

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

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

Marc Sato

#### ▶ To cite this version:

Avril Treille, Coriandre Emmanuel Vilain, Sonia Kandel, Jean-Luc Schwartz, Marc Sato. Speech in the mirror? Neurobiological correlates of self speech perception. SNL 2015 - 7th Annual Society for the Neurobiology of Language Conference, Oct 2015, Chicago, United States. hal-01297700

### HAL Id: hal-01297700 https://hal.science/hal-01297700

Submitted on 5 Apr 2016

**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.



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

*Avril Treille<sup>1</sup>, Coriandre Vilain<sup>1</sup>, Sonia Kandel<sup>1</sup>, Jean-Luc Schwartz<sup>1</sup> & Marc Sato<sup>2</sup>* <sup>1</sup> GIPSA-lab, Département Parole & Cognition, CNRS & Grenoble Université, Grenoble, France <sup>2</sup>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.

The same **stimuli** were used for both experiments : Task: passive listening and/or viewing of A, V, AV or Avi /pa/, /ta/, and /ka/ syllables (+ a resting face of the **Task**: three-alternative forced-choice identification task, with participants instructed participant of the unknown speaker serving as baseline). The stimuli were presented with (-6 dB SNR) or to categorize each perceived syllable with their left hand, after an audio "beep". without acoustic noise. Data acquisition: EEG data were continuously recorded from 64 scalp electrodes Syllables : /pa/, /ta/, /ka/ Data recording: 3T whole-body MR scanner (Philips Achieva TX) using a sparse sampling acquisition in order to (international 10–20 system) using the Biosemi EEG system operating at a sampling Modalities : auditory (A), visual (V), audio-visual (AV) and incongruent audiorate of 256 Hz. External reference electrode was at the top of the nose. The minimize scanner noise during speech perception (53 axial slices, 3 mm3; TR = 8 sec, acquisition from the visual (AVi, self auditory signal dubbing other visual signal) electrooculograms controlling for horizontal (HEOG) and vertical (VEOG) eye stimulus onset: 5s). T1-weighted whole-brain structural image for each participant after the last functional run Half of the stimuli were related to the participant (self condition), the other (MP-RAGE, sagittal volume of 256 x 224 x 176mm3 with a 1 mm isotropic resolution). movements were recorded using electrodes at the outer canthus of each eye as () half to an unknown speaker (other condition). well as above and below the right eye. **DOH** Data pre-processing: SPM8. A total of 1176 stimuli were created (1) rigid realignment of each EPI volume to the first of the session, Data pre-processing : EEG-Lab (2) coregistration of the structural image to the mean EPI, (1) Re-referenced off-line to the nose; (3) normalization of the structural image to common subject space (with a subsequent affine registration to **M** (2) Filtering: 2-30 Hz;



**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.

(3) Epoching: 1000ms (baseline from -500 to -400ms to the acoustic syllable	
onset);	

(4) Artifact Rejection:  $\pm 60 \mu V$ 

Data analyses:

ш

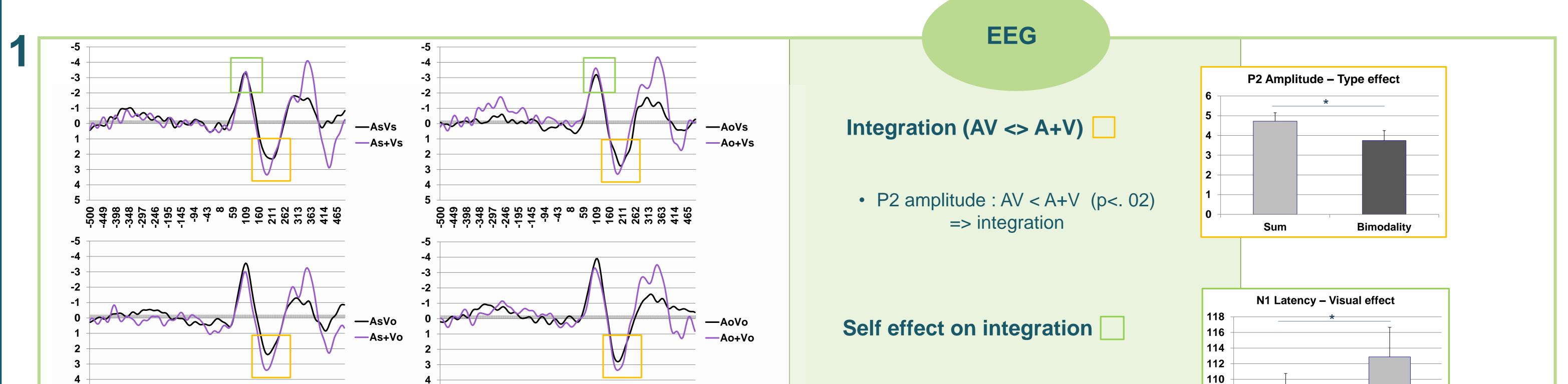
	Amplitude	Latency
Speaker's effect	ANOVA : Auditory modality (Self/other), Visual modality (Self/other/None)	ANOVA : Auditory modality (Self/other), Visual modality (Self/other/None)
Audio-visual Integration	ANOVA : Signal type (Bimodal/Sum), Auditory modality (Self/Other), Visual modality (Self/Other)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

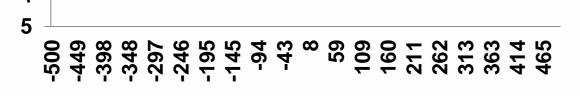
MNI space) using the group-wise DARTEL registration method,

(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:

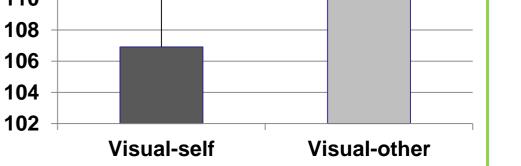
- 1<sup>st</sup> level: GLM: 16 regressors of interest (4 modalities (A, V, AV, AVi) x 2 speakers (self, other) x 2 noise (with, without) and six realignment parameters as nuisance regressors, with the 4 corresponding baselines (2 speakers x 2 noise levels) forming an implicit baseline. The BOLD response 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.
- 2<sup>nd</sup> 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.

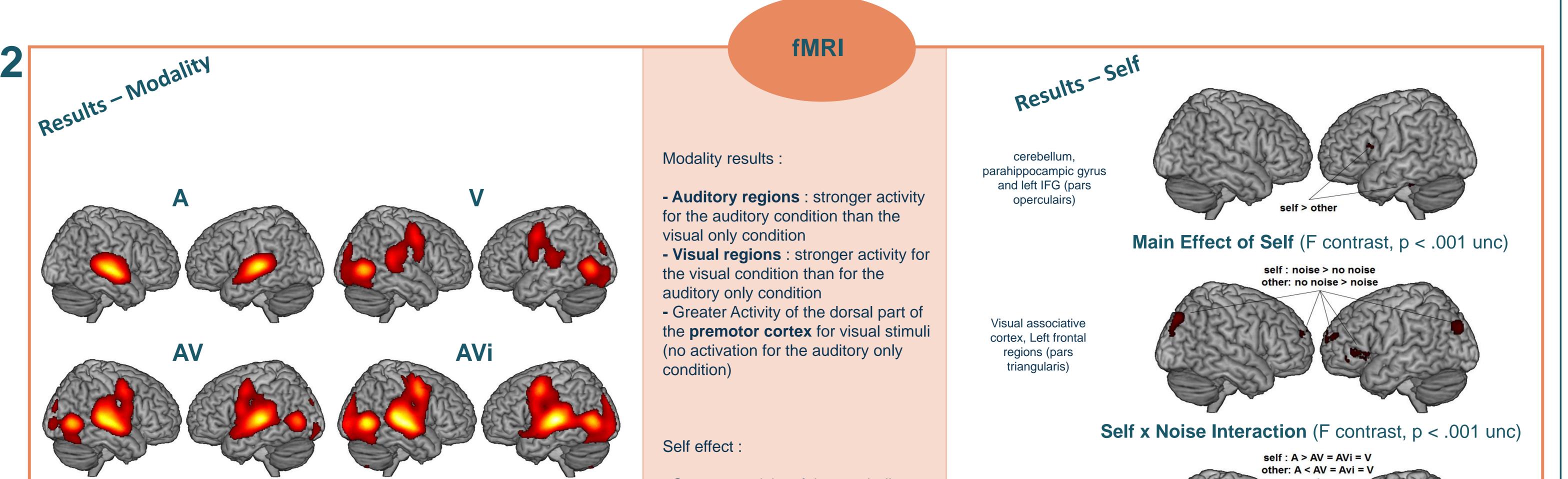




Averaged EEG signal on fronto-central electrodes for bimodal conditions (AsVs, AoVs, AsVo, AoVo) compared to the sum of unimodal conditions (As+Vs, Ao+Vs, As+Vo, Ao+Vo); s: self, o: other

 $\Rightarrow$  Visual-Self : reduced N1 latency (p<.02)

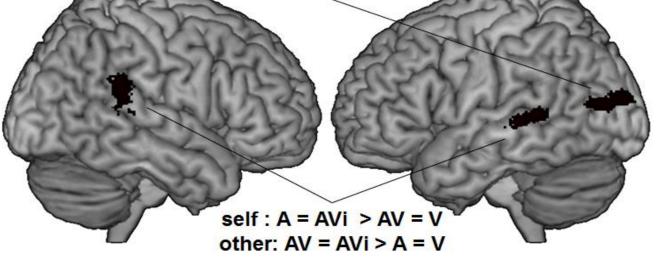




**Single Effects** (T contrasts, p < .05 FWE corrected): brain activity observed in each modality compared to baseline, irrespective of the speaker.

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

Auditory cortex Visual associative cortex



### **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