



HAL
open science

La collaboration multisensorielle en parole: Données et Modèles

Denis Beautemps

► **To cite this version:**

Denis Beautemps. La collaboration multisensorielle en parole: Données et Modèles. ASSISTH 2007 - 1ère Conférence internationale sur l'accessibilité et les systèmes de suppléance aux personnes en situation de handicaps, Nov 2007, Toulouse, France. hal-00264643

HAL Id: hal-00264643

<https://hal.science/hal-00264643>

Submitted on 17 Mar 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La collaboration multisensorielle en parole : Données et Modèles

Denis Beautemps

ICP, Département Parole et Cognition de GIPSA-lab, CNRS UMR 5216/UJF/Université Stendhal/INPG, 46 Avenue Félix Viallet, 38031 Grenoble Cedex 1

Denis.Beautemps@gipsa-lab.inpg.fr

RESUME

Cet article vise à définir la place de la vision dans le domaine de la communication parlée, en plaçant le débat dans le cadre de la parole multimodale. Nous illustrerons comment la parole multimodale est au cœur du dispositif de la communication parlée, en montrant que l'audition et la vision sont par nature complémentaires, que la vision peut parfois prendre le relais sur l'audition, que des gestes de la main en appont du mouvement des lèvres peuvent permettre la perception complète de la parole lorsque que l'information auditive n'est pas accessible. Nous rappellerons ainsi que l'information visuelle du visage du locuteur est utilisée pour améliorer la perception de la parole en environnement bruyé, que cette information est importante même en contexte de parole claire mais que même avec de très bonnes performances en lecture labiale, la parole sans connaissance du contexte sémantique reste incomplète. Nous illustrerons par des données issues de la production de parole les ambiguïtés du pattern labial. Et puisque la lecture labiale reste la modalité principale d'accès à la parole pour les personnes sourdes tournées vers l'oral, nous présenterons la méthode de la Langue Française Parlée Complétée, héritée du Cued Speech (CS) [1], et montrerons comment cette modalité manuelle se coordonne avec la parole pour sa perception complète.

Catégorie et termes généraux

Communication parlée; Parole multimodale; Perception et production audio-visuelle.

Mots-clés

Parole audiovisuelle, lecture labiale, surdit , langue Franaise Parl e Compl t e.

1. INTRODUCTION

La parole ne se r duit pas uniquement   du son transmis entre la bouche d'un locuteur et l'oreille de celui qui le reoit. La cha ne de production de la parole est un syst me complexe mettant en  uvre un ensemble d'articulateurs dont certains sont peu visibles car plac s   l'int rieur du conduit vocal, et d'autres visibles tels ceux engendrant des mouvements faciaux principalement les l vres. Des gestes de la main, peuvent aussi venir en appont du mouvement des l vres. C'est le cas de la Langue Franaise Parl e Compl t e. La parole met ainsi en  veil les sens du syst me de perception de la parole qui sait recruter

non seulement l'audition, mais aussi la vision, voire le toucher. C'est donc dans ce sens que la parole est multimodale et multisensorielle. Mais puisqu'apr s tout et pour  tre provocateur, on peut parler et comprendre sans se voir, la multimodalit  ne serait-elle pas un luxe pour la communication parl e, et m me un plaisir gratuit que s'offre le chercheur aux marges de l' tude de la parole sonore et auditive? Nous montrerons au contraire, par mouvements progressifs et convergents, que la multimodalit  est bien au c ur du dispositif de la communication parl e. Ainsi nous illustrerons comment l'audition et la vision sont par nature compl mentaires, que la vision peut parfois prendre le relais sur l'audition, comment des gestes de la main en appont du mouvement des l vres peuvent permettre la perception compl te de la parole lorsque que l'information auditive n'est pas accessible. Quelle est donc la place de la vision dans le domaine de la parole?

2. LA VISION EN RENFORT DE L'AUDITION

2.1 Dans le bruit

S'appuyant sur les travaux pr curseurs de Sumbly & Pollack [2], pour l'Anglais, Beno t et coll gues [3] ont compar  pour le Franais l'intelligibilit  dans le bruit des consonnes $C = [b, v, z, ʒ, ʁ, l]$ et des voyelles $V = [a, i, u]$ plac es   l'int rieur de logatomes de structure $[VCVCVz]$. Le signal audio des logatomes  tait d grad  par du bruit selon six niveaux de rapport signal sur bruit (S/N), pour constituer les stimuli de la condition audio seule A.

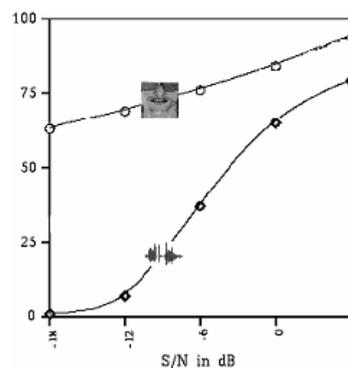


Figure 1. Identification moyenne (18 sujets normo-entendants) en pourcentage pour les stimuli de la condition audio seule A (courbe du bas) et audiovisuelle AV (courbe du haut), en fonction du rapport signal sur bruit.

Les stimuli audio de la condition A ont  t  mont s avec l'image du visage vue de face du locuteur pour constituer les stimuli de la

condition audio-visuelle AV. La figure 1 présente les réponses de 18 sujets normo-entendants pour les deux conditions d'expérience. Les deux courbes montrent une baisse de l'intelligibilité lorsque la proportion de bruit dans l'audio augmente (rapport signal sur bruit de plus en plus négatif) mais avec une meilleure résistance quand l'information visuelle est présente par rapport à la condition audio seule A. Les performances de la conditions AV atteignent un minimum à 60% correspondant à l'information visuelle seule, le niveau de signal sur bruit de -18 dB indiquant que le signal audio totalement bruité n'apporte aucune information. L'information visuelle, immédiatement disponible, récupérée via la lecture labiale, combinée à l'information audio permet donc des performances en identification supérieures à la condition audio seule.

2.2 Sans bruit

La vision est bénéfique même lorsque le son est parfaitement audible. Ceci est montré par le biais de tâches de Shadowing dans lesquelles il est demandé aux sujets de répéter le plus rapidement possible le texte entendu (Reisberg et al., [4]). Lorsque le visage du sujet est présenté en plus du son, les temps de réaction dans la tâche de répétition sont diminués en moyenne de 7,5 % pour des textes français sémantiquement difficiles (La critique de la Raison pure de Kant). Il est observé le même phénomène dans le cas de langues étrangères avec un gain moyen de 20,5 % pour la condition audiovisuelle. Enfin la présentation de l'image vidéo d'une séquence naturelle [aga], montée avec le son naturel de [aba] conduit à l'identification de la séquence /ada/ (effet Mc Gurk, [5]), démontrant ainsi que le système de perception ne privilégie pas plus l'information audio que visuelle, même si celles-ci sont conflictuelles, et qu'au contraire il sait les intégrer. De plus, cette expérience montre que même en étant averti, les sujets sont sensibles à l'illusion et qu'en conclusion on ne peut s'empêcher de voir.

2.3 Effet de l'apprentissage sur les performances en lecture labiale

On vient de le voir, un nombre important de travaux montre que les personnes normo-entendantes peuvent avoir des compétences en lecture labiale sans entraînement spécifique. Il est par contre bien établi que les performances initiales en lecture labiale – sans apprentissage ni entraînement spécifique – varient considérablement d'un individu à l'autre. Ainsi en 1990, McLeod et Summerfield ([6]) ont testé la compétence à lire sur les lèvres de 20 sujets normo-entendants qui ont obtenu des scores variant de 0 à 70 % d'identification correcte de mots dans des phrases. Les auteurs ont montré qu'en situation d'écoute bruitée, ces sujets tiraient profit de la vue du visage du locuteur d'une manière également variable : les sujets bons lecteurs labiaux supportaient – à performance d'identification égale – une augmentation du rapport signal sur bruit de 11dB en condition audiovisuelle par rapport à la condition auditive seule, tandis que les mauvais lecteurs labiaux ne supportaient que 2 dB supplémentaires. Plus récemment, Bernstein et collègues ([7]) ont comparé les performances de 96 normo-entendants et 72 sourds profonds. Ces auteurs retrouvent des performances extrêmement variables selon les individus, dans les deux groupes, mais démontrent clairement que les meilleurs lecteurs labiaux sont sourds. On a souvent dit

que ces bons lecteurs labiaux étaient nés ainsi car, en effet, les expériences d'entraînement à la lecture labiale ont souvent été moins concluantes que ce qui était attendu. Notons que dans les études où les sujets sont entraînés à décoder visuellement des segments de parole (des consonnes par exemple), une amélioration des performances est bien présente (Walden et collègues, [8]). La question reste de savoir si ces effets positifs sont seulement présents à court terme plutôt qu'à long terme (ainsi qu'ont pu le suggérer des auteurs comme Heider et Heider, [9], ou Farwell, [10]). Néanmoins, l'étude de Bernstein et collègues ([7]) semble clairement indiquer que de bonnes performances globales en lecture labiale (sur des phrases par exemple) sont reliées à de bons scores de perception phonétique visuelle (en identification de phonèmes ou de mots isolés). On peut donc inférer de cet ensemble d'informations qu'un entraînement régulier et répété à la lecture labiale doit permettre une amélioration des performances globales, même s'il ne faut peut-être pas espérer combler toutes les différences individuelles de départ.

2.4 Les ambiguïtés de la lecture labiale: Données issues de la production des voyelles

La lecture labiale ne permet pas toujours de percevoir complètement la parole, notamment quand le contexte sémantique est peu prédictible. Une des causes réside dans le fait qu'un certain nombre de phonèmes différents possèdent des formes aux lèvres similaires. C'est le cas par exemple des consonnes [p], [b] et [m] pour lesquelles la distinction sur le trait de voisement n'est pas possible à partir des seules formes labiales. Il en est de même pour les voyelles. Nous rapportons ci-dessous une analyse de la proximité des voyelles produites à partir de la distribution de leurs formes labiales.

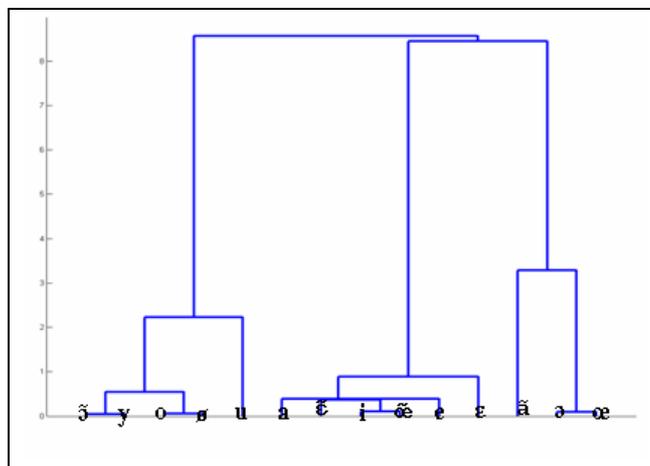


Figure 2. Dendrogramme des 14 voyelles du Français

La Figure 2 montre le résultat de l'analyse en dendrogramme des formes labiales de 1167 voyelles du Français (table 1) extraites d'un corpus de phrases prononcées par un participant. Les formes labiales sont caractérisées à partir de paramètres extraits du contour interne des lèvres : étirement « horizontal » (A) aux lèvres, ouverture « verticale » (B) aux lèvres et aire interlabiale (S). Sur l'axe des abscisses, les 14 voyelles sont présentées. L'axe des ordonnées représente la distance de Mahalanobis calculée à partir de la distribution des paramètres (A,B,S). La hauteur des

branches correspond à la distance entre les voyelles ou groupes de voyelles considérées. La Figure 2 montre clairement trois groupes contrastés de voyelles (visèmes), à l'intérieur desquels les voyelles sont proches en terme de formes labiales: visème 1 (les voyelles antérieures non arrondies [a, ɛ̃, i, œ̃, e, ɛ]), visème 2 (les voyelles arrondies et moyennement arrondies [ɔ̃, y, o, ø, u]), visème 3 (les voyelles peu arrondies [ã, ɔ, œ]). Ce groupement des 14 voyelles en trois visèmes est en conformité avec la description phonétique classique des voyelles et illustre la difficulté pour discriminer les voyelles avec la seule information des lèvres. Ce résultat illustre d'un point de vue « production de parole » les causes de l'ambiguïté de la lecture labiale à laquelle sont confrontées les personnes malentendantes.

| | | | | | | | | | | | | | | |
|-----------------|-----|----|----|----|-----|-----|----|----|----|----|----|----|----|-----|
| | a | o | œ | ɛ̃ | ø | i | ã | ɔ̃ | ɛ | u | ɔ | œ̃ | y | e |
| Effectif | 216 | 63 | 24 | 26 | 110 | 176 | 67 | 41 | 96 | 69 | 32 | 26 | 97 | 124 |

Table 1. Effectif des 14 voyelles étudiées.

3. La Langue Française Parlée Complétée (code LPC) comme exemple de complément de la lecture labiale.

3.1. Le système LPC

La Langue Française Parlée Complétée (LPC) héritée du Cued Speech (CS) (Cornett, [1]) est un code manuel utilisé pour désambigüiser la lecture labiale et ainsi améliorer la perception de la parole par les malentendants et sourds profonds. Dans cette méthode, le locuteur pointe des positions précises sur le côté de son visage ou à la base du cou en présentant de dos des formes de main bien définies. Cinq positions de la main sont utilisées pour coder les voyelles et huit formes de main sont utilisées pour les consonnes (Figure 3). Une même position de la main code plusieurs voyelles, celles pour lesquelles les formes labiales sont bien contrastées. Il en est de même pour les consonnes. Ainsi l'information de la main et de la forme labiale aux lèvres permettent l'identification d'un percept unique. Enfin ce système est syllabique dans le sens où la main pointant une position et présentant une forme de main précise fournit le code de la consonne C et celui de la voyelle V pour la syllabe. L'intérêt de cette méthode réside dans son efficacité d'une part à améliorer la perception de la parole (Nicholls, G., [11] ; Nicholls, G. and Ling, D., [12] pour l'identification des syllabes, Uchanski et collègues, [13] pour l'identification de phrases, taux entre 78 et 97 %), d'autre part à permettre l'accès à une représentation complète du système phonologique pour les malentendants exposés à cette méthode depuis leur plus jeune âge, avec un impact positif sur le développement du langage (Leybaert, J., [14]). Enfin, pour les enfants qui sont porteurs d'implants cochléaires, cette méthode contribue à faciliter l'accès à la modalité auditive.

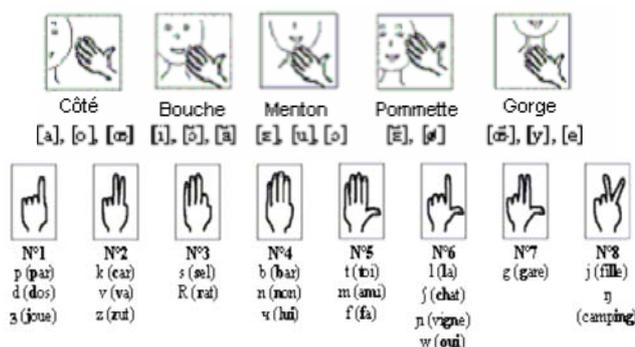


Figure 3. Positions de la main pour les voyelles et formes de mains pour les consonnes (issue de [16]).

3.2 Illustration de la complémentarité du système LPC avec la lecture labiale

Le groupement en trois visèmes des formes labiales des 14 voyelles étudiées dans la section 2.4 est maintenant comparé au système LPC.

| Voyelles | Visème 1 | | | | | Visème 2 | | | | Visème 3 | | |
|-------------------------|----------|---|---|---|---|----------|----|---|---|----------|---|---|
| | a | ø | i | ɛ | œ | o | ɔ̃ | u | y | œ̃ | ã | ɔ |
| Position LPC de la main | | | | | | | | | | | | |
| Côté | X | | | | | X | | | | X | | |
| Pommette | | X | | | | | X | | | | | |
| Bouche | | | X | | | | | X | | | X | |
| Menton | | | | X | | | | | X | | | X |
| Gorge | | | | | X | X | | | | X | | |

Table 2. Voyelles pour chaque visème et positions de main du code LPC associées.

La Table 2 présente les voyelles de chacun des visèmes et le code LPC correspondant en terme de position de main. Chacune des voyelles des visèmes 2 ou 3 est associée à une position de main LPC distincte, ainsi pouvant être discriminée avec l'ajout de la modalité LPC. Pour le visème 1, il apparaît une exception avec la position « Gorge » qui dans le système du LPC est associée à la fois à la voyelle [e] et à la voyelle [œ̃]. Dans ce cas particulier, il est impossible de discriminer ces deux voyelles malgré le code LPC. Mais à part cette exception, ces données sur les voyelles montrent que le code LPC permet de désambigüiser les formes labiales.

4. COORDINATION ORO-FACIALE ET MAIN EN LANGUE FRANÇAISE PARLÉE COMPLÉTÉE

Le fait que les clés manuelles (positions et formes de main) doivent être associées à la forme des lèvres pour que le code LPC soit efficace révèle probablement une véritable coordination entre la main et la mâchoire.

4.1 Etude de la production du code LPC

Aucune étude fondamentale de la production de ce code n'a été réalisée avant les travaux précurseurs d'Attina et collègues ([15]). Cornett ([1]) indique seulement que pour coder certains groupes de consonnes (cluster), la production de la parole doit être retardée pour laisser le temps à la main d'atteindre sa position: "When two consonants precede a vowel, as in the word steep, the first consonant is cued in the base [side] position and the hand moved quickly to the vowel position while the second consonant cue is formed, in synchronization with the lip movements. The lips should assume the position for the first consonant as it is cued, but one should not begin making the sound until the hand is approaching the position in which the contiguous consonant and the following vowel are to be cued. This makes it possible to pronounce the syllable naturally" (Cornett, [1], p. 9). Ceci signifie que l'instruction est clairement d'attendre la fin du mouvement du geste porteur de la voyelle [i] avant de commencer à prononcer le [s]. Excepté cette indication théorique, le problème d'instant de présentation de la clé n'est qu'incidemment évoqué au détour de développements technologiques pour le Cued Speech.

Le code LPC s'appuie sur une organisation syllabique CV, la main donnant conjointement l'information sur la consonne et la voyelle. Le mouvement de la main entre deux positions correspond à la transition vocalique et la mise en forme de la main à l'information consonantique.

Les travaux pionniers dans le domaine et menés à l'Institut de la Communication Parlée (ICP) consistent à déterminer comment le mouvement de la main co-produit l'information sur la consonne et la voyelle en LPC. Pour une première réponse à cette question, Attina et collègues ([15], [16]) se sont focalisés sur l'organisation temporelle des clés manuelles en relation avec le mouvement des lèvres et le signal acoustique correspondant. Dans ces études, à partir de l'analyse d'un codeur LPC diplômé en LPC, Attina et collègues montrent une avance du début du mouvement de la main d'une valeur moyenne de 200 ms sur la réalisation acoustique de la syllabe CV. Par ailleurs les données montrent une superposition complète du geste de formation de la clé digitale sur le geste propre de la main. Dans leurs expériences, les auteurs suivent la position 2D d'un point placé au milieu du dos de la main (Figure 4) à partir des images vidéo d'un enregistrement d'un codeur LPC ainsi que l'aire intérolabiale aux lèvres.

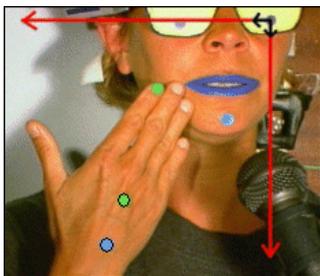


Figure 4. Image du codeur LPC avec les axes x et y en superposition.

Les mouvements de la main sont caractérisés par des transitions lentes entre plateaux selon les trajectoires x et y au cours du temps de la position 2D. Les auteurs délimitent les plateaux par la position du maximum de décélération (M2)

pour le début et le maximum d'accélération (M3) du début de la transition suivante qui marque aussi l'extrémité du plateau le long des axes x et y (Figure 5).

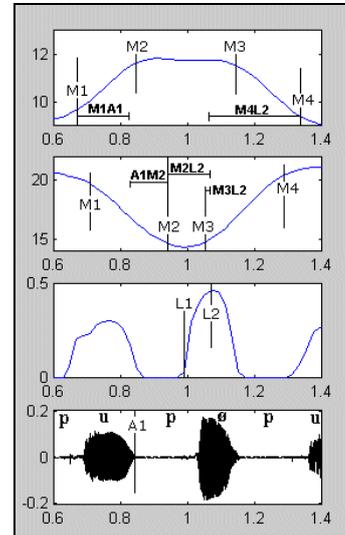


Figure 5. De haut en bas: (1). Trajectoires x (cm) et (2) y (cm) de la main pour une séquence [pupøpu], (3). Décours temporel de l'aire intérolabiale S (cm²); (4). Signal acoustique correspondant, d'après [16].

La cible de la main est ainsi supposée atteinte lorsque x et y atteignent simultanément leur plateau. Utilisant cette analyse cinématique du geste de la main, les auteurs montrent que la main arrive en position cible de manière quasi synchrone avec le début de la réalisation acoustique de la syllabe CV, donc en début de consonne et ainsi bien avant la réalisation acoustique de la voyelle; la main quitte sa position vers une nouvelle cible au niveau du climax vocalique des lèvres (Figure 6).

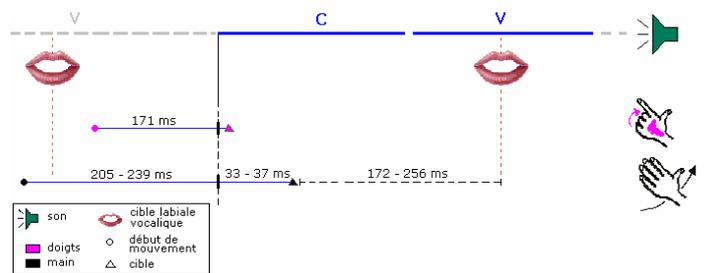


Figure 6. Schéma général de coordination de la main et des lèvres en relation avec le son de parole pour le code LPC, d'après [16].

Les résultats sur l'avance de la main ont été confirmés par l'analyse de la production de trois codeurs supplémentaires (Attina [17]; Attina et al., [18]), ainsi que dans un cadre plus complexe constitué de phrases (Aboutabit et al., [19]).

4.2 Est-ce que l'anticipation de la main est efficacement perçue par les personnes sourdes ?

Les auteurs (Attina et al., [17] ; Cathiard et al., [20]) ont ensuite étudié comment les personnes sourdes profondes, utilisateurs de LPC, traitaient cette coordination temporelle reliant l'information manuelle et labiale. L'hypothèse était que les personnes sourdes traitaient le geste de main avant celui des lèvres. Cette hypothèse serait confirmée si l'on observe : (1) pour l'information de voyelle, un score d'identification correct de la position de main avant celui d'identification de l'information labiale de la voyelle, et (2) pour la consonne, une identification quasi-simultanée de la forme de la main et de l'information aux lèvres.

Le paradigme de "Gating" (Grosjean, [20]), ou de "dévoilement progressif" en Français, a été utilisé pour l'expérience de perception. Cela a permis aux auteurs de tester spécifiquement l'intégration temporelle de l'information manuelle et labiale en tronquant des séquences vidéo de syllabes codées en LPC à différents instants, en fait six points clés comme l'instant d'atteinte de la position cible de la main ou l'instant d'atteinte de la cible aux lèvres. Dans l'expérience, les syllabes étaient composées de séquences Consonne-Voyelle (CV) avec C= [k,v,p,d] et V= [ē, ø, e, o]. Les CV à identifier étaient présentées dans des logatomes de type /mytymaCVma/ comme par exemple /mytymadema/, afin d'éviter la possibilité de prédiction par le contexte sémantique. Pour chacune des vidéos tronquées (présentées en ordre aléatoire), les 16 participants sourds devaient identifier la consonne et la voyelle à partir d'une liste de possibilités. Les résultats montrent que les participants sourds utilisent l'information de main disponible en avance pour anticiper l'information de lèvres et ainsi prédire la voyelle, étant donnée que la position de la main est identifiée correctement avant la cible aux lèvres (l'avance moyenne étant de 155 ms dans cette expérience). De plus, les résultats montrent une identification progressive des stimuli, illustrant ainsi une sélection perceptive d'un sous-ensemble de consonnes et de voyelles à partir de l'information manuelle, puis un choix parmi ces possibilités dès que l'information de lèvre est disponible.

Ainsi ce serait les lèvres qui désambiguïseraient l'information issue de la main. Ces résultats ont servi de référence pour l'élaboration d'un synthétiseur audio-visuel 2D (Attina et al., [16]) puis 3D du code LPC (Gibert et al., [22]).

5. REFERENCES

- [1] Cornett, R.O. Cued Speech, *American Annals of the Deaf*, Vol. 112, pp. 3-13, 1967.
- [2] Sumbly, W.H., Pollack, I. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America* 26 (2), 212-215, 1954.
- [3] Benoit, C., Mohamadi, & Kandel S.. Effect of phonetic context on audio-visual intelligibility of French. *Journal of Speech and Hearing Research*, 37, 11995-1203, 1994.
- [4] Reisberg, D., Mclean, J., Goldfield, A.. Easy to hear but hard to understand: a lipreading advantage with intact auditory stimuli. In: Dodd, R., Campbell, R. (Eds.), *Hearing by Eye : The Psychology of Lipreading*. Lawrence Erlbaum Associates Ltd, Hillside, NJ, pp. 97-113, 1987.
- [5] MacDonald, J., McGurk, H.. Visual influences on speech perception processes. *Perception and Psychophysics* 24, 253-257, 1978.
- [6] MacLeod, A., Summerfield, Q.. A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. *British Journal of Audiology* 24, 29-43, 1990.
- [7] Bernstein, L.E., Demorest, M.E., Tucker, P.E.. Speech perception without hearing. *Perception & Psychophysics* 62(2), 233-252, 2000.
- [8] Walden, B.E., Prosek, R.A., Montgomery, A.A., Scherr, C.K., & Jones, C.I.. Effects of training on the visual recognition of consonants. *Journal of Speech and Hearing Research*, 20,130-145, 1977.
- [9] Heider, F.K., & Heider, G.M.. An experimental investigation of lipreading. *Psychological Monographs*, 52, 124-153, 1940.
- [10] Farwell, R.M.. Speechreading : a research review. *American Annals of the Deaf*, February, 19-30, 1976.
- [11] Nicholls, G.. Cued Speech and the reception of spoken language. Montréal, McGill University: 163, 1979.
- [12] Nicholls, G., Ling, D.. Cued Speech and the reception of spoken language. *Journal of Speech and Hearing Research* 25, 262-269, 1982.
- [13] Uchanski, R.M., Delhorne, L.A., Dix, A.K., Braida, L.D., Reed, C.M., Durlach, N.I.. Automatic speech recognition to aid the hearing impaired: Prospects for the automatic generation of cued speech. *Journal of Rehabilitation Research and Development* 31(1), 20-41, 1994.
- [14] Leybaert, J.. Phonology acquired through the eyes and spelling in deaf children. *Journal of Experimental Child Psychology* 75, 291-318, 2000.
- [15] Attina V., Beautemps D. & Cathiard M.-A.. Coordination of hand and orofacial movements for CV sequences in French Cued Speech. *In proceedings of ICSLP conference*, Denver, USA, pp. 1945-1948, 2002.
- [16] Attina, V., Beautemps, D., Cathiard, M. A. & Odisio, M. A pilot study of temporal organization in Cued Speech production of French syllables: rules for Cued Speech synthesizer, *Speech Communication*, Vol. 44, pp. 197-214, 2004.
- [17] Attina, V., « La Langue Française Parlée Complétée (LPC) : Production et Perception ». Mémoire de thèse, INPG, Grenoble, 2005.
- [18] Attina, V., Cathiard, M.-A. & Beautemps, D.. Temporal measures of hand and speech coordination during French Cued Speech production. In S. Gibet, N. Courty, & J.-F. Kamp (Eds.) *GW 2005, Lecture Notes in Artificial Intelligence*, 3881, pp. 13-24, Springer-Verlag, 2006.
- [19] Aboutabit, N., Beautemps, D., Besacier, L., 2006. Hand and Lips desynchronization analysis in French Cued Speech: Automatic segmentation of Hand flow. In: *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)'06*, Toulouse, 2006.
- [20] Cathiard, M.A., Attina, V., Abry, C. and Beautemps, D.. La Langue française Parlée Complétée (LPC) : sa coproduction avec la parole et l'organisation temporelle de sa perception. *Revue Parole*, N° 29/30/31/32, n° spécial sur « Handicap langagier et recherches cognitives : apports mutuels. J.-L. Nespoulous & J. Virbel (Eds.), Université de Mons-Hainaut, Belgique, 31/32, pp. 255-280, 2004.
- [21] Grosjean, F.. Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28, 267-283, 1980.
- [22] Gibert, G., Bailly, G., Beautemps, D., Elisei, F. & Brun, R. Analysis and synthesis of the 3D movements of the head, face and hands of a speech cuer, *Journal of the Acoustical Society of America*, Vol. 118, No. 2, pp. 1144-1153, 2005.

