Etude perceptive du décours de l'information manuo-faciale en Langue Française Parlée Complétée

Marie-Agnès Cathiard, Florence Bouaouni, Virginie Attina et Denis Beautemps

Institut de la Communication Parlée, INPG/Université Stendhal,
Domaine Universitaire, BP 25, 38040 Grenoble Cedex 9, France
Tél.: ++33 (0)4 76 82 41 28 - Fax: ++33 (0)4 76 82 43 35
Mél: cathiard@icp.inpg.fr

ABSTRACT

In Cued Speech (CS), where the hand is used as an augment to desambiguate facial (mainly labial) information, the question of the influence of the temporal organisation of CS production on the time course of perception has not been adressed. In this study, we will first corroborate the anticipation of the hand over the mouth, a phenomenon which has not been documented except recently in our preceding studies. We found here again an appreciable advance ─ of the size of a fair syllable (i.e. about a quarter of a second) ─, what we have now confirmed for two other coders in addition to the one performing in this study. With a gating procedure, we obtained for 10 deaf cueing subjects the following results. Using CVs inserted in a carrier frame [mytymaCVma], we found that the manual information of hand position, whatever the vowel, and finger configurations, whatever the consonant, were both perceived in advance on the specific identification of the consonant, the specific identification of the vowel being the last one. This perceptual order corresponds to the time course we evidenced for Cued Speech production.

1. INTRODUCTION

La langue française parlée complétée ou LPC, adaptée du Cued Speech (Cornett [1]) est un codage manuel complétant l'information labiale de parole qu'un malentendant peut décoder visuellement. Ce codage, qui repose sur une unité syllabique CV, est composé pour le français de 5 positions de la main autour du visage qui codent les voyelles et de 8 configurations des doigts codant les consonnes (cf. figure 1 pour quelques exemples). Il a été montré de longue date que les malentendants perçoivent correctement par la vision seule cette parole codée (Nicholls & Ling [2]). En revanche la manière dont les informations visuelles – labiales et manuelles – sont intégrées au cours du décodage LPC n'a presque pas été étudiée. Leybaert [3] cite une étude non publiée d'Alegria et al. qui ont testé la perception de présentations congruente et conflictuelle entre main et lèvres : ils montrent que les enfants sourds qui ont été exposés précocement au LPC accordent un poids plus important aux informations manuelles que les sujets exposés plus tardivement. Cette étude ne donne cependant pas d'information sur le déroulement du processus d'intégration perceptive. Nous nous proposons de tester la perception des informations visuelles et manuelles, par un paradigme de gating (Grosjean [4]) qui permet de dévoiler progressivement l'information et de suivre ainsi pas à pas la perception du geste au fur et à mesure de sa production.

2. CORPUS ET ENREGISTREMENT

Les stimuli sont constitués de logatomes du type [mytymaCVma] avec C=[k, p, v, d] et V=[ɔ, ø, ɛ, ɛ̃]. Le choix de ces consonnes et de ces voyelles permet de combiner 2 configurations de doigts et 2 positions de main (figure 1). Les 16 séquences obtenues ont été répétées 3 fois. Notre locutrice est une codeuse professionnelle de 38 ans, qui pratique depuis 9 ans la LPC avec son enfant et en classe.

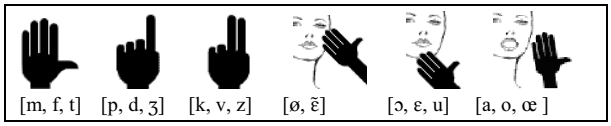

Figure 1 : Configurations de doigts ou clés pour coder les consonnes [m, f, t], [p, d, ʒ] et [k, v, z]; position de main "pommette" pour les voyelles [ø, ɛ̃], "menton" pour [u, ɔ, ɛ] et côté pour [a, o, œ]. La clé avec la main ouverte montrée sur les 3 icônes de droite est celle des consonnes [m, f, t].

La locutrice-codeuse a été enregistrée audiovisuellement à l'aide du poste Visage-Parole de l'ICP (Lallouache [5]), à 25 images/s, par 2 caméras en vue de face, la première filmant le visage et la main, la seconde les lèvres en gros plan. Les lèvres de la locutrice étaient maquillées en bleu pour détecter correctement leurs contours. Une pastille collée sur le dos de la main de la codeuse permettra de suivre les mouvements de la main en coordonnées horizontale (x) et verticale (y) par rapport à une référence sur le verre droit des lunettes portées par la locutrice. Le son est enregistré de façon synchrone avec les vidéos.

3. ANALYSE DE LA COORDINATION MAIN-LEVRES

4 types de signaux sont obtenus (figure 2): le signal acoustique (échantillonné à 22kHz), les trajectoires en x et en y du geste de la main et le décours de l'aire aux lèvres. Pour les trajectoires de la main et des lèvres, l'accélération a été calculée. Différents événements ont

été étiquetés: A1 correspond au début acoustique de la consonne C du logatome (fin de la structure formantique de la voyelle précédente); M1 marque le début du geste de la main pour la position correspondant à la voyelle V et M2 l'atteinte de la position cible; L1 est le début de la configuration labiale pour la voyelle V et L2 l'atteinte de la cible vocalique (les événements M et L sont déterminés aux pics d'accélération et de décélération des gestes).

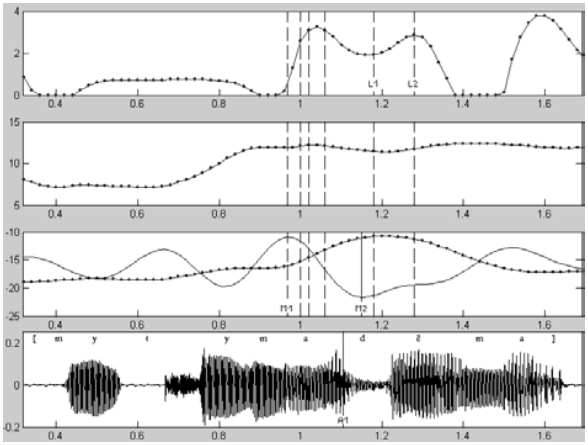

Figure 2 : Exemple d'une séquence : [mytymadɛ̃ma]. La fenêtre du haut représente le décours de l'aire aux lèvres. Les deux fenêtres du milieu présentent la trajectoire des coordonnées x (au dessus) et y (en dessous) du geste de la main. Le passage de la syllabe [ma] (codée avec la main sur le côté) à la syllabe [dɛ̃] (codée avec la main vers la pommette) impliquant un mouvement de main plus marqué verticalement qu'horizontalement, c'est sur la trajectoire en y que l'étiquetage a ici été fait. Pour cette fenêtre, l'accélération est visualisée (ligne fine). La fenêtre du bas donne le signal acoustique correspondant (voir texte pour l'explication des étiquettes A1, M1, M2, L1 et L2). Les 6 lignes pointillées indiquent les 6 points de troncature retenus pour l'expérience perceptive (cf. texte section 4.2.).

A partir de ces événements, nous avons calculé des relations temporelles: M1A1 correspond à la durée entre le début du mouvement de la main et le début de la consonne. Nous avons de même calculé les intervalles A1M2, M1L1 et M2L2. Le patron de coordinations qui se dégage à partir de l'analyse des 42 séquences [mytymaCVma] est le suivant: la main commence à se déplacer depuis sa position sur le côté du visage correspondant à la syllabe [ma], 139 ms (=M1A1; σ = 48) avant le début acoustique de la consonne C. La main atteint sa position M2 pendant la première moitié de la consonne C (A1M2 = 58 ms; σ = 44). En ce qui concerne la coordination main-lèvres, l'écart M1L1 nous indique une avance du début du geste de la main de 259 ms (σ = 113) par rapport au début de la forme labiale; l'écart M2L2 est moins important mais la main atteint néanmoins sa position avec une avance de 155 ms (σ = 111) sur la cible labiale.

Ainsi, au niveau de la production des gestes manuel et labial, nous mettons clairement en évidence une organisation temporelle spécifique, avec un geste de la main qui démarre bien avant la réalisation de la consonne et une atteinte de la position de la main avant la formation de la voyelle aux lèvres. Ce patron de coordinations confirme celui que nous avions précédemment obtenu (Attina et al. [6] [7]), sur un corpus plus étendu produit par la même codeuse et que nous retrouvons par ailleurs pour deux autres codeuses (article en préparation).

4. EXPERIENCE PERCEPTIVE

Comment est traitée cette coordination main-lèvres par le sujet qui perçoit la LPC? Est-il capable de traiter l'information apportée par la main avant l'information portée par les lèvres ou attend-il pour intégrer les deux informations? Autrement dit, un sujet sourd décodant la LPC va-t-il exploiter l'anticipation observée de la main sur les lèvres? Si oui, on devrait observer une identification correcte de la position de la main plus précoce par rapport à l'identification correcte de la voyelle. Nous utiliserons un paradigme de *gating*, éprouvé en perception de la parole [note 1], en tronquant en différents points nos séquences. Cette technique nous permettra d'obtenir une identification pas à pas, au fur et à mesure du dévoilement progressif des informations manuelles et labiales.

4.1. Sélection des séquences

Nous avons retenu une réalisation sur les 3 répétitions de nos 16 séquences, soit celle qui présentait des valeurs d'anticipation les plus proches par rapport aux valeurs moyennes données en 3.

4.2. Choix des points de troncature et montage du test

Nous avons déterminé, à partir des images numérisées puis détramées (1 trame toutes les 20 ms) de chaque séquence, 6 points de troncature (figures 2 et 3) dans le domaine de la syllabe CV pour chacune des 16 séquences. Les séquences tronquées commencent toujours par le début de la phrase porteuse [mytyma] et se terminent à l'image correspondant à chaque point de troncature de la syllabe CV. Pour chaque séquence tronquée, le sujet devra identifier la syllabe CV.

Le 1er point de troncature est au début du mouvement de la main (trame correspondant à notre étiquette M1) pour la syllabe CV. Ce premier point étant le tout début du geste de la main pour coder cette syllabe, la séquence tronquée à ce point ne délivrera aucune information sur la nature de CV : nous prévoyons, pour cette première troncature, que le sujet choisisse la réponse [ma]. Nous aurons ainsi des courbes d'identification démarrant quasiment à 0.

Le 2ème point correspond au début de la formation de la configuration de la main. A ce point, la main n'est plus complètement ouverte (comme elle l'était pour coder la consonne [m] précédente) mais on ne peut pas encore identifier quelle sera la clé suivante. La trame retenue a été sélectionnée par analyse visuelle.

Sur la trame au 3ème point de troncature, la clé est en cours de formation et la main est en train de se déplacer vers la position codant V ("pommette" ou "menton").

Le 4ème point correspond à une trame où la clé et la position sont nettement visibles et identifiables. A ce niveau, la forme labiale de la consonne n'est pas forcément achevée. (les points 2 à 4 précèdent toujours le début acoustique de la consonne, cf. figure 2).

En 5ème point, nous avons pris le point L1 du signal de la trajectoire des lèvres correspondant au début du mouvement des lèvres vers la cible vocalique. La clé est totalement formée, la main a atteint sa cible spatiale, la consonne est identifiable aux lèvres.

Le 6ème point est la trame correspondant à L2 soit à l'atteinte de la cible labiale. A ce point, la syllabe devrait pouvoir être identifiée quelle que soit la séquence.

Figure 3 : Exemple d'un découpage en 6 points de troncature de la syllabe [pɛ̃] pour la séquence [mytymapɛ̃ma] (de gauche à droite et de haut en bas).

Nous avons donc 16 séquences coupées progressivement en 6 points de troncatures, ce qui nous fait un total de 96 films. Les séquences sont présentées, en vision seule (sans le son), avec une interface Matlab, en ordre aléatoire et différent pour chaque sujet. Après visualisation d'une séquence, le sujet doit indiquer la consonne et la voyelle perçues en cliquant sur les boutons réponses à sa disposition ([m, p, v, k, d] et [a, ɔ, ø, ɛ, ɛ̃]. Une phase de familiarisation avec 6 exemples de séquences était proposée avant la passation complète du test.

4.3. Sujets et procédure

10 sujets sourds profonds ont passé le test: 9 adolescents âgés de 11 à 17 ans (âge moyen : 14 ans 11 mois) et une adulte de 35 ans. Tous ont des parents entendants qui codent leur parole et ils bénéficient de codage régulier en milieu scolaire (3-4 heures/jour en moyenne). L'adulte lit parfaitement sur les lèvres mais décode le LPC lors de réunions professionnelles. Ces jeunes ont été rencontrés soit dans leur milieu scolaire soit lors d'une journée organisée par l'association ADIDA réunissant des sourds et leur famille. L'anamnèse pour chaque jeune réalisée auprès des parents a été contrôlée par l'orthophoniste.

Chaque sujet lisait, avant le test, une consigne écrite qui décrivait la tâche et la manière de donner la réponse. L'expérimentateur restait auprès du sujet pendant la phase de familiarisation avec l'interface afin de vérifier que la consigne avait été bien comprise. Puis le sujet était laissé seul pendant la passation.

4.4. Résultats du test

Nous présentons les résultats tous sujets confondus (figure 4). Nous avons calculé à partir des matrices d'identification des stimuli, les pourcentages suivants : % d'identification de la syllabe [ma], % d'identification correcte de la position LPC ("pommette" ou "menton"), % d'identification correcte de la clé LPC (clé [p, d] ou [k, v]), % d'identification correcte de la consonne ([p, d, k, ou v]), % d'identification correcte de la voyelle [ɔ, ø, ɛ, ɛ̃] et enfin % d'identification correcte de la syllabe CV.

Au premier point de gating, les sujets ne peuvent encore rien identifier de la syllabe CV et répondent, comme prévu, majoritairement [ma]. Aux points 2 et 3, les identifications [ma] diminuent et ce sont les scores pour l'identification de la position qui croissent les plus rapidement, suivis de ceux de l'identification de la clé. Au point 4, clé et position (soit l'information LPC sur la consonne et la voyelle) atteignent un pourcentage d'identification élevé (respectivement 87.5% et 92.5%) tandis que l'identification de la syllabe et de la voyelle restent en dessous des 50%. A ce même point 4, l'identification de la consonne est presque à 70%. Ce n'est qu'au point 6 que voyelle et syllabe sont identifiées à plus de 80%. On ne s'étonnera pas que les scores totalement corrects n'atteignent que 80% puisque nous n'avons pas montré toute la tenue de la voyelle, après son climax (notons aussi que ce score est comparable aux 83.5% d'identifications correctes de syllabes CV obtenus par 18 enfants sourds profonds dans l'étude de Nicholls & Ling [2]).

Que se passe-t-il au point 4, là où l'identification [ma] commence à être voisine de zéro ? La différence de scores à cette troncature entre l'identification de la position (et de la clé) *versus* l'identification de la voyelle indique que l'intégration des informations des lèvres et de la main ne se fait pas simultanément : il semble que les sujets exploitent dès que possible les informations de la main puisqu'ils identifient correctement la position de la main pour la voyelle. Ce n'est qu'ensuite que les sujets se servent des informations labiales, pour déterminer une voyelle unique (meilleur score au point 6). Pour la consonne, le choix final, grâce aux informations labiales, se fait plus tôt (point 5). Nous avons donc d'abord un traitement des informations manuelles permettant au sujet d'obtenir un groupe de voyelles et de consonnes

possibles selon la position et la configuration, puis les informations labiales lui permettront de désambiguïser les informations manuelles et de déterminer la syllabe adéquate aux différentes informations visuelles, d'abord pour la consonne ensuite pour la voyelle.

5. CONCLUSION

Nous avons retrouvé, dans cette étude, l'avance systématique de la main dans la production du codage LPC par rapport au mouvement des lèvres que nous avions observé dans d'autres études (Attina et al. [6] [7]). En effet, la main se déplace bien avant le début de la tenue acoustique de la consonne, c'est-à-dire pendant la voyelle précédente. La cible spatiale de la main est atteinte avant la cible vocalique labiale suivante pour la syllabe CV. Notre test de gating montrent que nos sujets sourds tirent profit de cette organisation temporelle spécifique en récupérant perceptivement cette anticipation de la main par rapport aux lèvres. En effet leurs performances indiquent un décodage progressif avec, dans un premier temps, identification par les informations manuelles d'un groupe de consonnes et voyelles possibles, suivi du choix d'un percept unique une fois que les informations labiales pour la consonne et la voyelle sont disponibles. Ce résultat n'est qu'un premier pas dans la compréhension de l'intégration perceptive des deux informations visuelles de la LPC.

Remerciements : Nos remerciements s'adressent à Gladys Brunel, notre codeuse, à Martine Marthouret, orthophoniste au CHU de Grenoble, à l'association ADIDA, au collège des Buclos et au Lycée Louise Michel, aux jeunes sourds et à leur famille. A Christophe Savariaux et Alain Arnal pour leur aide technique. Cette étude est soutenue par un programme Cognitique du Ministère de la Recherche.

[note 1] Les doutes qui avaient été émis sur cette méthode du *gating*, notamment par Ohala & Ohala (1995, In B. Connell & A. Arvaniti, Eds., Papers in Lab. Phonology IV, 41-60) et McQueen (1995, *ib.*, 61-67) n'ont jamais concerné que l'accès lexical (ce qui n'est pas le test réalisé ici). Ils ont d'ailleurs été dissipés depuis (U. Frauenfelder, comm. pers.).

BIBLIOGRAPHIE

[1] R.O. Cornett. Cued Speech. *American Annals of the Deaf*, 112: 3-13, 1967.

[2] D. Nicholls and D. Ling. Cued Speech and the reception of spoken language. *Journal of Speech and Hearing Research*, 25: 262-269, 1982.

[3] J. Leybaert. The role of Cued Speech in language processing by deaf children: An overview. In *Proceedings of the Auditory-Visual Speech Processing (AVSP'O3*), pages 179-186, St Jorioz, France, 4-7 sept. 2003.

[4] F. Grosjean. Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28: 267-283, 1980.

[5] M.-T. Lallouache. *Un poste Visage-Parole couleur. Acquisition et traitement automatique des contours des lèvres*. PhD Thesis. I.N.P. Grenoble, 1991.

[6] V. Attina, D. Beautemps and M.-A. Cathiard. Coordination of hand and orofacial movements for CV sequences in French Cued Speech. In *Proceedings of the International Conference on Spoken Language Processing*, pages 1945-1948, Denver, sept. 2002.

[7] V. Attina, D. Beautemps, M-A. Cathiard and M. Odisio. Toward an audiovisual synthetizer for Cued Speech : Rules for CV French syllables. In *Proceedings of the Auditory-Visual Speech Processing (AVSP'03)*, pages 227-232, St Jorioz, France, 4-7 sept. 2003.

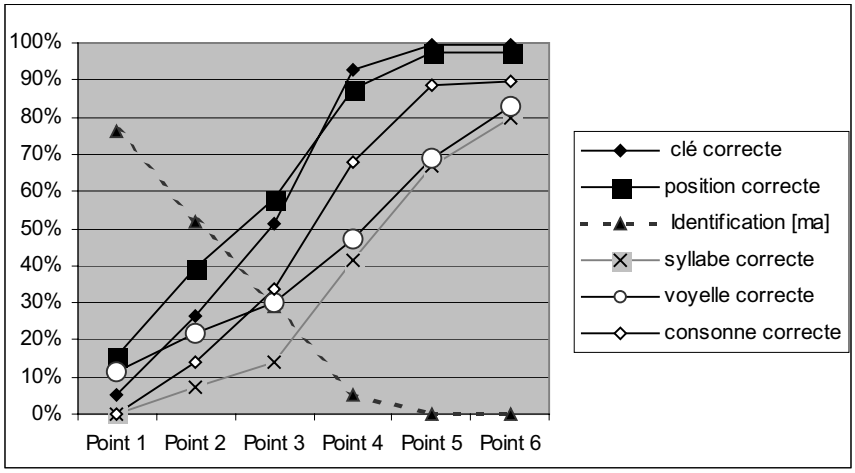

Figure 4 : Résultats d'identification obtenus aux différents points de troncature (voir texte pour l'explication de ces points) par les 10 sujets confondus.