

Speech Face Perception is Locked to Anticipation in Speech Production

Emilie Troille

CRI, EA 610, Université Stendhal Grenoble III
GIPSA-Lab-ICP, UMR 5216, CNRS-INPG-Université Stendhal Grenoble III
emilie.troille@gipsa-lab.inpg.fr

At the beginning of the 90s, it was definitively demonstrated that as early as the visual speech information is perceivable, speech identification can be processed. Cathiard & *al.* (1991; see also Cathiard, 1994, Cathiard & *al.*, 1996) used different V-to-V anticipatory spans, with articulatory measurements, along silent pauses, in a perceptual gating paradigm, and found that up to 200 ms **"speech can be seen before it is heard"**. The available lip image processing system (Lallouache, 1991) allowed to evidence that speech face perception was fairly locked to anticipation in production. Accordingly these results could be framed in the framework of a general anticipatory control model, the *Movement Expansion Model* (for the link with perception, see Abry & *al.*, 1996; for production testing of the model, since Abry & Lallouache, 1995a, b, for French, see recently Noiray & *al.*, 2006, 2008, for French children and English).

It is important to note that the more classical CVCV phonetic span remained until now poorly understood as concerns the audiovisual production-perception phenomenology, more specifically, the coordination of the vowel and consonant bimodal streams. Smeele & *al.* (1994; Smeele, 1994) in a pure perceptual gating experiment (with no articulatory measurements) assessed, plainly, that the most visible places for consonants, bilabials and labiodentals, were naturally better identified when vision was added, just along two 40 ms steps in the constriction phase before the release. To our knowledge the first preliminary attempt to measure the time course of audio and visual perception in CVCVs, with the control tracking of lip gestures, was published by Escudier, Benoît & Lallouache (1990). The test stimulus was [zizy], *i.e.*: (i) with a vowel-to-vowel rounding gesture, (ii) throughout a fricative voiced consonant, this (iia) in order to interrupt as less as possible the acoustic flow, (iib) while offering a sufficiently high frequency frication noise, which could carry the resonance changes corresponding to the lip rounding gesture, above the range of the formants characterizing the change in the vowels. They found roughly (apart from methodological problems they acknowledged) that the visible rounding anticipatory gesture was perceived 40-60 ms ahead of the acoustic change.

This is the paradigm we took up again more than ten years later. And for the first time in this research domain we found, repeatedly, that **"speech can be heard before it is seen"** (Troille & *al.*, 2007).

The main purpose of the present contribution will be to clear up apparent contradictions, essentially due to misconceptions of variance and lawfulness in speakers' behavior. A variability which perceivers have to cope with, since we demonstrated that with their ears and/or their eyes, be they aided or not, they succeed in recovering in due time the recoverable linguistic information in the modality or modalities they have access to (Troille & *al.*, 2008). Supporting this conception, an additional type of behavior (compared to Escudier & *al.*, 1990 and Troille & *al.*, 2008) will be presented.

REFERENCES

- Abry C., & Lallouache, T. (1995a). Le MEM : Un modèle d'anticipation paramétrable par locuteur. Données sur l'arrondissement en français. *Bulletin de la Communication Parlée*, 3, 85-99.
- Abry, C., & Lallouache, T.M. (1995b). Modeling lip constriction anticipatory behaviour for rounding in French with the MEM. In *Proceedings of ICPHS*, 4, Stockholm, Suède, 152-155.
- Abry, C., Lallouache, M.-T., & Cathiard, M.-A. (1996). How can coarticulation models account for speech sensitivity to audio-visual desynchronization? In D. Stork & M. Hennecke (Eds.), *Speechreading by Humans and Machines*, NATO ASI Series F: Computer and Systems Sciences, vol. 150, pp. 247-255, Springer-Verlag, Berlin Heidelberg New York London Paris Tokyo.
- Cathiard, M.-A., Tiberghien, G., Tseva, A., Lallouache, M.-T., & Escudier, P. (1991). Visual perception of anticipatory rounding during acoustic pauses: A cross-language study. In *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 19-24 Août 1991, Aix-en-Provence, France, 4, 50-53.

- Cathiard, M.-A. (1994). *La perception visuelle de l'anticipation des gestes vocaliques : cohérence des événements audibles et visibles dans le flux de la parole*. Thèse de Psychologie Cognitive, Grenoble.
- Cathiard, M.-A., Lallouache, M.-T., & Abry, C. (1996). Does movement on the lips mean movement in the mind? In D. Stork & M. Hennecke (Eds.), *Speechreading by Humans and Machines*, NATO ASI Series F: Computer and Systems Sciences, vol. 150, pp. 211-219, Springer-Verlag, Berlin Heidelberg New York London Paris Tokyo.
- Escudier, P., Benoît, C., & Lallouache, M.-T. (1990). Identification visuelle de stimuli associés à l'opposition /i/-/y/ : étude statique. *Colloque de physique, supplément au n° 2, tome 51*, 1er Congrès Français d'Acoustique, C2-541-544.
- Lallouache, M.-T. (1991). *Un poste «Visage-parole » couleur. Acquisition et traitement automatique des contours des lèvres*. Thèse. ENSERG, Grenoble.
- Noiray, A., Ménard, L., Cathiard, M.-A., Abry, C., Aubin, J. & Savariaux, C. (2006). Extending the Movement Expansion Model (MEM) for rounding from French to English. *Proceedings of the 7th International Seminar on Speech Production*, Ubatuba, 319-326.
- Noiray A., Ménard L., Cathiard M.-A., Abry C. & Savariaux C. (2008). Emergence of a vocalic gesture control: The tuning of the anticipatory rounding temporal pattern for French children. In Kern, S., Gayraud, F. et Marsico, E. (eds), *Emergence of Linguistic Abilities*, Cambridge Scholars Publishing : New Castle, 100-116.
- Smeele, P. M. T., Sittig, A. C., & Van Heuven, V. J. (1994). Temporal organization of bimodal speech information. In *International Conference on Spoken Language Processing*, Vol. 3, pp. 1431-1434.
- Smeele P.M.T. (1994) *Perceiving speech: Integrating auditory and visual speech*. Doct Dissertation, Techn Univ Delft, Delft.
- Troille, E., Cathiard, M.-A. & Abry, C. (2007). Consequences on bimodal perception of the timing of the consonant and vowel audiovisual flows. *Proceedings of the International Conference on Audio-Visual Speech Processing*, 281-286, 31 august - 3 September, Hilvarenbeek, Pays-Bas.
- Troille, E., Cathiard, M.-A, Abry, C., Ménard L. & Beautemps, D. (2008). Multimodal perception of anticipatory behaviour. Comparing blind, hearing and Cued Speech subjects. *International Conference on Auditory-Visual Speech Processing*, Tangalooma, Australia, 26-29 September.