début de la transition LPC de [i] vers [y] et l'instant d'atteinte de la cible labiale du [y] (M1L2) ;
- entre l'atteinte de la cible LPC du [y] et l'atteinte de la cible labiale du [y] (M2L2). (figure 3)

Les deux premiers écarts nous permettent de quantifier l'anticipation sur le plan labial par rapport au début acoustique de la voyelle ; les deux derniers nous donnent l'anticipation de la main sur les lèvres.

En ce qui concerne l'anticipation labiale par rapport au son, les résultats montrent que la constriction des lèvres démarre en moyenne 419 ms en longue pause et 280 ms en petit pause avant le début acoustique du [y] (L1A1 ; t=4.07 pour ν=10 ddl, supérieur à la valeur théorique $t_{0.01}$=3.169). Cela confirme les résultats de la littérature qui montrent que l'anticipation aux lèvres est de plus en plus importante avec l'augmentation de l'intervalle consonantique [Abr95a] ou avec l'allongement de la pause [Cat96]. Par contre, la cible labiale est atteinte en moyenne 134 ms en petite pause et 76 ms en longue pause avant le début acoustique (L2A1 ; t=4.08, ν=10 ddl). Ce dernier résultat surprenant peut s'expliquer par le comportement particulier de notre locutrice en longue pause qui, après la réalisation du [i], réalise un relâchement des lèvres, ce qui a pour conséquence une augmentation de l'aire intérolabiale (4,58 cm^2 en moyenne pour les réalisations en longue pause soit le double de l'aire moyenne donnée précédemment pour les cibles [i]). On peut donc comprendre que, bien que démarrant plus précocement qu'en petite pause son geste d'arrondissement, elle atteigne plus tardivement la cible fermée du [y]. Ce phénomène de relâchement des lèvres peut s'expliquer par les valeurs de grande pause particulièrement longues (1075 ms en moyenne).

En ce qui concerne le geste manuel LPC par rapport au geste des lèvres, M1L2 – soit le début du geste de la main par rapport à l'atteinte de la cible labiale – augmente avec la longueur de la pause (en moyenne 613 ms en longue pause contre 283 ms en petite pause ; t=9.3, ν=10 ddl). De même, si on considère l'atteinte de la cible LPC par rapport à la cible labiale (M2L2), on remarque une anticipation de 63 ms en petite pause contre 283 ms en longue pause (t=11.4, ν=10 ddl).

Etude de la transition de [y] vers [i]

Les écarts temporels suivants ont été calculés :

- entre le début de la transition labiale vers [i] et l'instant de son début acoustique (L3A2) ;
- entre l'atteinte de la cible labiale du [i] et l'instant de son début acoustique (L4A2) ;
- entre le début de la transition LPC vers [i] et l'instant d'atteinte de la cible labiale du [i] (M3L4) ;
- entre l'atteinte de la cible LPC du [i] et l'atteinte de la cible labiale (M4L4).

En considérant le début du geste labial vers [i] (L3A2), on remarque que celui-ci démarre en moyenne 133 ms avant le début acoustique, avec un effet non significatif de la condition de pause (t=1.7 inférieur à la valeur théorique $t_{0.01}$=3.169, ν=10 ddl), ce qui n'est pas surprenant. Cette valeur nous donne la durée moyenne incompressible du geste pour passer de [y] à [i] pour notre locutrice : elle est supérieure à celle observée dans une étude testant une transition identique mais sans codage LPC (80 ms [Cat92]). Les résultats pour L4A2 montrent un retard moyen de 30 ms de la cible labiale sur le son.

En ce qui concerne les gestes de la main en rapport avec ceux des lèvres, le début du geste LPC vers la cible du [i] démarre en moyenne 295 ms avant l'atteinte de la cible labiale, quelle que soit la condition de pause (M3L4). L'atteinte de la cible LPC (M4L4) pour [i] se fait avec une avance de 77 ms en moyenne sur la cible labiale (ces deux valeurs d'avance sont significativement différentes de 0 ; respectivement, t=30.8 et t=9.9, ν=11 ddl, valeur théorique $t_{0.01}$=3.106).

4. CONCLUSION

En conclusion de cette étude de l'anticipation d'arrondissement, nous constatons un phénomène d'avance de la main par rapport au geste vocalique correspondant. En condition de pause prosodique [i # y], l'atteinte de la cible LPC précède toujours (de 283 ms à 63 ms selon la pause) l'atteinte de la cible vocalique aux lèvres, qui peut elle même précéder le son (de 134 ms en petite pause à 76 ms en longue pause). En condition de deux voyelles [yi] sans pause, nous avons pu aussi montrer une avance de l'atteinte de la cible LPC sur la cible labiale vocalique de 77 ms en moyenne. Nous remarquerons donc que pour nos deux transitions le geste LPC est toujours en avance sur le geste labial.

On peut donc conclure que : (i) le codage manuel LPC suit les contraintes temporelles de la parole coarticulée, mises en évidence dans cette expérience par la manipulation prosodique de l'anticipation du geste d'arrondissement ; (ii) la main atteint sa position toujours plus précocement que la mise en forme des lèvres. On peut donc retenir une organisation des coordinations orofaciales et manuelles qui permettrait une désambiguïsation progressive – par restriction des choix possibles – par la main puis par les lèvres, du son articulé. Ainsi, en considérant par exemple la réalisation d'un [y], lorsque la main arrive en position gorge, le décodeur saura que la voyelle à venir peut être [y], [e] ou [œ], la sélection de [y] étant finalement opérée dès que les lèvres seront suffisamment arrondies pour éliminer les autres voyelles. Cette hypothèse de décodage progressif par désambiguïsation, formulée à partir de nos observations au niveau de la production du code manuel, mériterait d'être vérifiée par une étude perceptive s'appuyant sur un paradigme de dévoilement progressif du signal vidéo (ou gating).

BIBLIOGRAPHIE

[Abr80] Abry C., Boë L.-J., Corsi P., Descout R., Gentil M. & Graillot P. (1980), "Labialité et phonétique. Données fondamentales et études expérimentales sur la géométrie et la motricité labiales », Publications de l'Université des Langues et Lettres de Grenoble.

[Abr95a] Abry C. & Lallouache M.T. (1995), "Le M.E.M. : un modèle d'anticipation paramétrable par locuteur. Données sur l'arrondissement en français", Bulletin de la Communication Parlée, Vol. 3, pp. 85-99.

[Abr95b] Abry C. & Lallouache T. (1995), "Modeling lip constriction anticipatory behaviour for rounding in French with the MEM (Movement Expansion Model)", Proc. of the 13th International Congress of Phonetic Sciences, Vol. 4, pp. 152-155.

[Aud00] Audouy M. (2000), "Logiciel de traitement d'images vidéo pour la détermination de mouvements des lèvres", Projet de fin d'études, option génie logiciel, ENSIMA Grenoble.

[Bel81] Bell-Berti F. & Harris K.S. (1981), "A temporal model of speech production", Phonetica, Vol. 38, pp. 9-20.

[Cat92] Cathiard M.A. & Lallouache M.T. (1992), "L'apport de la cinématique dans la perception visuelle de l'anticipation et de la rétention labiales", 19ème Journées d'Etudes sur la Parole, Bruxelles, 19-22 Mai 1992, pp. 25-30.

[Cat96] Cathiard M.A., Lallouache M.T. & Abry C. (1996), "Does movement on the lips mean movement in the mind ?". In D.G. Stork & M.E. Hennecke (Eds), Speechreading by humans and machines : Models, Systems and applications, Vol. 150, NATO ASI Series F, pp. 211-219.

[Cor67] Cornett R.O. (1967), "Cued Speech", American Annals of the Deaf, Vol. 112, pp. 3-13.

[Hen67] Henke W. L. (1967), "Preliminaries to speech synthesis based on an articulatory model", Proc. 1967 IEEE Boston Speech Conference, pp. 170-171.

[Lal91] Lallouache M. T. (1991), « Un poste visage-Parole couleur. Acquisition et traitement automatique des contours des lèvres », PhDThesis, INP Grenoble.

[Per86] Perkell J.S. & Chiang C. (1986), "Preliminary support for a "hybrid model" of anticipatory coarticulation", Proc. of the 12th International Congress of Acoustics, A3-6.

[Per90] Perkell J.S. (1990), "Testing theories of speech production : implications of some detailed analyses of variable articulatory data". In W.J. Hardcastle & A. Marchal (Eds.), Speech Production and speech modelling, pp. 263-288. Dordrecht/ Boston/ London : Kluwer Academic Publishers.

Remerciements à Mme M. Marthouret, orthophoniste au CHU de Grenoble, pour ses conseils, à Mme G. Brunel, notre codeuse LPC, pour avoir accepté les contraintes de l'enregistrement, à C. Savariaux pour son aide technique.