Multimodal Perception of Prosodic Contrastive Focus in French: A Preliminary fMRI Study

Marion Dohen, Hélène Loevenbruck, Akiko Callan, Daniel E. Callan, Monica Baciu, Cédric Pichat, Harold Hill

Multimodal Perception of Prosodic Contrastive Focus in French: A Preliminary fMRI Study

Marion Dohen[1,2], Hélène Lœvenbruck[1], Akiko Callan[2], Daniel E. Callan[2], Monica Baciu[3], Cédric Pichat[3], Harold Hill[2,4]

[1] ICP, Speech and Cognition Department, GIPSA-lab, France
[2] ATR Cognitive Information Science Laboratories, Japan
[3] Laboratoire de Psychologie et Neurocognition, France
[4] School of Psychology, University of Wollongong, Australia

Contrastive focus is used to emphasize a word or group of words in an utterance as opposed to another. In French, it can be conveyed by prosody using a specific intonational contour on the constituent pointed at (*XXXf a mangé la pomme.* 'XXXf ate the apple.'). It remains unclear what neural processes underlie the perception of prosodic focus. Meanwhile studies have shown that prosodic processing in general cannot be restricted to the right hemisphere (see [1] for review). Moreover it appears ([2]) that even though the perception of prosodic focus was often considered as uniquely auditory, it is possible to perceive prosodic focus visually and the visual modality can enhance perception when prosodic auditory cues are degraded (whispered speech). This finding emphasizes the necessity to consider the perception of prosodic contrastive focus and speech prosody in general as multimodal. The aim of this study is to analyze the neural processing of prosodic focus from a multimodal point of view.

Methods

fMRI recordings were conducted for 12 native speakers of French at the ATR Brain Activity Imaging Center (Japan). Subjects were scanned while they were performing a prosodic focus detection task for three modalities (audio only A, visual only V and audiovisual AV). The stimuli were subject-verb-object (SVO) structured sentences uttered in both normal and whispered speech. In some cases, S was under prosodic contrastive focus. The speaker was a female native speaker of French and the audiovisual stimuli were recorded at ICP. During the functional MRI sessions, whenever no video stimulus was displayed on the screen (A only condition and Null Events), the subjects saw a mid-centered black cross that they were asked to fixate (fixation cross). After seeing/hearing/seeing and hearing each stimulus, subjects were (indirectly) asked to tell whether they had perceived focus or not. The fMRI procedure consisted of an event-related pseudo-random design. Four functional scans were acquired: two for normal speech and two for whispered speech. Each functional scan consisted of 12 sentences in six conditions: AV+focus, AV+no focus, A+focus, A+no focus, V+focus and V+no focus (72 stimuli). 14 null events (NE, fixation cross) were added to vary intertrial interval times. Trials were presented as events lasting 5.1s: stimulus (3s) + response delay (2.1s), NE also lasted 5.1s (total duration of a scan: approximately 7 mn). Functional data analysis (pre-processing and statistical analysis) was performed using the SPM2 software (Statistical Parametric Mapping-Wellcome Department of Cognitive Neurology, London, UK) running on a PC under MATLAB (Mathworks, Sherbon, USA).

Results

Behavioral results

Table 1 provides the percentages of correct answers for all conditions. It appears that subjects performed the task correctly: they were able to identify focus cases from non-focus cases (in all conditions the percentages of correct answers were well above chance). Moreover these

results agree with previous behavioral results on the audiovisual perception of prosodic contrastive focus (see [2]).

Table 1: mean percentages of correct answers and standard deviations (sd) across all subjects for each modality (chance level: 50%).

modality	normal speech		whispered speech	
	% correct	sd	% correct	sd
AV	98.4	2	89.6	6.5
A	97.4	3	69.9	12.8
V	86.4	7.5	88	8.6

fMRI results

The analysis of the fMRI data is still underway and the results presented here are the preliminary observations that were made after a first rough analysis of the data.

We first analyzed auditory alone processing of prosodic contrastive focus (vs. baseline). The aim was to be able to make comparisons with other studies which mainly deal with auditory only perception. We found bilateral activations of the primary and secondary auditory cortices (Superior Temporal Gyrus: BA 22, 41-42). The Left Middle Temporal Gyrus (BA 21) and Inferior Frontal Gyrus (BA 44/45) were activated. The Left Supplementary Motor Area and Cingulate Gyrus (BA 6, 32) were also activated probably corresponding to the press-button task. The activation patterns are very similar for normal and whispered speech apart from an additional activation of the Right Precentral Gyrus (BA 6) for whispered speech.

A preliminary analysis was conducted for the focus vs. no focus contrast using all the data (A, V and AV). Since this contrast is subtle, using more data ensures more robustness for the analysis. It appears that both for normal and whispered speech, identifying a focus case involves the Left Supramarginal Gyrus (BA 40) and the Left Inferior Temporal Gyrus (BA 37). Focus detection for whispered speech additionally requires the Left Superior and Middle Frontal Gyri (BA 8, 9, 11) but not the auditory cortices (unlike for normal speech).

Conclusions

This study shows that auditory processing of prosodic focus (linguistic prosodic task) is partially lateralized to the left which is consistent with the findings of [3] (production of prosodic contrastive focus).

Perception of prosodic focus (vs. no focus) appears to be essentially processed in left associative areas: SupraMarginal Gyrus (BA 40) and Inferior Temporal Gyrus (BA 37). This illustrates the necessity of associating various types of information to detect focus. The fact that the perception of focus (vs. no focus) does not involve auditory cortices for whispered speech shows that auditory cues are probably not crucial and maybe even not used to make a decision for this type of speech when other modalities are available.

Finally, it appears that when no pitch cues are available (whispered speech), prosodic discrimination mainly involves the left hemisphere. The right hemisphere may therefore mainly play a role in the interpretation of pitch cues but not for higher level prosodic structure interpretation.

[1] Baum, S. R. & Pell M. D., 1999. The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology, 13(8)*, 581-608.

[2] Dohen M., Loevenbruck H., under revision. Multimodal prosodic contrastive focus in French: production and perception studies. *The Journal of Phonetics*.

[3] Lœvenbruck, H., Baciu, M., Segebarth C. & Abry C., 2005. The left inferior frontal gyrus under focus: an fMRI study of the production of deixis via syntactic extraction and prosodic focus. *Journal of Neurolinguistics 18*, 237-258.