

L'ancrage de la main sur les lèvres :

Langue Française Parlée Complétée et anticipation vocalique

Attina Virginie, Cathiard Marie-Agnès & Beautemps Denis

Institut de la Communication Parlée (ICP)
46, av. Félix Viallet 38031 Grenoble Cedex 01
Tél.: ++33 (0)4 76 57 45 37 - Fax: ++33 (0)4 76 57 47 10
Mél: attina, cathiard, beautemps@icp.inpg.fr

ABSTRACT

This paper presents a temporal study of the two visible anticipatory gestures for vowels, i.e. vowel rounding and vowel height or aperture, in coordination with the anticipatory gesture of the hand in the framework of French Cued Speech. Using a silent pause paradigm, our results showed an advance of the hand on rounding and a manual synchronisation for the aperture gesture. Whatever the explanation of these two behaviours, it remains that the temporal order of information in Cued Speech is one where the hand leads or meets the mouth which leads the sound.

1. INTRODUCTION

Nos recherches sur la coordination de la main et des lèvres en langue Française Parlée Complétée (LPC) — augment manuel de la labiolecture à l'usage des malentendants, adapté du Cued Speech, Cornett [1] — ont mis en évidence, pour des syllabes CV, l'avance systématique de la main sur les gestes orofaciaux (Attina et al. [2]). La main — qui désambiguïse, dans un cadre CV, les voyelles par sa position autour du visage et les consonnes par la configuration de ses doigts (figure 1) — est-elle toujours anticipante dans des conditions d'avantage extrêmes qui nous permettent de tester son ancrage par rapport à l'anticipation des configurations visibles de la sortie du conduit vocal ? Nous avons choisi de tester cet ancrage dans le contexte de la pause acoustique où il a été déterminé que l'anticipation de la dimension d'arrondissement comme d'aperture peut atteindre des taux d'expansion inusités par rapport à ceux rencontrés dans les transitions V(C...)V (Abry et al. [3]). Ce que nous avons pu déjà démontrer, dans une étude préliminaire, pour l'arrondissement dans la LPC (Attina et al. [4]). Ce paradigme nous permet en outre d'étudier l'anticipation vocalique sans que le positionnement de la main ainsi que celui des lèvres ne soit *perturbé* par la production de consonnes intermédiaires.

Notre question se reformule ici plus précisément : dans le geste de position manuelle, l'initiation et la cible sont-elles respectivement en avance sur l'initiation et la cible du geste pour le geste d'arrondissement, mais aussi pour l'aperture, l'autre dimension visible des voyelles ? La quantification de cette anticipation, conçue comme reflet

d'une coordination motrice, sera ici testée en manipulant la longueur de la pause intervocalique.

| Configurations pour les consonnes | | | | Positions pour les voyelles | | |
|--|---|---------------------------|--------------------------------------|---|---------------------------------------|---|
| | | | | | | |
| N°1 p (par) d (dos) ɔ (joue) | N°2 k (car) v (va) z (zout) | N°3 s (sel) ʀ (raf) | N°4 b (bar) ʁ (rou) ʁ (rou) | Side a (ma) o (eau) ɔ (teuf) (**) | Mouth i (mi) ɛ (on) ā (yang) | Chin ɛ (mais) u (mou) o (fort) |
| | | | | | | |
| N°5 t (toi) m (ami) f (fa) (*) | N°6 l (la) ʃ (chat) ʁ (vigne) w (oui) | N°7 g (gare) | N°8 j (jille) ɲ (camping) | Cheek bone ɛ (main) ɔ (feu) | Throat ɑ (un) y (tu) e (tée) | égale- ment utilisée pour coder une voyelle isolée ***) égale- ment utilisée pour coder une consonne isolée ou suivie d'un schwa. |

Figure 1 : Clés LPC pour les consonnes et les voyelles en langue Française Parlée Complétée.

2. MÉTHODE

2.1. Corpus

Deux corpus ont été enregistrés. Le corpus 1 étudie le geste de constriction dans des transitions [i#y] (avec une pause plus ou moins longue) insérées dans une phrase porteuse du type « T'as mis : Uhi ise ? » [tami#yii:z], dans laquelle « Uhi » représente un nom propre et « ise » un pseudo-verbe à la 3ème personne du singulier. De la même manière, le corpus 2 étudie le geste d'aperture dans des transitions [i#a], insérées dans la phrase porteuse « T'as mis : Ahi ise ? » [tami#aiz].

En ce qui concerne le codage LPC, la transition [i#y] est codée par la main depuis la position bouche correspondant au [i] jusqu'à la position cou correspondant au [y] avec la configuration n°5 qui code les voyelles isolées (cf. figure 1). La transition [i#a] est codée depuis la position bouche pour le [i] vers la position côté qui code le [a] avec la même configuration (n°5, cf. figure 1). Au total 24 séquences pour chaque corpus ont été enregistrées selon les deux conditions de pause : 12 réalisations de petite pause [#] et 12 réalisations de longue pause [#:].

2.2. Locuteur-codeur

La locutrice est une femme française âgée de 36 ans qui pratique la LPC depuis plus de 8 ans. Elle a eu son diplôme de codeuse professionnelle de LPC en 1996 qui

atteste de la bonne qualité de son codage et depuis code régulièrement en classe selon les besoins des élèves.

2.3. Matériel et enregistrement

L'enregistrement a été effectué dans la chambre sourde anéchoïque de l'ICP. La locutrice est assise, la tête maintenue par un casque fixe afin de rester dans le champ des caméras. Elle porte une paire de lunettes opaques qui lui protège les yeux contre un fort éclairage nécessaire au bon contraste des images. Une pastille de couleur est repérée sur la lunette gauche servant par la suite de point de référence pour les différentes mesures. Deux caméras en vue de face ont été utilisées : l'une en gros plan pour les lèvres et l'autre en vue d'ensemble pour le visage et la main. La synchronisation des deux caméras est assurée par le repérage d'un pavé de diodes allumé sur une trame vidéo. Afin de récupérer avec précision les contours labiaux, les lèvres du sujet sont maquillées en bleu. Pour suivre les mouvements de la main, une pastille de couleur était posée sur le dos de la main droite du sujet.

2.4. Traitement des données

Grâce au poste Visage-Parole de l'ICP, les signaux audio et vidéo ont été acquis de manière synchrone. Le son a été numérisé à une fréquence de 22050 Hz. Les images correspondant aux séquences ont été numérisées à une fréquence d'échantillonnage de 25 Hz puis détrimées de façon à avoir une information toutes les 20 ms. Nous avons utilisé un système de traitement d'images TACLE [5] pour mesurer pour chaque séquence le déroulement de l'aire intérolabiale S . Ce paramètre est en effet pertinent à la fois sur le plan articulaire et sur le plan acoustique pour l'arrondissement (Abry et al., [6]). Il est de plus fortement corrélé à l'aperture des lèvres (séparation verticale des lèvres) (Abry et al., [7]). Dans le cas de transitions $[i\#a]$, la variation de l'aire aux lèvres est due à

l'abaissement de la mâchoire : Cathiard [8] a en effet montré, dans une étude de la perception visuelle du geste d'aperture pour des transitions comparables aux nôtres, l'exacte superposition du déroulement temporel de ces deux gestes. Pour le traitement de la main, nous avons mis au point un logiciel de suivi de pastille colorée nous délivrant les coordonnées en x et y du barycentre de la pastille de la main.

Après traitement, nous obtenons trois signaux synchrones : (i) le déroulement temporel de l'aire aux lèvres avec un point toutes les 20 ms, (ii) les coordonnées x et y au cours du temps du centre de la pastille de la main mesurée en référence au point sur la lunette et (iii) le signal acoustique. Pour notre analyse, nous prendrons en compte uniquement la coordonnée y de la pastille pour les transitions $[i\#y]$ et la coordonnée x pour les transitions $[i\#a]$. En effet, pour la transition $[i\#y]$, le passage de la position bouche à la position cou implique un déplacement majeur en y et très peu de mouvement en x . Inversement, pour la transition $[i\#a]$, nous avons choisi de ne traiter que la coordonnée x car le déplacement se fait de la position bouche à la position côté et implique très peu de mouvement en y (cf. figure 2).

Sur chacune des trajectoires, pour les portions $[i\#y]$ et $[i\#a]$ des deux corpus, le début des transitions a été repéré par la position du pic d'accélération et la fin par la position du pic de décélération [9]. Nous obtenons ainsi pour le signal de la main les repères suivants : M1 est le démarrage du geste LPC à partir de la position bouche codant le $[i]$ vers la position suivante, soit le cou pour $[y]$ ou le côté pour $[a]$ et M2 l'atteinte de la position cible. Pour les lèvres nous avons les repères : L1 pour le démarrage du geste labial de constriction ou d'aperture et L2 pour l'atteinte de la cible vocalique du $[y]$ ou du $[a]$. Pour le signal acoustique, nous avons A1 qui indique le début des voyelles $[y]$ ou $[a]$ suivant le corpus.

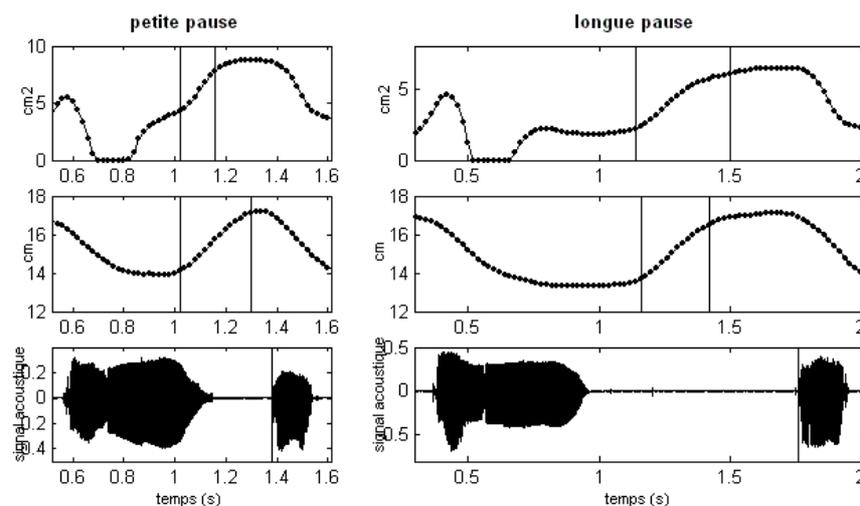


Figure 2 : Exemple de signaux pour la séquence $[tami\#a]$ dans la condition petite pause (gauche) et longue pause (droite). De haut en bas, (i) le déroulement temporel de l'aire aux lèvres, (ii) la coordonnée x au cours du temps de la pastille sur la main et (iii) le signal acoustique. Sur chacun des signaux, tracé des étiquettes servant pour l'analyse temporelle : dans l'ordre, L1 et L2 pour les lèvres, M1 et M2 pour la main et A1 pour le son (cf. texte).

3. RÉSULTATS

Pour l'analyse temporelle, nous avons calculé les intervalles ci-dessous à partir des instants repérés sur les différents signaux (la valeur de l'intervalle résulte de la différence du premier terme (événement 1) moins le second (événement 2) ; une valeur négative signifie donc que l'événement 1 est antérieur à l'événement 2) :

- L1A1 est l'intervalle entre le début du geste labial vocalique et le début acoustique de la voyelle ;
- L2A1 est l'intervalle entre l'atteinte de la cible vocalique aux lèvres et le début acoustique de la voyelle ;
- M1L1 est l'intervalle entre le début du geste manuel et le début du geste labial pour la voyelle ;
- M2L2 est l'intervalle entre l'atteinte de la position LPC par la main et la réalisation de la cible vocalique aux lèvres.

3.1. Anticipation vocalique d'arrondissement

Pour les séquences [i#y], la durée de la pause mesurée à partir du signal acoustique est en moyenne de 389 ms ($\sigma=47$) pour la condition petite pause et de 961 ms ($\sigma=239$) pour la condition longue pause (différence significative $t_{\text{calc}}=8.1$ pour $v=22$ ddl, supérieur à $t_{0,01}=2.819$). Il est à noter que ces durées importantes de pauses correspondent au rythme de parole assez lent de la locutrice.

Pour l'analyse temporelle, nous avons séparé les mesures selon les deux conditions de pause. En ce qui concerne le geste labial par rapport au début acoustique de la voyelle, nous obtenons pour L1A1 une valeur moyenne de -322 ms ($\sigma=44$) pour la condition petite pause et de -369 ms ($\sigma=70$) pour la condition longue pause. Ces deux moyennes ne sont pas statistiquement différentes ($t_{\text{calc}}=1.95$, $v=22$ ddl), ce qui signifierait que la locutrice n'anticipe pas davantage quand la pause est plus longue. Si on regarde maintenant l'atteinte de la cible labiale du [y] par rapport à son début acoustique, nous obtenons pour L2A1 une valeur moyenne de -171 ms ($\sigma=34$) en condition petite pause, significativement différente ($t_{\text{calc}}=4.68$, $v=22$ ddl) de la valeur de -110 ms ($\sigma=28$) obtenue en condition longue pause. Comment expliquer qu'au final, l'anticipation des lèvres sur le son ne soit pas plus importante avec l'allongement de la pause ? En regardant plus particulièrement les décours temporels de l'aire aux lèvres dans les deux conditions, on peut remarquer pendant la réalisation des longues pauses, une augmentation importante de l'aire avant la phase de constriction (cf. figure 3 pour un exemple). Ceci indique que la locutrice, marquant une pause extrêmement longue, effectue un relâchement des lèvres (pour reprendre son souffle) avant de produire le [y].

En ce qui concerne le geste manuel par rapport aux lèvres, nous obtenons pour M1L1 une valeur moyenne de -145

ms ($\sigma=26$) en condition petite pause et de -288 ms ($\sigma=102$) en condition longue pause. Ces moyennes sont testées significativement différentes ($t_{\text{calc}}=4.73$, $v=22$ ddl), ce qui suggère que la main initie son mouvement avant celui des lèvres, en commençant encore plus tôt quand la pause est longue. En ce qui concerne l'atteinte de la cible manuelle par rapport à la cible labiale, une valeur moyenne pour M2L2 de -82 ms ($\sigma=33$) est obtenue en condition petite pause et de -272 ms ($\sigma=69$) en condition longue pause. Ces moyennes sont également significativement différentes ($t_{\text{calc}}=8.58$, $v=22$ ddl) et montrent ainsi une large anticipation de la main sur les lèvres et donc sur le son, l'anticipation étant plus importante quand la pause est plus longue.

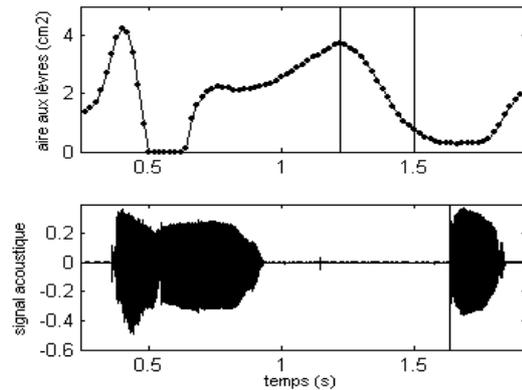


Figure 3 : Exemple de décours temporel de l'aire aux lèvres pour une séquence [tami# :y] en condition longue pause

3.2. Anticipation vocalique de hauteur

Pour les séquences [i#a], la durée de la pause mesurée à partir du signal acoustique est en moyenne de 260 ms ($\sigma=66$) pour la condition petite pause et de 860 ms ($\sigma=202$) pour la condition longue pause (durées significativement différentes, $t_{\text{calc}}=9.7$ pour $v=22$ ddl, supérieur à $t_{0,01}=2.819$).

En ce qui concerne l'analyse temporelle du geste labial par rapport au son, nous obtenons pour L1A1 une valeur moyenne de -313 ms ($\sigma=44$) en condition petite pause et de -593 ms ($\sigma=110$) en condition longue pause. Ces valeurs sont significativement différentes ($t_{\text{calc}}=8.2$, $v=22$ ddl) et montrent que le geste labial vocalique est d'autant plus précoce que la pause est longue. En ce qui concerne la cible labiale, nous obtenons pour L2A1 une valeur moyenne de -154 ms ($\sigma=50$) en condition petite pause et de -280 ms ($\sigma=115$) en condition longue pause. Ces moyennes sont statistiquement différentes ($t_{\text{calc}}=3.46$, $v=22$ ddl) et démontrent une anticipation labiale pour le geste vocalique d'aperture par rapport au son. De plus les lèvres anticipent davantage lorsque la pause est plus longue.

En ce qui concerne le geste manuel par rapport aux lèvres, nous obtenons pour M1L1 une valeur moyenne de -5 ms ($\sigma=28$) en condition petite pause et de 20 ms ($\sigma=110$) en condition longue pause. Ces moyennes ne sont pas

statistiquement différentes entre elles ($t_{\text{calc}}=0.76$, $v=22$ ddl). Si on mélange les deux conditions, on obtient une moyenne générale de 7.5 ms ($\sigma=80$) pour M1L1, statistiquement non différente de 0 ($t_{\text{calc}}=0.46$, $v=23$ ddl, inférieur à $t_{0.01}=2.807$). Ceci suggère que la main et les lèvres initient leur mouvement en même temps. Ce résultat pourrait s'expliquer par le fait que les lèvres anticipent beaucoup sur le son. Ainsi la main ne pourrait anticiper davantage. Pour l'atteinte de la cible manuelle par rapport à la cible labiale, nous obtenons pour M2L2 une valeur moyenne de 50 ms ($\sigma=59$) en condition petite pause et de 7 ms ($\sigma=129$) en condition longue pause, moyennes statistiquement non différentes ($t_{\text{calc}}=1.06$, $v=22$ ddl). En mélangeant les deux conditions, on obtient une moyenne générale de 28.5 ms ($\sigma=100$) pour M2L2, statistiquement non différente de 0 ($t_{\text{calc}}=1.38$ pour $v=23$ ddl, inférieur à $t_{0.01}=2.807$). Ces résultats montrent donc que la main atteint sa cible en même temps que la réalisation de la cible labiale. Dans ce cas la main n'anticipe pas sur les lèvres.

Comment pouvons-nous comprendre cette absence d'anticipation ? Ce résultat pourrait, selon nous, s'expliquer par la position codant la voyelle [a]. En effet, pour cette voyelle, le codeur pointe simplement sur le côté du visage et non pas sur une partie bien précise de celui-ci comme il le fait quand il pointe pour [y] en allant toucher le cou au niveau de la pomme d'Adam. Ainsi il se pourrait que la main ait tendance à continuer sa course en se synchronisant avec les lèvres pour ne s'arrêter qu'en même temps que la cible labiale : ceci explique que les atteintes des gestes manuel et labial soient synchrones.

4. CONCLUSION

En conclusion de cette étude sur l'ancrage de l'anticipation de la position de la main sur celle des configurations labiales, la manipulation de la pause prosodique confirme, sur un corpus plus étendu que celui de notre étude de 2002 [4], l'anticipation de la cible manuelle pour la dimension d'arrondissement. En ce qui concerne l'aperture, nous mettons en évidence un patron de relations différent puisque nous observons une synchronie des gestes manuel et facial. Notons que dans ce cas la main ne prend cependant pas de retard par rapport aux lèvres.

Il semble donc qu'en Langue Française Parlée Complétée, la coordination des deux gestes soit caractérisée par une avance de la position de la main par rapport à la configuration vocalique. Ceci est la règle pour l'unité CV courante du codage LPC ([2]), et nous venons de montrer que cette règle est préservée même dans le cas particulier d'une anticipation déjà importante aux lèvres, comme c'est le cas pour la dimension d'arrondissement testée ici lors de pauses acoustiques conséquentes. Pour la dimension d'aperture, la coordination particulière, qui se traduit par une synchronie du geste manuel d'éloignement du visage et du geste d'aperture buccale, nous semble davantage s'expliquer par la particularité de la position

côté, plutôt que par la spécificité du geste labial. Il se trouve en effet que cette position côté de la main, étant aussi la position qui permet de coder des consonnes isolées (non insérées dans une syllabe CV), est bien celle qui est concernée lors de l'anticipation vocalique à travers une suite de consonnes (comme dans [ikstRy] par exemple). Pour de telles suites à consonnes multiples, Alloati [10] a montré que l'anticipation de la position de la main existe sur l'anticipation vocalique labiale correspondante tant que le nombre de consonnes codées ne dépasse pas trois ; au-delà la main a tendance à se synchroniser avec les lèvres. Ainsi le geste manuel se coordonne avec le geste labial, soit en anticipant soit en se synchronisant avec lui. C'est en ce sens que nous parlons de main ancrée sur les lèvres.

BIBLIOGRAPHIE

- [1] R.O. Cornett. Cued Speech. *American Annals of the Deaf*, 112:3-13, 1967.
- [2] V. Attina, D. Beautemps., M.-A. Cathiard and M. Odision. Toward an audiovisual synthesizer for Cued Speech : rules for CV French syllables. In *Proceedings of AVSP*, pages 227-232, St Jorioz, France, 4-7 September 2003.
- [3] C. Abry, and M.-T. Lallouache. Modeling lip constriction anticipatory behaviour for rounding in French with the MEM (Movement Expansion Model). In *Proceedings of the 13th International Congress of Phonetic Sciences*, volume 4, pages 152-155, Stockholm, Sweden, 13-19 August 1995.
- [4] V. Attina, M.-A. Cathiard and D. Beautemps. Contrôle de l'anticipation vocalique d'arrondissement en Langage Parlé Complété. In *XXIVèmes J.E.P.*, pages 161-164, Nancy, 24-27 Juin 2002.
- [5] M. Audouy. Logiciel de traitement d'images vidéo pour la détermination de mouvements des lèvres. Projet de fin d'études, option génie logiciel, ENSIMA Grenoble, 2000.
- [6] C. Abry, L.-J. Boë, P. Corsi, R. Descout, M. Gentil and P. Graillot. *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, 1980.
- [7] C. Abry and L.-J. Boë. Laws for lips. *Speech Communication*, 5:97-104, 1986.
- [8] M.-A. Cathiard. La perception visuelle de l'anticipation des gestes vocaliques: coherence des événements audibles et visibles dans le flux de la parole. Thèse de Psychologie Cognitive, Université Grenoble 2, 1994
- [9] R. A. Schmidt. *Motor Control and Learning: A Behavioral Emphasis*. Champaign, Illinois, Human Kinetics Publishers, 1988.
- [10] D. Alloati. Anticipation d'arrondissement et LPC. Mémoire de Maîtrise IDL. Université Stendhal, Grenoble, 2002.

Remerciements. A 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 et à C. Savariaux pour son aide technique. Cette étude est soutenue par un programme Cognitique du Ministère de la Recherche.