Auditory to visual matching of speech and nonspeech oral gestures: an fMRI - study

Gabriele Schmid*,Valentin Riedl°, Afra Wohlschläger° & Wolfram Ziegler*

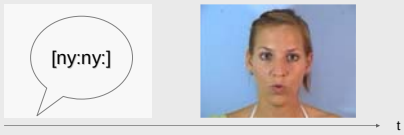
*Clinical Neuropsychology Research Group, Neuropsychological Dept., Municipal Hospital Munich – Bogenhausen, Germany

°Depts. of Neurology, Nuclear Medicine, Radiology, Technical University, Munich

email: Gabi.Schmid@extern.lrz-muenchen.de

Objective. Viewing speech is beneficial to listening to speech. The basis for audiovisual speech processing is likely to be a common supramodal representation. So far it is widely unknown if the audiovisual processing of mouth movements has a specific organisation in the speech as compared to the nonspeech domain. In a former study (Schmid & Ziegler, 2006) we found that subjects have no problems to decide if a heard speech sound matches a seen articulatory gesture, while they show great problems doing the same in soundproducing nonspeech oral gestures (e.g. clicks, whistle, etc.). This difference is likely to be determined by the degree to which sound and articulatory movement are interlinked through perceptual learning. Therefore, speech and nonspeech oral movements offer the possibility of direct comparison between the audiovisual processing in a highly overlearned (speech) and a novel domain (nonspeech). We expect the behavioral differences to be reflected in different patterns of fMRI activity for speech and nonspeech audiovisual processing.

Subjects and Design. In an fMRI study with 19 neurologically healthy subjects (m: 9 / f: 10 ; age (md.) 25 years, range: 21-36) we applied a sequential matching task using speech (syllables) and sound producing nonspeech oral gestures (e.g. clicks) to identify brain areas involved in the matching of information across modalities. The task consisted of two unimodal conditions (visual to visual and auditory to auditory matching) and two cross-modal conditions (visual to auditory and auditory to visual matching). So two stimuli were presented sequentially with an interstimulus intervall of 700 ms and subjects had to indicate via key-press if the two stimuli were identical or different.

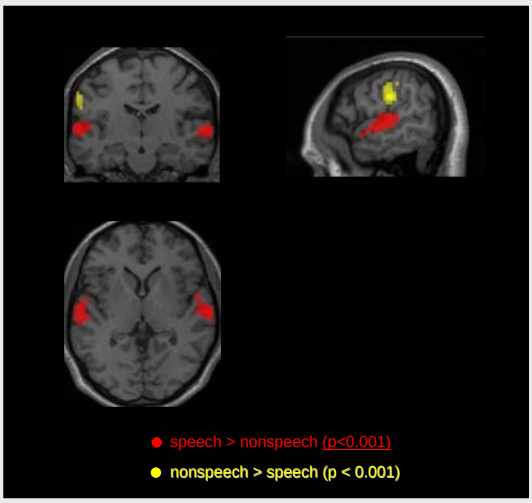


Data acquisition and Analysis. Functional imaging data were acquired on a 1.5T Siemens Symphony scanner with a multislice gradient-echo EPI sequence (TE 50ms, TR 3000ms, flip angle 90°, FoV 224mm^2, matrix 64^2, slice thickness 4mm and 10% interslice gap) at the Neuroradiology Dept., Klinikum rechts der Isar, of the Technical University Munich. Thirty-two axial slices covering the whole brain were acquired every 3 sec over a total scanning period of 36 min yielding four runs. A T1-weighted anatomical dataset was obtained from each subject using a magnetization prepared rapid acquisition gradient echo-sequence (TE=3,93ms, TR=1520ms, angle 15°, FoV 256mm^2, 160 slices, voxel size 1x1x1 mm).

Using SPM5 all images were slice-time corrected, realigned, co-registered, normalized and smoothed (8x8x8 mm^3 Gaussian kernel).

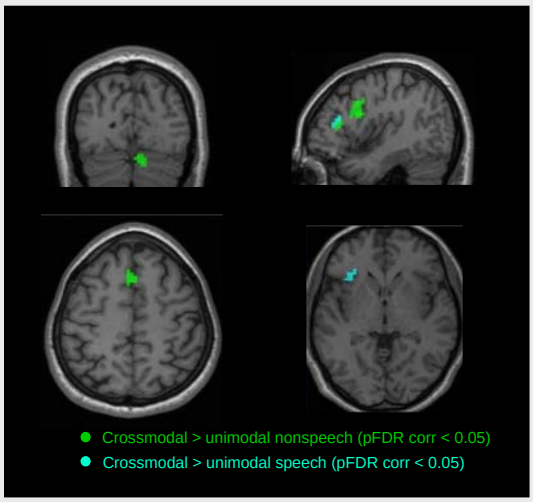
Results

speech versus nonspeech



Comparison for nonspeech > speech in the crossmodal condition led to activation of the left inferior parietal gyrus (IPG), while speech stimuli compared to nonspeech stimuli elicited activation of the STG bilaterally.

crossmodal versus unimodal processing



In the nonspeech domain crossmodal processing as compared to unimodal processing elicited activation of structures in the left inferior frontal cortex (IFG), the pre-SMA and the right cerebellum. In the speech domain comparison of the two conditions resulted in activation of the left IFG and the the left insula.

Discussion

Comparing crossmodal matching for speech > nonspeech led to bilateral STG activation, while nonspeech > speech engaged the left inferior parietal gyrus. It is known that the IPG is activated by observation of movements, motor imagery and movement preparation and is devoted to the perceptual analysis of body movements and closely related to imitation (Mühlau et al., 2005). STG is associated not only with auditory processing, but is known to be involved in audiovisual speech processing (Campbell et al., 1999). So we can conclude that while the crossmodal processing of *speech* is relying on audiovisual representations of speech sounds, in the *nonspeech* domain the brain has to analyse the single components of the stimulus.

Where is the difference in matching *across* as compared to *within* modalities? While in unimodal matching a simple pattern-comparison is possible, crossmodal matching requires a transfer of the information from one modality into the other. This seems to involve different parts of the motor network: in the case of speech, higher order motor representations stored in left anterior language areas (IFG, Insula) are available, whereas in the nonspeech domain a more basic motor network, icluding the cerebellum and left IFG and pre-SMA is required, since nonspeech oral movements are not as highly overlearned as speech.

References

Campbell et al. (1999) Activation in auditory cortex by speechreading in hearing people: fMRI-studies. *avsp99*, August 1999, Santa Cruz

Mühlau et al. (2005) Left inferior parietal dominance in gesture imitation: an fMRI-study. *Neuropsychologia*, 43 (7), 1086-98

Schmid & Ziegler (2006) Audio-visual matching of speech and non-speech oral gestures in patients with aphasia and apraxia of speech. *Neuropsychologia*, 44, 546-555