Perception multimodale de l'anticipation en parole chez le sourd et l'entendant, adulte et enfant

Emilie Troille[1,2], Marie-Agnès Cathiard[1], Lucie Ménard[3] & Denis Beautemps[2]

[1] CRI, EA 610, Université Stendhal Grenoble III, emilie.troille@gipsa-lab.inpg.fr, marieagnes.cathiard@u-grenoble3.fr
[2] GIPSA-Lab-ICP, UMR 5216, CNRS-INPG-Université Stendhal Grenoble III, denis.beautemps@gipsa-lab.inpg.fr
[3] Laboratoire de Phonétique, Université du Québec à Montréal, menard.lucie@uqam.ca

ABSTRACT

In a previous study [1] on vowel rounding anticipatory perception through a fricative consonant stream, we demonstrated that auditory information could be, in this VCV frame, ahead of visual information, *and even of audiovisual information*. We propose here to extend this experiment to 7-11-year-old children and to adults, both with normal and impaired hearing. Results showed that, if normal and impaired hearing subjects, both adults and children, had the same late performance in vision only, impaired hearing practicing French Cued Speech succeeded in the identification task *as precociously as normal hearing in the auditory condition*. Hence given the best available information, provided it anticipates on vision, be it the audio channel or manual CS coding, subjects take the best advantage of the natural timing of the different sensory informations. So the hand can substitute *in time* for the spoken sounds.

Keywords: Speech perception, hearing impairment, Cued Speech

1. INTRODUCTION

Nous proposons dans cette étude une extension d'une première recherche présentée aux JEP'06 (Troille et Cathiard [1]) sur la perception de l'anticipation vocalique d'arrondissement par le suivi de la voyelle [y] à travers la consonne [z], particulièrement perméable aux effets de la coarticulation labiale. Nos séquences étaient de type [zizy] et [zizi], cette dernière étant utilisée comme séquence contrôle (fig. 1). Les identifications de la voyelle [y] en condition auditive, visuelle et audiovisuelle ont été comparées grâce à un test de *gating* (dévoilement progressif des informations temporelles) dont la zone couvrait l'ensemble de la consonne [z] jusqu'à [y].

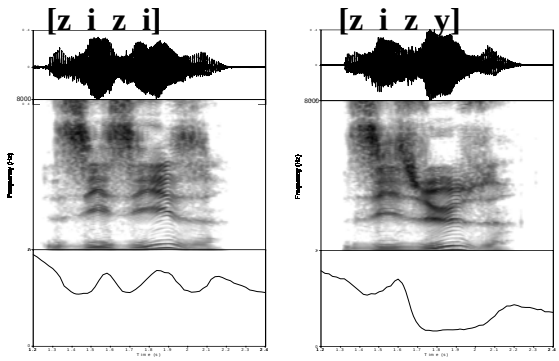


Figure 1 : Signal acoustique et spectrogramme (échelle de fréquence de 0 à 8KHz) avec l'aire aux lèvres (de 0 à 2 cm²) pour ZIZI (à gauche) et ZIZU (à droite)

Nos résultats sur 26 sujets ont démontré un ordre d'anticipation A>AV>V, l'identification auditive, portée par le bruit de friction de la fricative, étant en avance sur l'identification visuelle et *même, dans ce cadre VCV, sur l'identification audiovisuelle*. Ces résultats ont attesté que l'évolution temporelle de l'information bimodale en parole dépendait fortement du timing de la coordination des gestes articulatoires. En effet, le mouvement d'établissement de la voyelle est audible très tôt dans le bruit de friction de la consonne, alors que les changements dans les groupements formantiques vocaliques se produisent plus tardivement.

Nous proposons d'élargir ces résultats, dans un premier temps, à une population d'adultes entendants, ainsi qu'à des enfants entendants âgés de 7 à 11 ans. En effet, peu d'études ont exploré la perception A, V et AV chez les enfants, hormis dans le cadre du paradigme de l'effet McGurk [2]. Nous étudierons dans un second temps ce phénomène d'anticipation vocalique chez des sujets malentendants, adultes et enfants de 7 à 11 ans, en modalité visuelle seule ou avec la Langue française Parlée Complétée (fig. 2). Puis nous comparerons leurs performances à celles des sujets entendants.

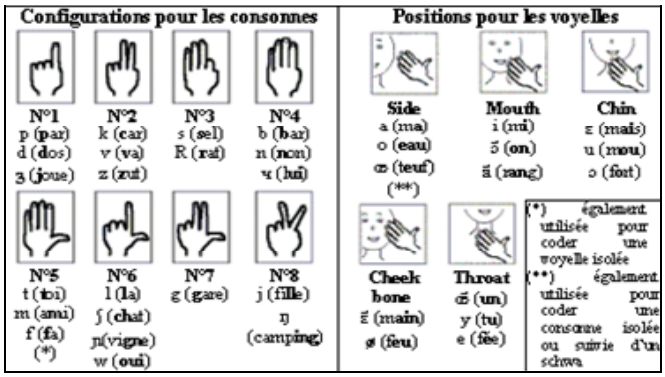


Figure 2 : Les clés en Langue française Parlée Complétée ([4])

La Langue française Parlée Complétée (LPC), adaptée du *Cued Speech* créé en 1967 par le docteur Cornett [3] pour l'anglais américain, est un système de codage manuel artificiel destiné aux sourds pour désambiguïser la lecture labiale. Ce codage permet de délivrer des informations phonologiques précises pour percevoir efficacement l'ensemble de la parole. Le principe du code LPC consiste à convertir la chaîne parlée en syllabes CV et à associer aux formes labiales des clés manuelles, la consonne de la syllabe étant codée par des configurations digitales tandis que la voyelle est indiquée par la position de la main par rapport au visage.

Les études de suivi des mouvements manuels et faciaux lors de la production de ce code par Attina et al. [4] ont mis en évidence une anticipation du mouvement de la main sur celui des lèvres de l'empan d'une bonne syllabe. Ainsi la LPC s'est révélée être en réalité l'inverse d'un langage labial

complété par la main tel que Cornett l'avait conçu. Ce patron d'organisation des gestes étant établi, restait à tester sa prise en compte au cours de la perception de la LPC. Et rappelons que les études sur l'intégration des informations manuelles et faciales en LPC sont très peu nombreuses. Celle d'Alegria et Lechat [5], utilisant des présentations discordantes (paradigme inspiré de l'illusion McGurk) a montré que l'information manuelle même discordante ne peut être ignorée. L'expérience par *gating* de Cathiard et al. [6] a démontré que l'anticipation naturelle de la main était directement récupérée perceptivement par les décodeurs.

2. METHODE

2.1. Données

La méthodologie est celle de notre étude de 2006 [1]. A partir de l'enregistrement audiovisuel en chambre sourde d'un locuteur masculin français, nous avons retenu 2 séquences « T'as dit ZIZU ze ? » [tadizizyz] et « T'as dit ZIZI ze ? » [tadiziziz], afin d'explorer la transition de la voyelle [i] vers la voyelle [y] avec une consonne intervocalique [z].

2.3. Montage des tests

Afin de tester l'identification auditive, visuelle - avec ou sans LPC - et audiovisuelle de l'arrondissement vocalique à travers la consonne, nous utilisons la technique de *gating*, largement connue en perception de parole, qui consiste à dévoiler progressivement les informations temporelles, en ordre aléatoire.

Tests A, V et AV

Dans les 3 conditions de présentation A, V et AV, les séquences commencent toutes au début de « T'as dit... » et se terminent aux différentes dates de la zone de *gating*. Ce domaine d'exploration démarre 40 ms avant la fin du [i] et se termine à la fin de la consonne [z] (soit de 1560 ms à 1720 ms). Le pas du gating étant de 20 ms, nous obtenons 9 séquences tronquées [zizy], auxquelles nous ajoutons 3 séquences tronquées [zizi]. Le découpage audio du domaine perceptif toutes les 20 ms est réalisé sous Praat. Pour le test audiovisuel, nous ajoutons sous Adobe Première Pro 1.5., les images détramées. Pour le test visuel, nous montons les images avec seulement le début de la séquence audio, soit « T'as dit », pour servir d'amorce attentionnelle au sujet.

Test LPC

Le test avec LPC est obtenu à partir du test visuel, auquel nous avons supprimé le son. Il s'agit d'un test LPC de montage : la main n'est pas celle du locuteur, elle a été découpée d'après l'image d'une main codant la clé n°2 pour [z], puis intégrée artificiellement trame par trame à chaque séquence, en prenant soin de reproduire le geste de transition de la main d'une position à une autre, soit, par exemple, de la position-bouche à la position-cou pour [zizy]. Pour effectuer la synchronisation des mouvements labiaux et manuels, nous avons retenu la valeur moyenne d'anticipation de la main sur les lèvres établie par Attina & al. [4], soit 160 ms. La zone de gating adoptée est plus large que dans les tests précédents. Elle s'étend de 1460 ms à 1720 ms, soit depuis 140 ms avant la fin du [i] jusqu'à la fin de [z], pour être sûr d'englober l'ensemble des gestes à la fois manuels et labiaux. Enfin nous avons joint aux séquences [zizy] et [zizi] un montage [zize], la voyelle [e] ayant la même position manuelle que [y] et une forme labiale extrêmement proche de [i]. Ainsi les sujets sourds auront le choix entre 3 voyelles [i], [y] et [e], le choix [e] permettant de contrôler que leurs résultats correspondent bien à une intégration des gestes de la main et des lèvres, et non à la simple détection d'un mouvement de la main vers le cou. Nous obtenons 14 séquences tronquées pour [zizy], auxquelles nous ajoutons 2 séquences tronquées pour [zizi] et pour [zize].

L'ensemble des tests est inséré dans le logiciel Multimédia Toolbook, qui permet le tri aléatoire des séquences présentées aux sujets.

2.4. Sujets et passation

Adultes et enfants entendants

24 adultes entendants de langue maternelle française, sans déficience ni auditive ni visuelle (ou corrigée) sont testés. Le groupe est composé de 21 femmes et de 3 hommes, de 18 à 25 ans (20 ans et 9 mois en moyenne).

Les 18 enfants entendants, sans déficience auditive ou visuelle, sont québécois de langue maternelle française. Ils sont répartis en 14 filles et 4 garçons et ont de 8 à 11 ans (9 ans et 7 mois en moyenne).

Les sujets passaient les 3 tests A, V et AV en ordre aléatoire et devaient identifier la voyelle finale [i] ou [y] de chaque séquence tronquée. Les 9 séquences [zizy] et les 3 séquences [zizi] sont respectivement répétées 10 fois et 5 fois pour les adultes ; chaque séquence est répétée 2 fois seulement pour les enfants, pour éviter des tests trop longs.

Adultes et enfants malentendants

12 adultes sourds sévères ou profonds, 6 femmes et 6 hommes, de 15 à 34 ans (23 ans et un mois de moyenne) ont été testés. 8 d'entres eux sont sourds de naissance, les autres sont devenus sourds avant l'âge de 2 ans et demi (moment du dépistage). L'âge moyen de début d'exposition au LPC est de 3 ans. Il est à 7 mois pour notre sujet le plus jeune, à 10 ans pour le sujet le plus âgé, et à 15 ans pour un sujet qui avait refusé de l'apprendre plus tôt. La durée d'exposition moyenne au code est de 19 ans. Tous ont un bon niveau de décodage.

Les 9 enfants sourds se répartissent en 6 filles et 3 garçons de 7 à 11 ans (8 ans et 10 mois en moyenne). 3 sont français et 6 sont québécois de langue maternelle française. Leur surdité est sévère ou profonde (un sujet a une surdité moyenne). 4 sont sourds de naissance. Ils ont été dépistés entre 18 mois et 5 ans et 11 mois, et leur âge d'exposition au LPC dépend de leur nationalité : les enfants français pratiquent le LPC depuis leur 2 ans, alors que la moyenne est de 5 ans pour les enfants québécois. Ils pratiquent le code depuis 5 ans en moyenne.

La tâche est d'identifier [i], [y] ou [e] en condition visuelle seule ou avec LPC, ces conditions étant contrebalancées. Les 14 séquences [zizy] et les 2 séquences [zizi] et [zize] sont

répétées respectivement 8 fois et 3 fois pour les adultes ; 2 fois pour les enfants.

3. RESULTATS

Nous ne traitons que les réponses [y] dans les séquences [zizy] (les séquences contrôles [zizi] et [zize] ayant toujours été correctement identifiées). Nous rappelons que le début acoustique de la voyelle [y] se situe à 1720 ms. Les frontières d'identification [y] à 50% ainsi que leur comparaison sont déterminées par une analyse Probit (différence significative admise à p<0.01). Pour les adultes, pour lesquels nous avons un nombre élevé de répétitions, nous avons pu aussi calculer des frontières individuelles par Probit et comparer les conditions par une analyse de la variance.

3.1. Résultats des sujets entendants

L'ordre d'anticipation A>AV>V est trouvé pour nos 24 sujets adultes entendants. Une analyse de variance menée avec nos 26 sujets de 2006 et ces 24 nouveaux sujets ne révélant pas d'effet du groupe (F<1), les résultats de l'ensemble de nos 50 sujets peuvent donc être regroupés (fig. 3). La condition de présentation a un effet significatif ($F(2,98)=278$; $p<0,001$). Les frontières à 50% moyennées pour les 50 sujets sont situées à 1628 ms en A, 1645 ms en AV et 1669 ms en V, soit respectivement à 92, 75 et 51 ms avant le début acoustique de la voyelle [y]. La voyelle [y] est donc clairement identifiée dans la consonne [z] qui précède, le plus précocement en A et le plus tardivement en V, la position intermédiaire de la condition AV reflétant bien l'influence des deux modalités auditive et visuelle.

L'ordre d'anticipation A>AV>V est également constaté chez nos 18 enfants entendants (fig. 4). Les frontières d'identification A, AV et V sont respectivement situées à 1651 ms, 1671 ms et 1709 ms. [y] reste donc identifié par les enfants dans la consonne [z], mais avec un retard significatif par rapport aux adultes, et ce dans les trois conditions de présentation : retard de 23 ms en A, 26 ms en AV et 19 ms en V.

3.2. Résultats des sujets malentendants

Nous comparons les courbes d'identification en visuel seul et avec LPC, en ne présentant que les 9 pas correspondants dans les 2 conditions de présentation, depuis 1560 ms jusqu'à 1720 ms.

Pour les adultes (courbes moyennées en fig. 5), la voyelle [y] est identifiée en LPC avec une avance de 64 ms par rapport à la condition visuelle seule : frontière à 1609 ms en LPC et 1675 ms en V (différence significative $F(1,11)=54,839$, $p<0,001$; différence de 33 ms). La comparaison des courbes d'identification individuelles montre que l'ordre d'anticipation LPC>V est retrouvé chez 10 de nos 12 sujets.

Chez les 9 enfants (fig. 6), cette avance en LPC est également significative : [y] est ainsi perçu à 1642 ms en LPC et 1705 ms en V, soit une avance de 64 ms. L'analyse Probit indique une différence significative de 33 ms entre enfants et adultes sourds en condition LPC et une différence non significative en condition visuelle seule.

3.3. Comparaison entendants et malentendants

Si nous comparons les frontières d'identification en modalités V et LPC chez les adultes sourds à celles en modalités A, AV et V chez les adultes entendants, nous constatons tout d'abord que leur performances sont les mêmes en V (1675 ms vs. 1669 ms ; F<1). Nous comparons aussi la frontière LPC et la frontière A, puisque ce sont les deux plus précoces (1609 vs. 1628 ms) sans trouver de différence significative ($F(1,61)=5,520$; $p=0.02$). On notera cependant une courbe d'identification moins catégorielle chez les sourds. Nous obtenons donc pour l'ensemble de nos sujets adultes, l'ordre d'anticipation suivant : LPC=A>AV>V.

La frontière en visuel seul est similaire chez les enfants sourds et entendants (1705 vs. 1709 ms ; différence NS) et la frontière LPC des enfants sourds rejoint celle des enfants entendants en A (1642 ms et 1651 ms ; NS). Notons à nouveau des courbes plus irrégulières chez les sourds. Comme pour les adultes, nous obtenons donc l'ordre d'identification LPC=A>AV>V.

4. DISCUSSION

L'objectif de cette étude était de comparer, avec un paradigme d'anticipation permettant de mettre en concurrence plusieurs modalités, la perception multimodale de la parole chez le sourd et l'entendant, adulte et enfant.

Chez nos 12 adultes malentendants, l'identification visuelle de la voyelle [y] avec le code LPC est nettement plus précoce que l'identification visuelle de la voyelle sans le code. Conformément à l'étude de Cathiard et al. [6], l'avance de la main sur les lèvres, ici recréée artificiellement, est bien utilisée perceptivement par les sujets, comme au naturel. La comparaison des courbes d'identification [i], [e] et [y] en condition LPC permet de constater que le décodage s'opère bien en 2 temps, avec d'abord l'identification d'un groupe de voyelles possibles d'après le mouvement descendant de la main vers le cou (soit [e, y]), puis par le choix d'une de ces voyelles une fois la forme labiale disponible. Ainsi, la figure 7 illustre, pour les adultes, ce passage de la perception [i] jusqu'à 1520 ms à la perception [y] à partir de 1620 ms, avec une zone intermédiaire de près de 100 ms où [e] et [y] sont en concurrence. La même observation est faite chez les enfants sourds (fig. 8), même si les courbes sont nettement plus irrégulières.

La perception des informations dans les 3 modalités A, AV et V pour les enfants entendants, et faciales et manuelles pour les enfants malentendants, apparaît conforme à celle des adultes, avec cependant un retard temporel (jusqu'à 30 ms). Nous pouvons dire que nos enfants, âgés de 7 à 11 ans, n'ont pas encore atteint le développement d'un locuteur adulte. Ce résultat rejoint celui de Sekiyama et Burnham [2], montrant une prise en compte progressive avec l'âge de l'information visuelle.

Les sujets malentendants (adultes et enfants) ne sont pas meilleurs en visuel seul par rapport à leurs pairs entendants, conformément à ce qui est défendu par Bernstein et al. [7]

quand on oppose globalement les deux populations, sans tenir compte d'éventuelles différences individuelles. Notre constat essentiel est que la perception avec LPC chez les sujets sourds peut rivaliser avec la perception la plus en avance chez les entendants, soit, dans le cas de notre étude avec un cadre universellement valide comme VCV avec la perception auditive. Ces résultats indiquent que tous nos sujets, sourds comme entendants, ont tiré le maximum de bénéfice du timing anticipatoire des gestes perçus. Pour les sourds décodeurs de la LPC, la main peut ainsi à temps suppléer au son.

BIBLIOGRAPHIE

[1] E. Troille & M.-A. Cathiard. L'intégration bimodale de l'anticipation du flux vocalique dans le flux consonantique. *XXVIèmes JEP*, Dinard, 367-370, 2006.

[2] K. Sekiyama & D. Burnham. Issues in the Development of Auditory-Visual Speech Perception: Adults, Infants, and Children. In *INTERSPEECH-2004*, 1137-1140, Jeju Island, 2004.

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

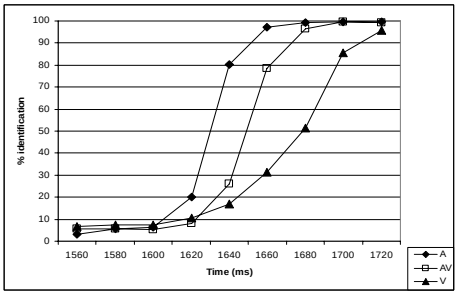
[4] V. Attina, M.-A. Cathiard, D. Beautemps & M. Odisio. A pilot study of temporal organization in Cued Speech production of French syllables : Rules for a Cued Speech synthetizer. *Speech Communication, 44(1-4*), 197-214, 2004.

[5] J. Alegria & J. Lechat. Phonological processing in deaf children : When Lipreading and cues and incongruent. *Journal of Deaf Studies and Deaf Education, 10(2)*, 122-133, 2005.

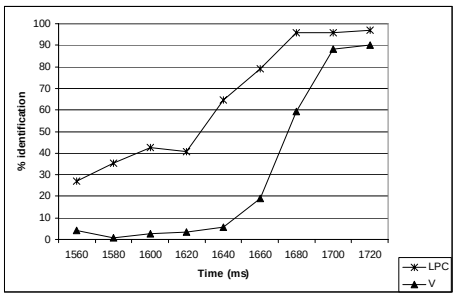
[6] M.-A Cathiard, V. Attina, C. Abry & D. Beautemps. La Langue française Parlée Complétée (LPC) : sa coproduction avec la parole et l'organisation temporelle de sa perception. *Revue Parole, 31/32*, 255-280, 2004.

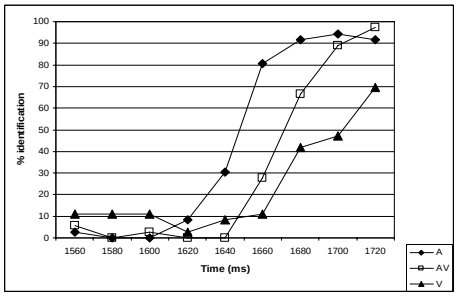
[7] L.E. Bernstein, M.E. Demorest & P.E. Tucker. Speech perception without hearing. *Perception & Psychophysics, 62(2)*, 233-252, 2000.

Remerciements : Ce travail a été soutenu par une bourse de la Région Rhônes Alpes et une bourse de voyage Explora'Doc au Québec. Nous remercions l'école Saint-Jude de Longueuil (Québec) et sa directrice Mme Lyne Gravel qui nous ont accueilli et permis de tester les enfants sourds et entendants. Merci à l'association ADIDA 38 et à Mme Martine Marthouret. Enfin, merci à tous nos sujets, du Québec et de France.

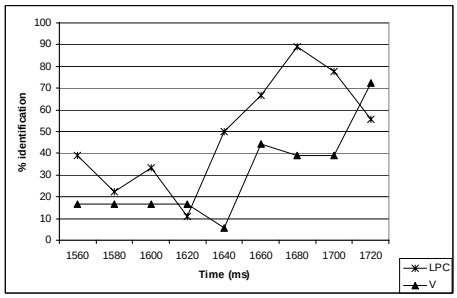


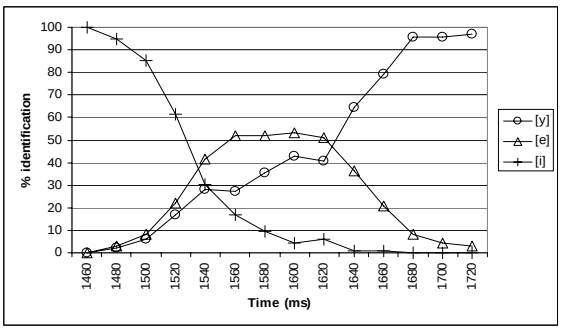
Figures 3 et 4 : Courbes d'identification [y] moyennées en A, AV et V des adultes (en haut) et des enfants entendants (en bas)

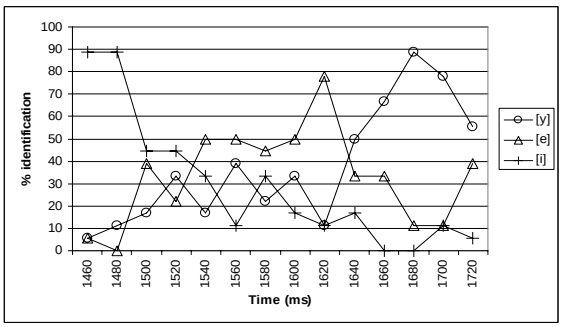




Figures 5 et 6 : Courbes d'identification [y] moyennées en LPC et V des adultes (en haut) et des enfants sourds (en bas)







Figures 7 et 8 : Courbes d'identification [y, e, i] au cours des 14 pas de gating en condition LPC des adultes (à gauche) et des enfants sourds (à droite)