Speech in the mirror? Neurobiological correlates of self-speech perception

Avril Treille, Coriandre Emmanuel Vilain, Sonia Kandel, Jean-Luc Schwartz, Marc Sato

SPEECH IN THE MIRROR?
NEUROBIOLOGICAL CORRELATES OF SELF-SPEECH PERCEPTION

Avril Treille[1], Coriandre Vilain[1], Sonia Kandel[1], Jean-Luc Schwartz[1] & Marc Sato[2]
[1] GIPSA-lab, Département Parole & Cognition, CNRS & Grenoble Université, Grenoble, France
[2] Laboratoire Parole & Langage, CNRS & Aix Marseille Université, France

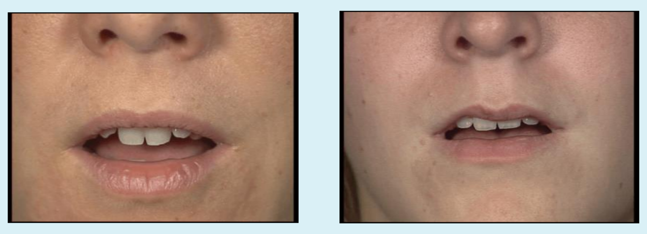
Self-awareness and self-recognition during action observation may partly result from a functional matching between action and perception systems. This perception-action interaction is thought to enhance the integration between sensory inputs and our own sensory-motor knowledge.

We present a combined EEG and fMRI study that examines the impact of self-knowledge on multisensory integration mechanisms during auditory, visual and audio-visual speech perception. Our working hypothesis was that hearing and/or viewing oneself talk might facilitate the bimodal integration process and activate sensory-motor plans to a greater degree than does observing others.

METHODS

*The same **stimuli** were used for both experiments :*

- **Syllables** : /pa/, /ta/, /ka/
- **Modalities** : auditory **(A)**, visual **(V)**, audio-visual **(AV)** and incongruent audio-visual (**AVi**, **self** auditory signal dubbing **other** visual signal)
- Half of the stimuli were related to the participant (**self** condition), the other half to an unknown speaker (**other** condition).
- A total of 1176 stimuli were created



Example of self (left) and other (right) stimuli for one subject

- **Subjects**:
  - EEG: 18 healthy adults, right-handed native French speakers.
  - fMRI: 12 healthy adults, right-handed native French speakers.

EEG

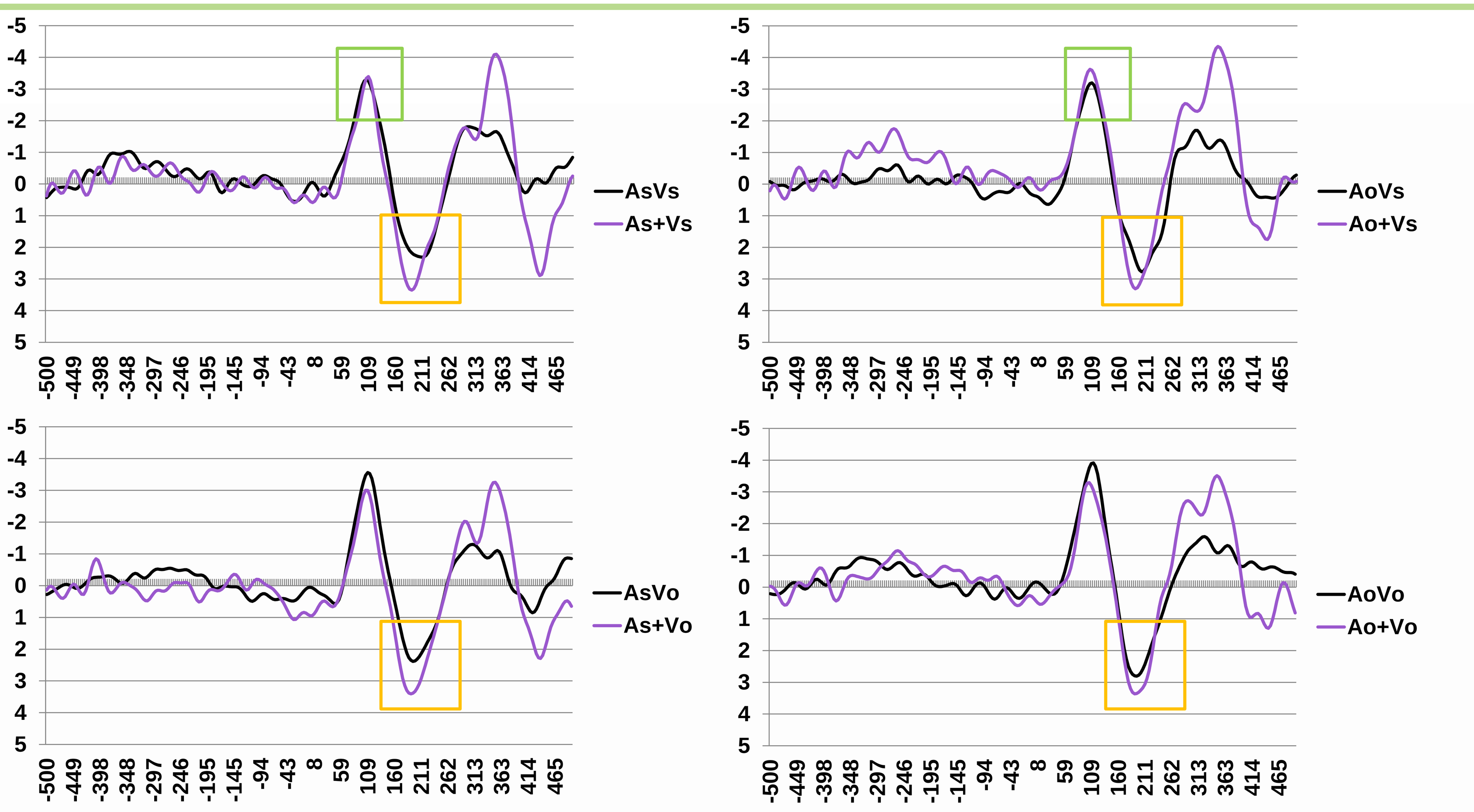
- **Task**: three-alternative forced-choice identification task, with participants instructed to categorize each perceived syllable with their left hand, after an audio "beep".
- **Data acquisition**: EEG data were continuously recorded from 64 scalp electrodes (international 10–20 system) using the Biosemi EEG system operating at a sampling rate of 256 Hz. External reference electrode was at the top of the nose. The electrooculograms controlling for horizontal (HEOG) and vertical (VEOG) eye movements were recorded using electrodes at the outer canthus of each eye as well as above and below the right eye.
- **Data pre-processing : EEG-Lab**
  (1) Re-referenced off-line to the nose;
  (2) Filtering: 2-30 Hz;
  (3) Epoching: 1000ms (baseline from -500 to -400ms to the acoustic syllable onset);
  (4) Artifact Rejection: ±60 µV
- **Data analyses**:

	Amplitude	Latency
Speaker's effect	ANOVA : Auditory modality (Self/other), Visual modality (Self/other/None)	ANOVA : Auditory modality (Self/other), Visual modality (Self/other/None)
Audio-visual Integration	ANOVA : Signal type (Bimodal/Sum), Auditory modality (Self/Other), Visual modality (Self/Other)	ANOVA : Signal type (Bimodal/Sum), Auditory modality (Self/Other), Visual modality (Self/Other)

fMRI

- **Task**: passive listening and/or viewing of **A**, **V**, **AV** or **Avi** /pa/, /ta/, and /ka/ syllables (+ a resting face of the participant of the unknown speaker serving as baseline). The stimuli were presented with (−6 dB SNR) or without acoustic noise.
- **Data recording**: 3T whole-body MR scanner (Philips Achieva TX) using a sparse sampling acquisition in order to minimize scanner noise during speech perception (53 axial slices, 3 mm3; TR = 8 sec, acquisition from the stimulus onset: 5s). T1-weighted whole-brain structural image for each participant after the last functional run (MP-RAGE, sagittal volume of 256 x 224 x 176mm3 with a 1 mm isotropic resolution).
- **Data pre-processing**: SPM8.
  (1) rigid realignment of each EPI volume to the first of the session,
  (2) coregistration of the structural image to the mean EPI,
  (3) normalization of the structural image to common subject space (with a subsequent affine registration to MNI space) using the group-wise DARTEL registration method,
  (4) warping of all functional volumes using deformation flow fields generated from normalization step,
  (5) affine registration for transformation into the Montreal Neurological Institute (MNI) space ,
  (6) spatially smoothing with a three-dimensional Gaussian kernel with a full width at half-maximum of 9 mm.
- **Data analyses**:
  - 1st level: GLM: 16 regressors of interest (4 modalities (A, V, AV, AVi) x 2 speakers (self, other) x 2 noise (with, without) and six realignment parameters as nuisance regressors, with the 4 corresponding baselines (2 speakers x 2 noise levels) forming an implicit baseline. The BOLD response for each event was modeled using a single-bin finite impulse response (FIR) basis function spanning the time of acquisition (3s) and a high-pass filtering with a cutoff period of 128s was applied.
  - 2nd level: group effect: modality (4 levels: A, V, AV, AVi), the speaker (2 levels: self, other) and the noise (2 levels, without, with) as within-subject factors and the subjects treated as a random factor.

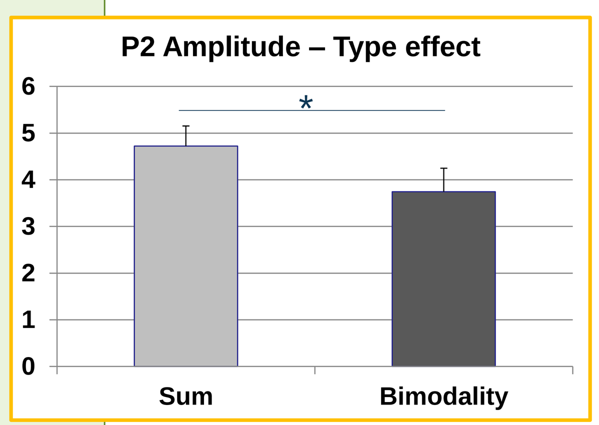
1 EEG



Averaged EEG signal on fronto-central electrodes for bimodal conditions (AsVs, AoVs, AsVo, AoVo) compared to the sum of unimodal conditions (As+Vs, Ao+Vs, As+Vo, Ao+Vo); **s:** self, **o:** other

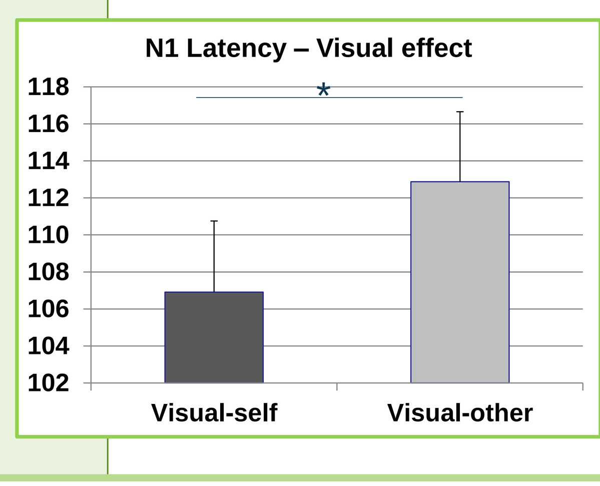
Integration (AV <> A+V)

- P2 amplitude : AV < A+V (p<. 02) => integration



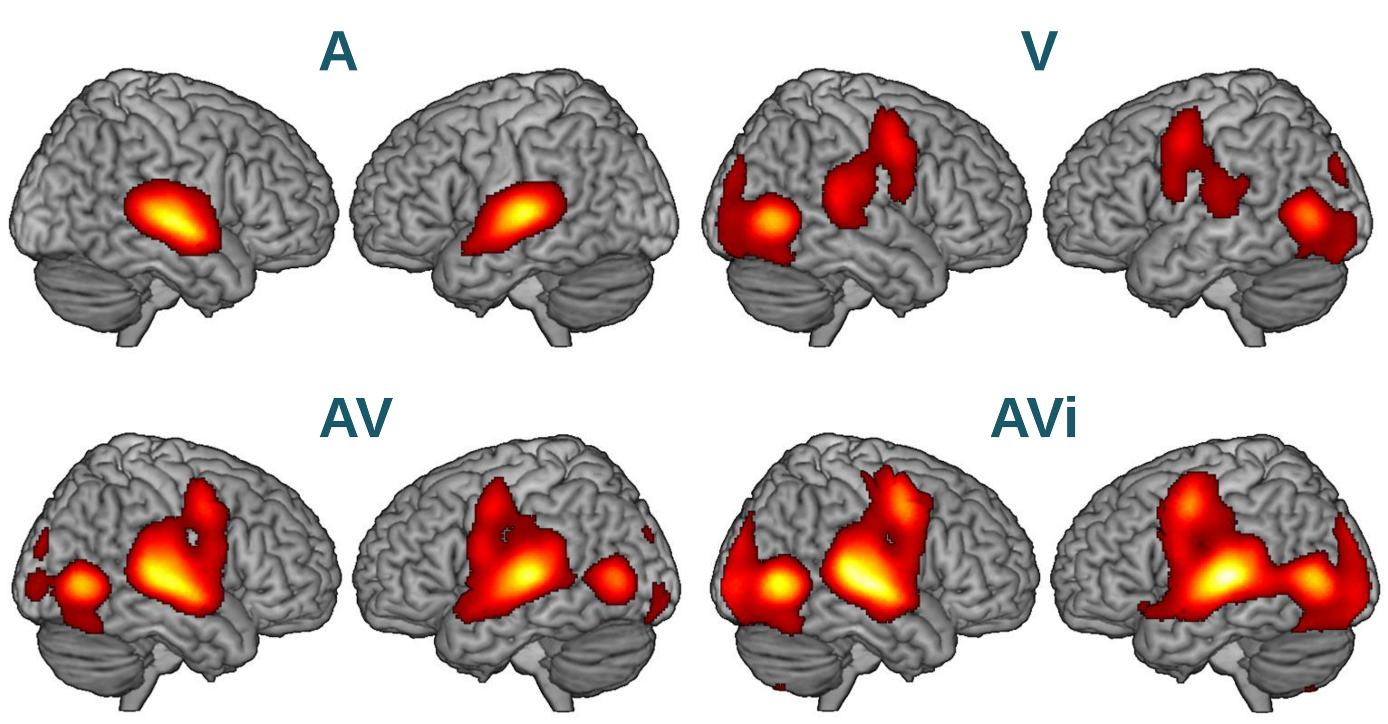
Self effect on integration

⇒ Visual-Self : reduced N1 latency (p<.02)



2 fMRI

Results – Modality



Single Effects (T contrasts, p < .05 FWE corrected): brain activity observed in each modality compared to baseline, irrespective of the speaker.

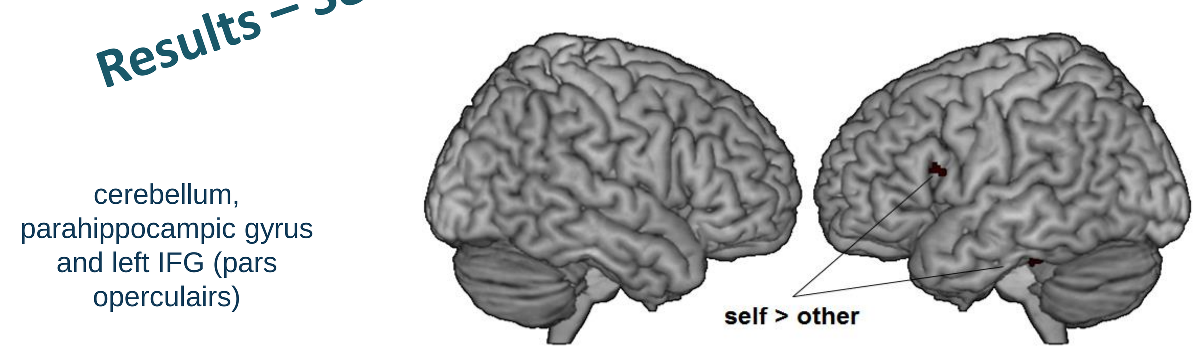
Modality results :

- **Auditory regions** : stronger activity for the auditory condition than the visual only condition
- **Visual regions** : stronger activity for the visual condition than for the auditory only condition
- Greater Activity of the dorsal part of the **premotor cortex** for visual stimuli (no activation for the auditory only condition)

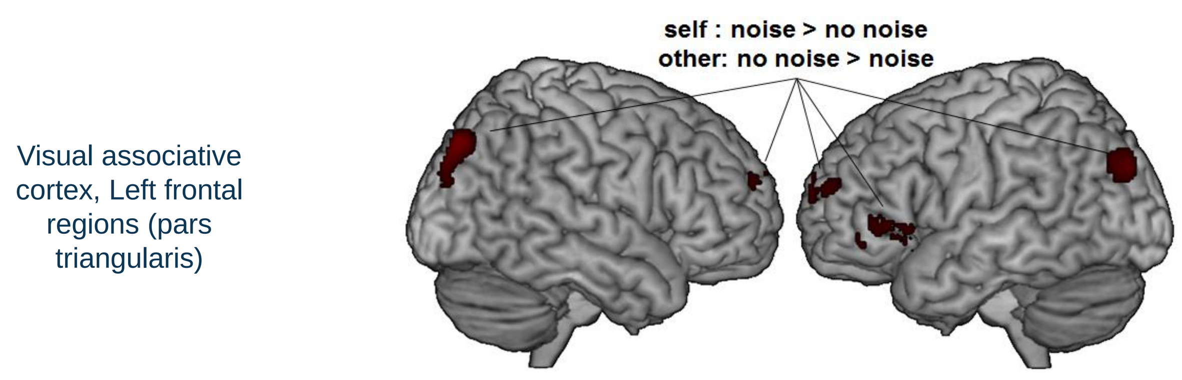
Self effect :

- Stronger activity of the cerebellum, the parahippocampic gyrus and the left inferior frontal gyrus (pars opercularis)
- Small effect but we'll test more subjects

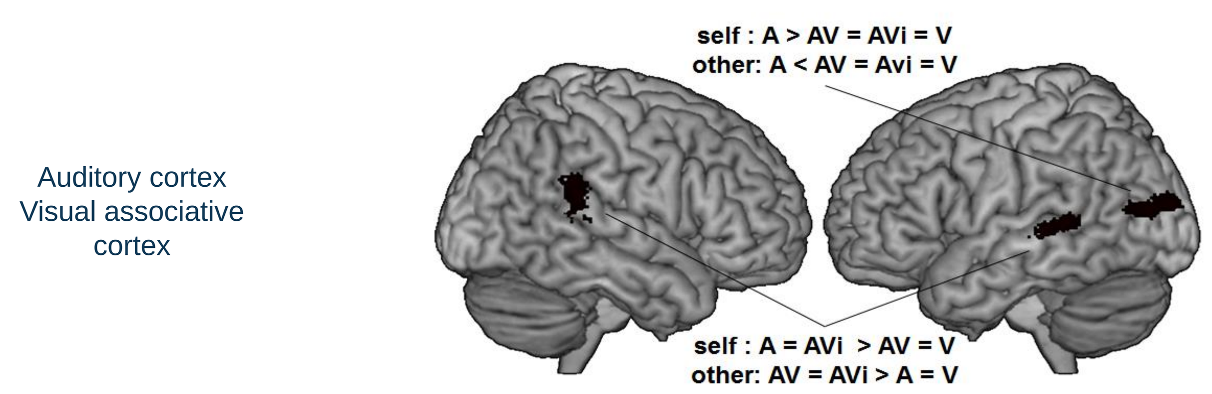
Results – Self



Main Effect of Self (F contrast, p < .001 unc)



Self x Noise Interaction (F contrast, p < .001 unc)



Self x Modality Interaction (F contrast, p < .001 unc)

(1) a) In line with previous studies on multimodal speech perception => integration mechanisms of auditory and visual speech signals.
b) A visual processing advantage when the perceptual situation involves our own speech production.
(2) a) Global coherent activations of the single effects during auditory, visual and audio-visual speech perception.
b) hearing and/or viewing oneself talk increased activation in the left posterior inferior frontal gyrus (pars opercularis) and cerebellum.
➔These regions are generally responsible for predicting sensory outcomes of action generation.

Altogether, these results suggest that viewing our own utterances leads to a temporal facilitation of auditory and visual speech integration and processing afferent and efferent signals in sensory-motor areas gives rise to self-awareness during speech perception.

Part of this research was supported by a grant from the European Research Council (FP7/2007-2013 Grant Agreement no. 339152, "Speech Unit(e)s")
Correspondence: avril.treille@gipsa-lab.grenoble-inp.fr