An fMRI study of multimodal deixis:
preliminary results on prosodic, syntactic, manual and ocular pointing

Francesca Carota[1], Hélène Lœvenbruck[1],Coriandre Vilain[1], Monica Baciu[2], Christian Abry[1],
Laurent Lamalle[3], Cédric Pichat[2], Christoph Segebarth[4]

[1] Institut de la Communication Parlée, UMR CNRS 5009, INPG, Univ. Stendhal, Grenoble, France
[2] Laboratoire de Psychologie et NeuroCognition, UMR CNRS 5105, UPMF, Grenoble, France
[3] INSERM IFR n° 1, RMN biomédicale, Unité IRM 3T, CHU de Grenoble, France
[4] Unité Mixte INSERM / Univ. J. Fourier, U594, Grenoble, France

```
{Francesca.Carota; Christian.Abry; Helene.Loevenbruck; Coriandre.Vilain}@icp.inpg.fr;
       {Monica.Baciu; cedric.pichat}@upmf-grenoble.fr; {Laurent.Lamalle;
              Christoph.Segebarth}@ujf-grenoble.fr
```

Abstract

Deixis or pointing plays a crucial role in language acquisition and speech communication. In this paper we present an innovative fMRI approach in order to examine deixis, conceived as a unitary communicative strategy which employs different verbal and non-verbal speech devices to achieve the pragmatic goal of bringing relevant information to the interlocutors' attention. We designed a unified fMRI paradigm for multimodal deixis, integrating four conditions of verbal and non-verbal pointing: 1) prosodic focus, 2) syntactic extraction, 3) index finger pointing, 4) eye pointing. Sixteen subjects were examined while they gave oral, manual and ocular responses inside the 3T magnet imager. Preliminary results based on a random effect analysis with a group of 10 subjects show that all pointing conditions recruit a left temporo-parieto-frontal network, with respect to the control condition. Further analyses are being carried out to distinguish between different modalities of pointing.

1. Introduction

Deixis, or the ability to draw the listener's attention to an object or person – gradually acquired by children, first by pointing with the eyes, then the finger, then with intonation and finally with syntax –, is crucial in speech communication [1, 2, 3]. Independently of their verbal or non-verbal modality, pointing activities share the same communicative purpose of bringing relevant information to the interlocutors' attention to create a common interest on it. Pointing is thus conceived here as a unitary multimodal pragmatic strategy, which could exhibit by a common cerebral correlates for its modalities.

The existence of a specific cerebral network shared by different pointing modalities is indirectly suggested in previous works. In particular, non-verbal deixis (i.e., manual pointing) would recruit a network including the left posterior parietal and frontal cortex [4]. Verbal deixis, such as prosodic deixis (i.e. focus) and syntactic deixis (i.e. syntactic extraction) would share a common neural network. Prosodic deixis recruits a left temporo-parieto-frontal network, including Wernicke's area, the left supramarginal gyrus, and Broca's region. Syntactic deixis only involves Broca's region. Broca's region is therefore recruited in both types of verbal deixis. The activation of the inferior parietal lobule during prosodic deixis suggests the existence of a functional continuum with manual deixis. To test the hypothesis that multimodal deixis involves a common network of cerebral regions, we designed an fMRI paradigm for multimodal pointing, integrating four conditions of verbal and non-verbal pointing: 1) prosodic focus, 2) syntactic extraction, 3) index finger pointing, 4) eye pointing. During fMRI acquisition, all behavioural responses, including overt speech responses, were systematically recorded. The aim of this fMRI study was to investigate the cerebral substrate of multimodal pointing. Given that pointing activities share the same communicative purpose our hypothesis was that different pointing modalities could also share the same cerebral substrate.

2. Methods

2.1. Subjects

Sixteen healthy right handed volunteers, aged between 18 and 55 years old, native speakers of French, with normal or corrected-to-normal vision and normal neurological history were examined. The study was performed in accordance with the institutional review board regulations.

2.2. Stimuli

Stimuli consisted of two types (1 and 2) of images illustrating a girl (Lise) and a boy (Jules), alternatively placed one on the right and the other on the left side of the screen. Type 1 images show the girl holding a book, while the boy does not; Type 2 images show the boy keeping the book while the girl does not. In the middle of all images, a red or green fixation cross was displayed (see Fig. 1). A blank screen with a mid-centered black cross preceded each stimulus.

Figure1: *A type 2 image, designed to elicit pointing to the boy Jules on the right.*

2.3. Tasks and procedure

Experimental tasks consisted of verbal and non-verbal pointing to a character on the screen. Verbal and non-verbal control conditions were also included. The same underlying question: "Est-ce que Lise tient le livre?" (Does Lise hold the book?) was used for all tasks. Subjects were instructed to perform several tasks: (a) to confirm this target question when type 1 images were presented (*control condition*), (b) to reply by a contrastive pointing to the character Jules when type 2 images were presented (*pointing condition*), (c) to keep the gaze on the black fixation cross when a blank screen was presented. Each of these tasks consisted of two phases: a *preparation* phase, indicated to subjects by the red fixation cross, and an *execution* phase indicated by the green fixation cross. These phases were aimed at better defining the cerebral network involved in an action of pointing (see [4]).

Tasks: During pointing preparation, subjects prepared to perform a contrastive pointing, without real execution. During pointing execution, they actually performed a contrastive pointing. Thus, during prosodic pointing, the subjects uttered « JULES tient le livre » (Jules holds the book), with a contrastive focus on the constituent "Jules". During syntactic pointing, they uttered « C'est Jules qui tient le livre » (It is Jules who holds the book), with a syntactic extraction of the pointed constituent. During manual pointing, subjects pointed with the right index finger to the target character (Jules), and then returned to the start index position. Subjects maintained their right shoulder and arm lying along their side. The elbow was kept in a flexed position using a pillow. The forearm was oriented towards the middle of the abdomen. The index finger was kept stretched. During ocular pointing, subjects pointed with their eyes in the direction of the target character, and then quickly returned to the start eye position (fixation cross).

Control: In control preparation, subjects prepared to perform a confirmation. In control execution, they performed a confirmation. In both prosodic and syntactic controls, subjects neutrally uttered: « Lise tient le livre » (Lise holds the book). In manual control, they performed a rapid downward index finger movement, as if clicking on an invisible mouse. In ocular control, they performed a rapid downward ocular movement starting from the fixation cross.

To avoid any loss of vigilance during the performance and preserve the communicative goal underlying all tasks, the subjects were informed that the experimenter would continuously monitor their behavioural responses. During the prosodic, syntactic, and manual tasks, subjects maintained their eyes fixed towards the fixation cross while the images were projected on the screen, and used the peripheral visual field to determine the position of the character holding the book. During all tasks, the subjects were asked not to perform any other movement than the gesture specific to the (verbal, manual, or ocular) pointing.

2.4. fMRI paradigm

A pseudorandom, event-related fMRI paradigm was employed. The paradigm was based on two types of trials, *i.e.* compound trials (preparation phase + execution phase) and single trials (preparation phase exclusively). Four functional scans were acquired, one for each type of pointing. The same visual stimuli were used in all these scans.

Each functional scan included a sequence of the following four conditions: Preparation of the control task (Pc), Preparation + Execution of the control task (PEc), Preparation of the pointing task (Pp), Preparation + Execution of the pointing task (PEp). A null event (NE, black fixation cross) was added to the four conditions. As a result, five conditions (Pc, PEc, Pp, PEp, NE) were included in total.

The five conditions were alternated between scans and between subjects. 24 repetitions of each condition were presented. Trials were presented as events lasting 4.5 s. More specifically, the duration of the PEp and PEc conditions was of 0.5 s for the initial fixation cross + 2 s for the preparation phase + 2 s for the execution phase. The Pp and Pc conditions consisted of: 0.5 s for the fixation cross + 4 s for the preparation phase alone. The NE condition lasted also 4.5 s. The trial sequences were presented following a pseudo-random order calculated to optimise the following contrasts: condition PEp – condition PEc, Pp - Pc, PEp – Pp, PEc – Pc. Null events allowed us to vary intertrial interval times. Total duration of a scan was approximately 9 mn (4.5 s × 5 conditions × 24 trials). Each scan started with a written instruction reminding the task. Ocular scans started with a 5mn eye-tracking calibration aimed at adjusting eye position recordings with predefined positions on the screen.

2.5. Apparatus

A special apparatus was specifically designed to allow recording subjects' overt responses in three modalities: oral, manual and ocular. Stimuli were presented to subjects by using the Presentation software [6]. They were projected using a video-projector onto a transparent projection screen situated behind the magnet and viewed through a mirror attached to the head coil in front of the subject's eyes. Ear plugs and anti-noise headphones protected subjects against the scanner noise during the 3T fMRI sessions (up to 115 dB). Three types of behavioural responses were recorded: vocal production, eye movement and index finger movement.

2.5.1. Audio data recording

Verbal responses were recorded using an fMRI-compatible microphone. To minimize the amount of noise recorded, the microphone was positioned out of the scanner, at one extremity of a wave guide consisting of a soft plastic tube. The other extremity of the wave guide was connected with a mask positioned over the subjects' mouth. This apparatus sufficiently reinforced the signal-to-noise ratio. An illustration of the audio set-up is provided in Fig. 2.

Figure 2: *Experimental set-up designed to record subjects' verbal production during an fMRI session.*

The names to be pronounced (Lise /liz/ and Jules /ʒyl/) were chosen so as to be sufficiently distinct in the background noise and to minimize mouth opening, since subjects had to wear a tight mask.

2.5.2. Eye-tracking data

Eye position was monitored using an ASL 504 eye-tracker (*Applied Science Laboratories*, Bedford, MA) coupled with the 3T scanner. The recording was synchronised with stimuli presentation using signals sent by the Presentation software.

2.5.3. Finger movement data

Right index finger responses were recorded using a digital camera positioned in front of the scanner out of the fMRI room, behind the glass window. The recording was synchronized with stimuli presentation using a sound signal which was recorded by the camera.

2.6. Subject training

Two days before the experiment, the subjects were intensively trained with the experimental tasks. The fMRI session was partially simulated during the training. After having read the instructions, the subjects rehearsed with slowed-down presentation of stimuli. Then they lied supine in a cardboard tunnel reproducing the dimensions of the magnetic resonance scanner. The stimuli were projected at a normal pace on a projection screen behind the tunnel using the Presentation software. In the tunnel, a mirror attached in front of the subjects' eyes allowed them to view the image stimuli. Recorded audio noise of the fMRI scanner was played during the simulation of the fMRI experience. The subjects wore the headphones and the mask with the plastic tube allowing their oral productions to be monitored. They were trained to use the right intonation and syntactic patterns in the verbal tasks, as well as to perform the appropriate manual and ocular gestures. Their ocular and manual responses could also be followed.

2.7. fMRI data acquisition and processing

A whole-body 3 Tesla MRI imager (Bruker) with gradient echo (EPI) acquisition was used to measure blood oxygenation level-dependent (BOLD) contrast over the whole brain (repetition time: 2.5s; echo time: 30 ms; field of view: 216 x 216 mm, acquisition matrix: 72x72; reconstruction matrix: 128 x 128; 7 dummy scans). Forty-one 3.5 mm axial interleaved slices were imaged adjacent and parallel to the bi-commissural plane. The whole brain and the cerebellum were encompassed.

Between the second and the third functional MR scans, a high-resolution 3D anatomical MR scan was obtained from the volume functionally examined. Anatomical images were acquired using a sagittal magnetization-prepared rapid acquisition gradient echo (MP-RAGE) sequence (inversion time 900ms, volume: 176 x 224 x 256 mm; resolution : 1.375 x 1.750 x 1.33 mm; acquisition matrix: 128 x 128 x 192; reconstruction matrix: 256 x 128 x 128).

A B0 fieldmap was acquired between the first and the second as well as between the third and the fourth scans.

Functional data analysis was performed using Matlab SPM2 software (Statistical Parametric Mapping-Wellcome Department of Cognitive Neurology, London, UK) running on a PC under MATLAB (Mathworks, Sherbon, USA).

Functional MR images were subjected to the following pre-processing steps. Functional data were realigned within functional scans to correct for head motion using a rigid body transformation. A spatial normalisation was applied. Finally, to conform to the assumption underlying SPM that the data are normally distributed, and to allow for inter-subject variability during group analysis, the functional images were spatially smoothed.

To examine cerebral activation during the crucial part of the trials (the second part, after an interval of 2.5s including fixation + preparation), the haemodynamic response to the onset of each event was modelled with a delayed haemodynamic response function (HRF) shifted to onset 2.5 s later than the canonical HRF.

Contrasts between conditions were determined voxelwise using the General Linear Model. In order to perform a random effect analysis, the contrast images (pointing vs control) were calculated for each subject individually and they were then entered into a one-sample t test with significance threshold of p=0.005 uncorrected. Preliminary results of the random effect analysis are reported here for a group of 10 subjects.

3. Results

3.1. Behavioural results

3.1.1. Audio data

Audio data analysis was performed using the Praat software [7] on each of the 1536 utterances (16 subjects, 24 repetitions of both the prosodic pointing condition + its respective control condition = 768, and the syntactic pointing condition + its respective control condition = 768).

For the syntactic pointing, the audio file was carefully examined and listened to, so as to assess that the oral productions corresponded to the expected sentences in both the pointing and the control conditions. For the prosodic pointing, fundamental frequency (F0) measurements were semi-automatically carried out. Peak F0 values were automatically measured using a peak-detection algorithm on the F0 traces provided by the Praat software. Data are still under analysis, but an informal assessment shows that, overall, the subjects behaved according to the instructions.

Two examples of intonational analysis are given in Fig. 3 and 4, which display the acoustic waveform (top panel), the spectrogram with the superimposed F0 curve (middle panel) and the syllabic and prosodic labelling (bottom panels). In the first figure, an utterance pronounced in the control condition is shown. The sentence is divided into two accentual phrases (see [8] for a description of the intonational model used): {Lise} [tient le livre]}. The first accentual phrase bears a typical LH* pattern while the second part of the sentence is marked by a HiLH% pattern, with a continuation rise.
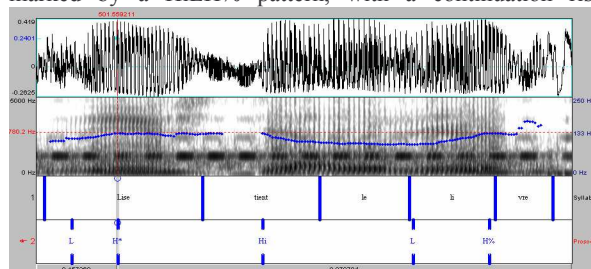

Figure 3: *Intonational analysis of one repetition of the control condition by speaker 5. The same neutral rendition is observed across the 24 repetitions.*

The second example shows an utterance bearing a typical prosodic pointing (focus). The typically high F0 peak on the syllable /ʒyl/ is labelled as LHf (Low High sequence with focus). The postfocal F0 trace falls to reach a flat floor (the

fall is labelled as two low Accentual Phrase boundary tones, L%). These two saliently different F0 patterns are observed across the different repetitions of the prosodic pointing and neutral conditions by this speaker.
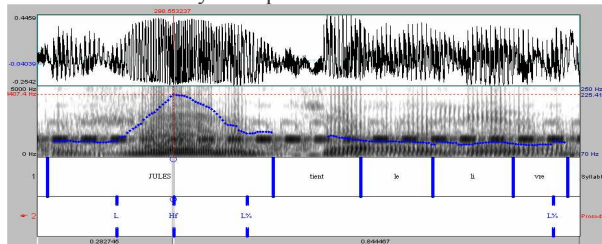

Figure 4: *Intonational analysis of one repetition of the prosodic pointing condition by speaker 5. The same focused rendition is observed across the 24 repetitions.*

3.1.2. Eye-tracking data

Eye-tracking data were processed using the « EYENAL» software (*Applied Science Laboratories* Bedford, MA) integrated in the eye-tracker system. Horizontal positions of the eyes were checked using a Matlab script. Data are still under assessment, but preliminary results suggest that subjects behaved according to the expectations.

3.1.3. Video recording of finger data

The recorded video data were saved in an MPEG format on PC. Data analysis was performed using the Adobe Premiere software (Adobe Systems Inc.). Finger movements are still under examination, but a preliminary overview suggests that the subjects performed adequately.

3.2. FMRI results

Table 1 represents the peaks of activations and their corresponding stereotaxic Talairach coordinates provided by the random effect group analysis on 10 subjects during pointing versus control condition.

Table 1: *Talairach coordinates and Z-scores of activated regions in the pointing vs. control conditions, on a group of 10 subjects*

Region	Talairach coord. (x, y, z in mm)	Z-scores
Right Sup. Par. Lob. (BA7)	12 -50 58	4.07
Left Mid. Temp. Gyr. (BA 41)	-42 -31 10	3.80
Right SupraMarg. Gyr (BA 40)	61 -30 26	3.79
Right Occipital Lobe	26 -89 7	3.69
Left Cerebellum	-38 -46 -28	3.65
Left Insula	-36 8 5	3.54
Right Insula	32 12 8	3.51
Left Precent. Gyr. (BA 4)	-28 -28 57	3.49
Left Sup. Par. Lobule (BA 7)	-38 -47 66	3.29
Right Cerebellum	34 -44 -33	3.23
Left Occipit. Lobe	-20 -94 -5	3.20
Right Occipit. Lobe	42 -84 -1	3.19
Left Front. Gyr. (BA 6)	-8 -14 64	3.16
Right Sup. Par. Lobule (BA 7)	12 -67 53	3.07

Fig. 5 represents functional activations (rendered on a 3D anatomical template) provided by the group analysis during all pointing conditions vs. all control conditions in 10 subjects. The pattern of common activation for all pointing conditions included to the left, the occipital lobe, the middle temporal gyrus (BA 41), the superior parietal lobule (BA 7), the left insula, the left precentral gyrus (BA 4) and the left premotor cortex (BA 6). To the right, activation was detected within the supramarginal gyrus (BA 40), the superior parietal lobe (BA 7), the insula and the occipital gyrus.
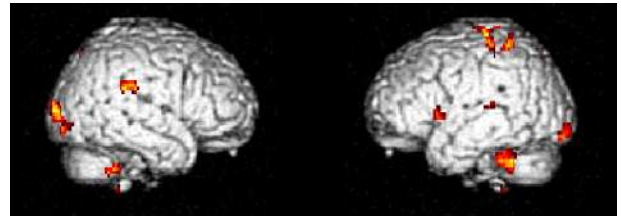

Figure 5: *Activations during pointing vs. control conditions. The left hemisphere is on the right, the right one on the left.*

4. Discussion and conclusion

A specially designed apparatus allowed us to monitor systematically multimodal deictic responses during fMRI sessions. Crucially, overt speech responses were recorded during noisy 3T fMRI session. Audio signal was adequate to be analyzed using classical speech analysis tools. The fMRI results obtained for 10 among the 16 examined subjects suggest the activation of a left temporo-parieto-frontal network for multimodal deixis. Further analyses of the data are currently carried out on more subjects, in order to compare the cerebral regions activated during pointing in different modalities, as well as during pointing preparation vs. execution.

5. References

[1] Kita, S., 2003. Pointing: where language, culture, and cognition meet. Mahwah, NJ: Lawrence Erlbaum.
[2] Capirci, O., Caselli, M.C., Iverson, J.M., Pizzuto, E., Volterra, V., 2002. Gesture and the Nature of Language in infancy: the role of gesture as a transitional device en route to two-word speech. In D. A. Armstrong, M. A. Karchmer, & J. Van Cleeve. *The Study of Sign Languages. Essays in Honour of William C. Stokoe* (pp. 213-246). Washington, D.C.: Gallaudet University Press.
[3] Goldin-Meadow, S., 1999. The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 3, 419-429.
[4] Astafiev S. V., Shulman, G. L., Stanley, C. M., Snyder, A. Z., Van Essen, D. C., Corbetta M., 2003. Functional Organization of Human Intraparietal and Frontal Cortex for Attending, Looking, and Pointing. *Journal of Neuroscience*, 23, 4689-4699.
[5] Lœvenbruck H., Baciu M., Segebarth C., Abry C., 200. 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.
[6] http://nbs.neurobs.com/presentation/
[7] Boersma P., 2001. PRAAT, a system for doing phonetics by computer. Glot International, 5, 341– 345.
[8] Jun S.-A., Fougeron C., (2002). Realizations of Accentual Phrases in French Intonation. *Probus*, 14, 147-172.