Through the looking-glass: Neural basis of self representation during speech perception A V AV AVi AV AVi Results -Self

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

Through the looking-glass:
Neural basis of self representation during speech perception

Marc Sato[1], Avril Treille[2], Coriandre Vilain[2], Jean-Luc Schwartz[2]

[1] Speech and Language Laboratory & CNRS, Aix-en-Provence, France

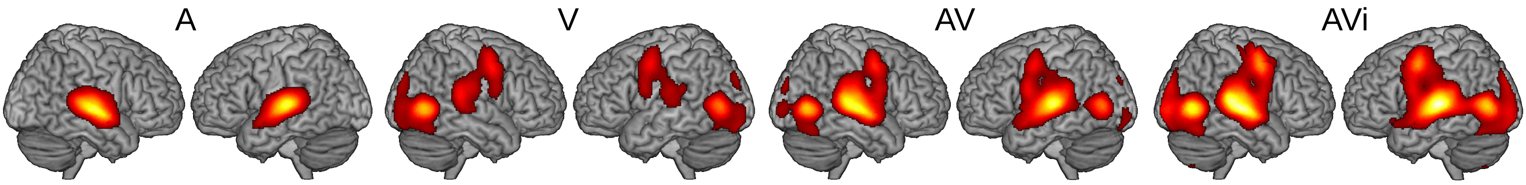
[2] GIPSA-lab & CNRS, Grenoble, France

To recognize one's own face and voice is key for our self-awareness and for our ability to communicate effectively with others. Interestingly, several theories and studies suggest that self-recognition during action observation may partly result from a functional coupling between action and perception systems and a better integration of sensory inputs with our own sensory-motor knowledge. The present fMRI study aimed at further investigating the neural basis of self representation during auditory, visual and audio-visual speech perception. Our working hypothesis was that hearing and/or viewing oneself talk might activate sensory-motor plans to a greater degree than does observing others.

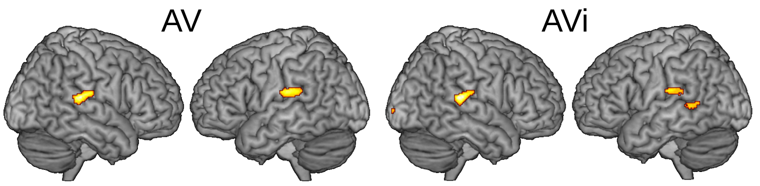
Materials & Methods

- Participants were 12 healthy adults (25±6 years, 9 females).
- A total of 1176 stimuli were created. During the scanning session, **participants were asked to passively listening and/or viewing auditory (A), visual (V), audio visual (AV) and incongruent audio-visual (AVi, self auditory signal, other visual signal) syllables (/pa/, /ta/ or /ka/). Half of the stimuli were related to themselves, the other half to an unknown speaker, and they were either presented with (−6 dB SNR) or without noise. In addition, a resting face of the participant or of the unknown speaker, presented with and without acoustic noise, served as baseline**.
- Functional MRI images were acquired with a 3T whole-body MR scanner (Philips Achieva TX) using a sparse sampling acquisition in order to minimize scanner noise (53 axial slices, 3 mm^3; TR = 8 sec, delay in TR = 5 sec). A high-resolution T1-weighted whole-brain structural image was acquired for each participant after the last functional run (MP-RAGE, sagittal volume of 256 x 224 x 176mm^3 with a 1 mm isotropic resolution
- fMRI data were processed and analyzed using SPM8. Data preprocessing steps for each subject included: (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 and (6) spatially smoothing them with a three-dimensional Gaussian kernel with a fullwidth at half-maximum of 9 mm.
- BOLD responses were analyzed using a general linear model, including 16 regressors of interest (4 modalities (A, V, AV, AVi) x 2 speakers (self, other) x 2 noise (with, without) levels) and the six realignment parameters, 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.
- **A second-level random effect group analysis was carried-out, with the 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.**

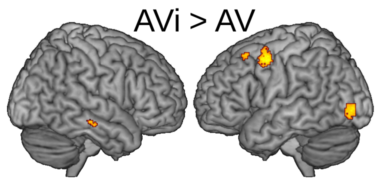
Results – Modality & Noise



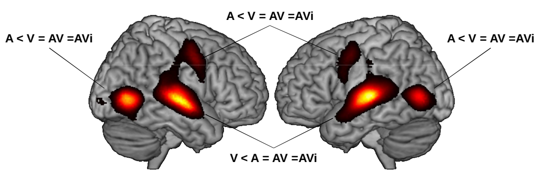
Single Effects (T contrasts, p < .05 FWE corrected): brain activity observed in each modality compared to baseline, irrespective of the speaker and noise.



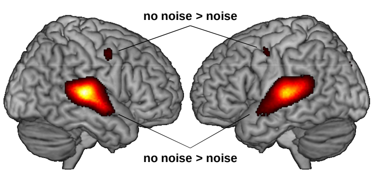
Supra-additivity Effects (T contrasts, p < .001 unc): brain activity observed in the bimodal compared to unimodal conditions (AV > A ∩ AV > V, AVi > A ∩ AVi > V).



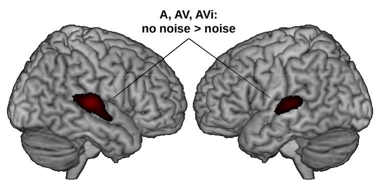
Incongruency Effect (T contrast, p < .001 unc): brain activity observed in AVi compared to AV conditions .



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

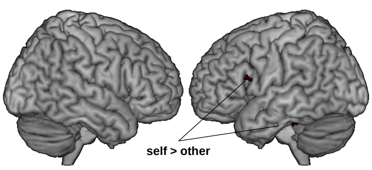


Main Effect of Noise (F contrast, p < .001 unc)

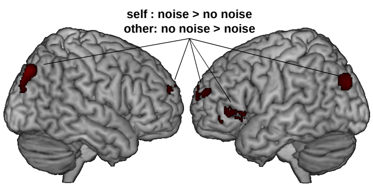


Modality x Noise Interaction (F contrast, p < .001 unc)

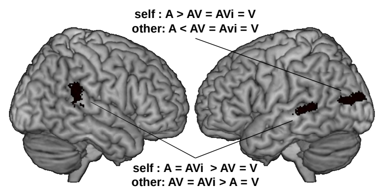
Results – Self



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)

Listening and/or viewing oneself talk was found to activate to a greater extent the left posterior inferior frontal gyrus and cerebellum, two regions thought to be responsible for predicting sensory outcomes of action generation constraining perceptual recognition. In addition, activity in associative auditory and visual brain areas was also found to be modulated by the speaker identity depending on the modality of presentation and the acoustic noise. Altogether these results suggest that self-awareness during speech perception is partly driven by afferent and efferent signals in sensory-motor areas.

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.inpg.fr