Pointing is 'special'

Hélène Loevenbruck, Marion Dohen, Coriandre Emmanuel Vilain

9

Pointing is 'special'

HÉLÈNE LŒVENBRUCK
MARION DOHEN
CORIANDRE VILAIN

Abstract: Deixis, or pointing, is the ability to draw the viewer/listener's attention to an object, a person, a direction or an event. Pointing is involved at different stages of human communication development, in multiple modalities: first with the eyes, then with the finger, then with intonation and finally with syntax. It is ubiquitous and probably universal in human interactions. The *'special'* role of index-finger pointing in language acquisition suggests that all pointing modalities may share a common cerebral network. This chapter aims at better grounding linguistic pointing in somatosensory as well as cerebral domains and at suggesting that it shares features with other pointing modalities. It is shown that manual and ocular pointings seem to recruit left posterior parietal and frontal cortices. Then vocal pointing is presented in details, for both production and perception. It is suggested that integrated multisensory representations may be needed in order to produce and perceive prosodic pointing and that these representations require the activation of associative cerebral areas. The results of a previous study support this hypothesis by showing that prosodic pointing seems to recruit a left temporo-parieto-frontal network whereas grammaticalized syntactic pointing mainly involves frontal regions. The involvement of the left parietal lobe is also apparent in our preliminary fMRI study of the perception of prosodic pointing. Finally, the exploratory results of a new study of multimodal pointing (digital, ocular, prosodic and syntactic) are presented. The common left parietal activation in ocular, digital and prosodic pointing is discussed in the framework of the link between gesture and language.

1. INTRODUCTION

1.1. WHAT IS SO SPECIAL ABOUT POINTING?

Recent research works suggest that deixis, ostension or pointing is at the root of human communication (e.g. Bates et al., 1975; Bruner, 1975; Corballis, 1991; Hewes, 1981; Kita, 2003; Rolfe, 1996; Tomasello et al., 2007). Pointing is the ability to draw the viewer/listener's attention to an object, a person, a direction or an event. Pointing is a communicative device which orients the attention of another person so that an object/person/direction/event becomes the shared focus of attention. It serves to single out, to individuate what will become the referent. In linguistics, deixis is defined as the way language expresses reference to points in time, space or events (Fillmore, 1997). In its narrow sense, it refers to the contextual meaning of deictic words, which include pronouns (me, you), place deictics (here, there), time deictics (now, then) and demonstratives (this, that). In its broad sense, deixis is a referential operation, i.e. it provides a mean to highlight relevant elements in the discourse, to designate, identify or even select an element. It is therefore tightly linked with topicalization, focus and extraction (see e.g. Berthoud, 1990; Jackendoff, 2002). Therefore, pointing, in its broad sense, can be conveyed in several ways: gesturally, with specific hand gestures, or verbally, with specific phonatory and articulatory movements (prosodic deixis) or through the use of a specific syntactic construction (syntactic deixis).

Pointing is 'special[1]' because it is…

… ubiquitous,

A pointing gesture is most often performed with the index finger and arm extended in the direction of the interesting object and with the other

1 The expression "Pointing is special" refers to the debate on the specialised nature of speech. This debate originates from Chomsky's (1957, 1966) position concerning the special nature of the human language faculty. This position has been taken over by researchers from the Haskins Laboratories, and in particular by Alvin Liberman, who wrote "On Finding that Speech Is Special" in 1982. During the summer school, the "speech-is-special" stance was the object of many questions and debates, driven by the Haskins Lab. alumnus Rudolph Sock.

fingers curled inside the hand (Butterworth, 2003). The 'canonical' index finger pointing is *ubiquitous* in everyday interactions, in most cultures of the world. It is observed in oral as well as signed communication. In some sign languages, index pointing is used linguistically for referential indexing. Noun phrases, for instance, may be associated with loci in space. Reference to a previously mentioned noun is performed by pointing again to its specific locus (Klima & Bellugi, 1988).

… universal,

Index finger pointing has been claimed to be a universal ability shared by all human beings (e.g. Povinelli & Davis, 1994). In some cultures, index finger pointing can be replaced with lip pointing (in the Barai in Papua New Guinea, see Wilkins, 2003). But pointing gestures, be they conveyed with the index-finger, the hand, or even the lip or the chin, remain a key *universal* communication tool in humans.

… has a long historical past,

Although its scientific study is relatively recent, index-finger pointing did not emerge in the XXth century. Traces of it can be found in early European art. The most famous paintings of points are those of Leonardo da Vinci, dating back from the late XVth century (such as the Virgin of the Rocks, 1483-1486, with the angel Uriel pointing towards baby John-the-Baptist; or St John the Baptist, 1513-1516, with his right hand pointing up toward heaven). But even earlier points can be found, like in the Bayeux Tapestry, which was embroidered towards the end of the XIth century, and which tells the 1066 Norman invasion of England (e.g. William of Normandy, on his ducal throne, pointing at Harold making an oath; or men pointing at the ominous Halley's comet). Furthermore pointing was not only used in European culture. Sculptures of pointing buddhas can be found in XVIIIth century Burmese art, for instance.

… performed by evolved animals,

It has long been claimed that pointing is a human-specific ability and that great apes, for instance, do not point (e.g. Povinelli et al., 2003). Works by Leavens, Hopkins and colleagues show, however, that chimpanzees in captivity can in fact point at unreachable food (Leavens et al., 2005). They sometimes use 'canonical' index finger pointing, but most often, they point with all fingers extended. And, like human infants, chimpanzees point spontaneously, without explicit training.

Therefore, although pointing does not seem to be specific to human primates – since apes, and other animals (think of the vocal reference calls of the vervet monkeys), may well be capable of it – it seems to be a *sophisticated social* behaviour, used by animals capable of some degree of imitation and inter-individual communication (Pollick & de Waal, 2007). Some researchers have even suggested that the index-finger pointing ability might be related to the thumb-index finger opposition and the pincer grip capacity (Butterworth, 2003).

… multimodal,

Pointing involves multiple modalities: it can be manual or digital (with the index finger), labial, facial (chin), ocular, and vocal.

As described above, pointing is most often, at least in Western cultures, *digital*, i.e. it is conveyed with an extended index finger. It is also very often *manual*, i.e. the whole extended hand is used to point. But when the hand is used, the handshape may vary (open flat hand held palm up, or palm facing laterally, thumb: see Kendon, 1996). Pointing can also be performed with protruded lips. *Labial* pointing has been observed in several communities in all inhabited continents (the Kuna indians in Panama, the Arrernte in Australia, the Awtuw and Barai in Papua New Guinea, the Ewe in Ghana, the Navajo in North America; cf. Wilkins, 2003). In fact, as Kendon (1996) notes, many different body parts can be used to perform pointing (head, lip, chin, elbow, foot, arm and hand).

Ocular pointing or deictic gaze is also frequent. It is the ability to (alternately) look at the 'object' then at the viewer's eyes. It is an invitation for the viewer to look at the object which becomes the shared focus of attention. A wonderful example of ocular pointing is the famous painting by Georges de la Tour, "Le Tricheur à l'as de carreau" (The cheat with the ace of diamonds, 1635). According to our interpretation of this painting, the woman standing next to the cheater gazes at him, thereby designating him for the woman sitting at her left. The sitting lady understands the ocular point, as she responds with a digital point.

Vocal pointing is what is sometimes called deixis or focus. In French, as in many languages, vocal deixis can be conveyed by syntactic extraction, using a deictic presentation form such as in the example below:

C'est Madeleine qui m'amena
*(**It's** Madeleine **who** brought me along.)*

It can also be conveyed by prosodic focus, i.e. by using a specific intonational contour on the pointed item, such as in the example below:

MADELEINE$_F$ m'amena.
(MADELEINE$_F$ brought me along.)

The effect of this intonational contour, which will be described in more detail in section 2.1., is to highlight the pointed item ('Madeleine'), the rest of the utterance bearing a flat, post-focal contour.

… involved in stages of language development

Ocular pointing (or deictic gazes, at 6 - 9 months) and, later, index finger-pointing (deictic gestures, at 9 - 11 months) have been shown to be two key stages in infant cognitive development that are correlated with stages in oral speech development.

Pointing with the eye is first observed at 8 to 9 months, mutual attention takes place between the adult and the baby. When the adult looks at something, the baby follows his/her gaze and looks in the same direction. At this stage, babies therefore have the ability to look where someone else is looking. Conversely, the baby seems to invite the adult to look at an object, by alternately looking at the object then at the adult's eyes. Quite early in development, babies can therefore use gaze to manipulate the attention of their carer.

Then, at 9 to 11 months, when infants start to be able to understand a few words, they produce pointing gestures, most often *index-finger pointing*. The emergence of pointing is a good predictor of first word-onset, and gesture production is related to gains in language development between 9 and 13 months (Bates et al., 1979; Butcher & Goldin-Meadow, 2000; Caselli, 1990). As explained in Butterworth (2003), pointing not only serves to single out the object but also to build a connection between the object and the speech sound. Finger pointing therefore seems clearly associated with lexicon construction.

Then at 16 to 20 months, during the transition from the one-word stage to the two-word stage, *combinations of words and deictic gestures* can be observed (such as a pointing gesture to a location and pronouncing 'dog', to indicate that a dog is there). Furthermore, the number of gestures and gesture-word-combinations produced at 16 months are predictive of total vocal production at 20 months (Morford & Goldin-Meadow, 1992; Capirci et al., 1996; Goldin-Meadow & Butcher, 2003;

Volterra et al., 2005). Finger pointing therefore clearly seems associated with morphosyntax emergence.

As for *vocal pointing*, only a few studies have examined the development of *prosodic focus* and *syntactic extraction* in children.

Concerning the *acoustic realisation of prosody*, pre-school children have been shown to master contrastive focus in English, in absence of any formal teaching (Hornby & Hass, 1970). Little is known about the development of prosodic focus in French, but the ability to realise adult-like pitch movements emerges quite early (Konopczinsky, 1986). Also, it has been shown that adult-like syllable lengthening is realised by young children (from 2 years of age) on accented vowels (Konopczinsky, 1986).

As concerns *articulatory production of prosodic pointing*, what is known is that lexical stress in English is differently realised in young children than in adults (Goffman & Malin, 1999; Connaghan et al., 2001). Children aged 3 to 6 years tend to produce larger movements, with reduced velocities, and increased durations, characteristics of slowly and carefully articulated speech. In a recent video study on French, Ménard et al. (2006) showed that French-speaking children (aged 4 and 8) do not differentiate articulatory and acoustic patterns across focused and unfocused syllables as much as adult speakers do. To summarise the findings, although the production of prosodic focus, or vocal pointing, seems to be mastered quite early (as early as 3 years of age) in the acoustic domain, and without formal teaching, its articulatory correlates seem to be acquired much later. But what seems lacking in children is not the capacity to hyper-articulate focused phrases, but to hypo-articulate surrounding phrases. The delay in articulatory performance could thus be related to general articulatory proficiency rather than to specific linguistic mastery. Since manual pointing emerges spontaneously before the end of the first year, it could be that acoustic correlates of prosodic focus are in fact mastered much earlier than 3 years of age.

The development of syntax in young children has been extensively studied by Tomasello and colleagues, among others (see also Brown, 1971; Sheldon, 1974; Tavakolian, 1981; MacWhinney & Pléh, 1988; McKee et al., 1998; Kidd & Bavin, 2000). In a study of the speech of English-speaking children between 1;9 and 5;2 years of age, Diessel & Tomasello (2000) showed that the earliest and most frequent relative clauses that children learn occur in *presentational constructions* that are propositionally simple.

They express a single proposition in two finite clauses, such as: ”Here’s a tiger **that’s** gonna scare him” or “That’s the sugar **that** goes in there”. According to the authors, one of the factors that might explain why relative clauses emerge in these presentational constructions is the prefabricated character of the main clause. Interestingly, the main clause contains a deictic pronoun (i.e. this, that, here, there, it). Diessel & Tomasello argued that *“the early use of relative constructions involves a very simple procedure by which the child combines a prefabricated (main) clause (i.e. a clause of the type That’s X, There’s X, It’s X) with a second component […].”* (p.144). They observed presentational constructions in children aged 2 and younger (1;11). They quoted a study by Jisa and Kern (1998) on French children (aged 5;0-5;11) who also reported extensive use of presentational constructions in young children.

To sum up, pointing is involved at several stages of human communication development. It seems to be one of the first communicative tools used by babies. It is a key part of the shared attention mechanism in child-adult interaction. It seems to emerge spontaneously, in stages, and in association with oral productions. The crucial role of index-finger pointing in language development and the involvement of multiple forms of pointing at different stages of human communication development, first with the eyes, then the finger, then intonation and finally syntax, suggest that all pointing modalities may share a common cerebral network. The aim of this chapter is to better ground linguistic pointing in somatosensory and cerebral domains and to show that it shares features with other pointing modalities.

First, we will discuss the results of several studies in the literature related to manual and ocular pointing that can provide clues on the cerebral correlates of these two pointing modalities and can shed light on the potential general pointing network.

Then we will focus on oral linguistic pointing. Among the two forms of linguistic pointing, prosodic focus will be described in particular, since it bears interesting physiological characteristics. Its acoustic and articulatory correlates will be given in detail. It will be shown that prosodic focus is signalled using very specific acoustic and articulatory features and that speakers use reliable strategies to organise phonation and articulation together in order to convey prosodic pointing. On the perception side, results from auditory-visual experiments will be

presented that indicate that the phonatory and articulatory patterns are not just productive habits but that they are in fact well recovered by listeners and viewers. It will be suggested that integrated representations (acoustic, articulatory, proprioceptive) may be needed in order to produce prosodic pointing adequately and that these multisensory representations may be formed via the activation of associative cerebral areas. It will be hypothesised that these areas should also be involved during the perception of prosodic focus. Two preliminary fMRI studies of the production and perception of prosodic focus will be presented. Finally, we will present the first results of an fMRI study on the different pointing modalities that suggest that manual, ocular and prosodic pointing may well be grounded in a same cerebral network.
The organisation of the chapter is as follows. In section 1.2., a review of the literature provides elements for a description of the cerebral networks of manual and ocular pointings. These networks are the grounding red thread for the study of oral pointing presented in section 2. A summary of several works carried out in our laboratories on the production and perception of prosodic pointing, both in the acoustic and articulatory (or visible) domains, is provided in sections 2.1. and 2.2. This summary is then used to make hypothesis on the potential cerebral correlates of prosodic pointing in both production and perception. Results of a previous work on the cerebral correlates of the production of prosodic and syntactic pointing are provided in section 2.3.1. A work-in-progress on the cerebral correlates of the auditory-visual perception of prosodic focus is presented in section 2.3.2. Finally, preliminary results of an fMRI study of pointing in different modalities (manual, ocular, prosodic, syntactic) are presented in section 3. The implications of these findings are discussed in the conclusion.

1.2. MANUAL AND OCULAR POINTING

As explained above, the crucial role of pointing in communication, be it with the eyes, the finger, intonation or syntax, suggests that all pointing modalities may be grounded in a common cerebral network. The first question we want to address is: what are the cerebral correlates of manual and ocular pointing? Several observations provide preliminary

answers to this question. The first set of observations deals with the role of the posterior parietal regions of both hemispheres in manual pointing tasks. It has been suggested that the spatial representations formed in the posterior parietal and premotor frontal regions could provide "*perceptual-premotor interfaces for the organization of movements (*e.g. *pointing, locomotion) directed towards targets in personal and extrapersonal space*" (Vallar, 1997, p.1401). Patients with left unilateral neglect have been shown to present deficits in pointing tasks (e.g. Edwards & Humphreys, 1999) while PET studies on normal subjects show activation within the left and/or right inferior parietal lobule (IPL) during pointing tasks (e.g. Lacquaniti et al., 1997; Kertzman et al., 1997). The role of the right and left posterior/inferior parietal regions in pointing tasks may further be related to data on brain-damaged deaf signers. Bellugi and colleagues presented a study of deaf signers of American Sign Language (ASL), two of which presented lateralized parietal lesions, one in the right hemisphere and the other in the left (Bellugi et al., 1989). Space in ASL is handled in two ways. The first is topographic: in the description of the layout of objects in space, spatial relations among signs reproduce the actual spatial relations among the objects. The second is deictic: space is used for referential indexing (as explained in the introduction). The right-lesioned signer had difficulty in the use of space for topography: room description was distorted spatially, with left side of signing space neglected. In the use of space for syntax, however, the entire signing space was covered and consistent reference to spatial loci was preserved. By contrast, the left-lesioned signer produced room descriptions without spatial distortions but made errors in the deictic use of space.

More recently a study on *finger pointing and looking* suggests that pointing with the finger and pointing with the eye seem to recruit a common left lateralized network including the frontal eye field and the posterior parietal cortex (Astafiev et al., 2003). In this study, the manual pointing task included a preparation phase and an execution phase. During the preparation phase, the subjects had to prepare to point at a cued location with the right index finger. During the execution phase, the subjects were asked to point towards the target as soon as the target flashed. The saccade task also included preparation and execution phases. Since the instruction was to (prepare to) look at a cued location, it can thus, according to us, be assimilated to ocular pointing. *In saccade*

preparation, a transient sensory response to the cue was observed in the bilateral occipital cortex, a sustained response was observed in the bilateral frontal cortex, at the junction of precentral and superior frontal sulci, i.e. in the human frontal eye field (FEF) and a sustained response was observed along the horizontal segment of the intraparietal sulcus (IPS). *In pointing preparation*, bilateral FEF and IPS activations were observed, just as in saccade preparation. These results therefore suggest that the posterior parietal cortex and the frontal cortex contain regions that code preparatory signals for pointing independently of the effector used (eye, index-finger). They are consistent with previous fMRI experiments that reported common activity in the IPS and FEF for visually guided saccades and pointing movements (Connolly et al., 2000; Simon et al., 2002). Additional left hemisphere activation was observed for the manual pointing task, in the angular gyrus, the supramarginal gyrus, the superior parietal lobule, the dorsal precentral gyrus, and the superior temporal sulcus. The authors checked that the left-lateralization held, even when the left hand was used. Therefore manual pointing-specific responses were observed in the left lateral and medial posterior parietal and frontal cortex. Similar posterior parietal and superior frontal activations were observed for eye as well as arm pointing tasks by Hagler et al., 2007. But no evidence was found for manual-pointing-specific maps.

To summarise, recent neuroimaging studies provide clues to the description of the cerebral networks involved in manual and ocular pointing. They suggest that posterior parietal cortex and frontal cortex are bilaterally activated in ocular and manual pointing tasks. Additional left lateralized activation in the medial posterior parietal cortex could be specific to manual pointing. The question we want to address now is whether a similar network is involved in linguistic pointing. First, we will review several studies carried out in our laboratory that indicate what the speaker's somatosensory representations might be for linguistic pointing as well as to what extent the listener/viewer is able to decode audiovisual signals of vocal pointing. This will enable us to make hypothesis as to which regions the cerebral network for the production and perception of vocal pointing may include.

2. WHAT DO WE KNOW ABOUT VOCAL POINTING?

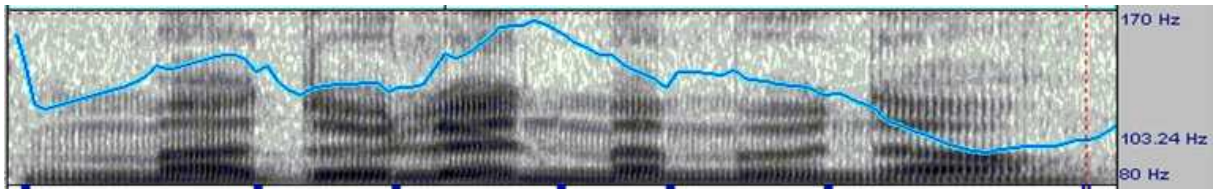
As mentioned in the introduction, vocal pointing in French (or deixis) can be conveyed by syntactic extraction or prosodic focus (Berthoud, 1990). Prosodic focus, as will be shown in the following, involves specific intonational and articulatory patterns. Syntactic extraction involves the use of a cleft form (*"c'est …. qui"*, "it's … who"). It may or not be accompanied by prosodic focus. It should be mentioned here that there are two types of prosodic focus in French: *contrastive focus* which selects a constituent in the paradigmatic dimension and *intensification focus* which makes a contrast along the syntagmatic axis (Séguinot, 1976). This chapter is only concerned with *contrastive* prosodic focus and the term 'prosodic focus' will be used hereafter to refer to this type of focus.
Compared with a broad focus or non-deictic rendition of an utterance, *prosodic* focus involves precise acoustic and articulatory modifications, whereas *syntactic* extraction may simply use additional words, but no specific somatosensory change. We will first describe what we know about the somatosensory characteristics of *prosodic* focus, namely its acoustic and articulatory correlates. We will also examine how prosodic focus is perceived visually. This will lead to conjectures on the cerebral regions involved in the production and perception of prosodic pointing.

2.1. ACOUSTIC CORRELATES

Acoustic correlates of contrastive prosodic focus have been extensively described (see e.g. Astésano, 2001; Astésano et al., 2004; Dahan & Bernard, 1996; Delais-Roussarie et al., 2002; Di Cristo, 1998; Di Cristo & Jankowski, 1999; Dohen & Lœvenbruck, 2004; Jun & Fougeron, 2000; Rossi, 1999; Touati, 1987). Prosodic deixis in French involves an increase in fundamental frequency (F0) on the focused constituent as well as a lengthening of the focal constituent. It is also associated with an F0 lowering on the pre-focal constituent and a deaccentuation of post-focal constituents. Figure 1 illustrates the effect of prosodic focus on F0. Figure 1a shows the spectrogram and F0 for the neutral rendition of the sentence *"[MADELEINE]$_{AP}$ [m'amena]$_{AP}$"* (MADELEINE brought me along). Figure 1b shows the same utterance with a focused subject. F0

raising on the first syllable of the Accentual Phrase[2] 'Madeleine' and post-focal deaccentuation are clearly visible. Metrical and rhythmical aspects of prosodic focus have also been studied. Following Di Cristo, Astésano (2001) postulates that the emphatic initial accent in contrastive focus is the 'hyper' realisation of the metrical, secondary initial accent in French. In that view, the initial contrastive accent shares common phonetic features with the metrical initial accent: both have similar infra-syllabic lengthening, with significant lengthening of the onset over the rhyme, and similar overall F0 configurations, which clearly distinguish them from final accents. However, these acoustic correlates are twice as large for the emphatic initial accent as for the metrical initial accent, confirming that it is a 'hyper' version of the initial accent (Astésano et al., 2007). In addition to the increase in duration of the focused phrase (typically of the focused syllable onset), an anticipatory increase in duration of the pre-focal syllable can sometimes be observed (Dohen & Lœvenbruck, 2004).

a.



b.

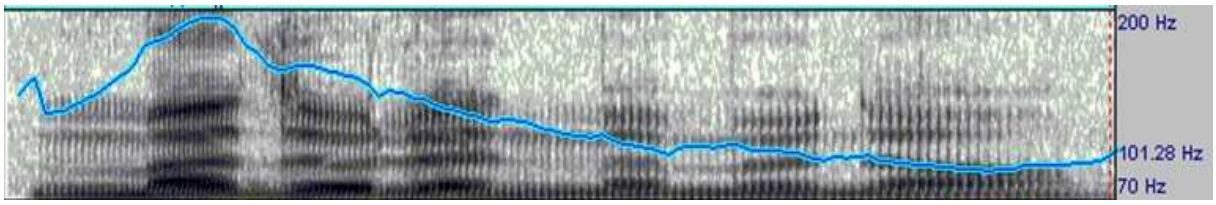


Figure 1: a) Spectrogram with superimposed F0 trace for a neutral rendition of the sentence "Madeleine m'amena". **b)** Focused rendition of the same sentence. The increased F0 peak on the first syllable /ma/ is typical of a focused phrase. The post-focal F0 trace falls to reach a flat floor. The second (highest) F0 peak observed in the neutral rendition (typical of the end of an Accentual Phrase) is suppressed.

2 For the definition of an Accentual Phrase (AP) in French, see Jun & Fougeron (2000). The AP has the default tonal representation /LHi LH*/, with an initial high tone Hi (also described as the peak of the 'accent secondaire'), a final high tone H*, realised on the phrase-final full syllable (peak of the 'accent primaire'), and two low L tones realised on the syllable preceding the H-toned syllable.

But there is no decrease in the duration of the post-focal items (no dephrasing). Restructuring of the focused Accentual Phrase is sometimes observed: it can be grouped with other phrases into one Intonational Phrase or divided into several Accentual Phrases.

2.2. ARTICULATORY CORRELATES, INCLUDING VISIBLE CUES

Summarised results of two research works on French, carried out in our laboratory, are presented here. The first one concerned tongue movements and was carried out using an electromagnetometer. The second set of studies examined lip and other facial movements. It used video and optical recordings. Perceptual tests checked whether the articulatory cues are perceived visually.

2.2.1. TONGUE MOVEMENT

Articulatory movements of the tongue were tracked using an electromagnetometer during the production of prosodic focus in several conditions (Lœvenbruck, 1999). Detailed methods are provided in Appendix A.1. A preliminary F0 analysis showed that, in all focused utterances, the F0 peak was carried by the phrase-initial /la/ syllable. The articulatory characteristics of the phrase-initial /la/ syllable in the neutral conditions were thus compared with those of the focus conditions. Articulatory analyses, shown in Figure 2, consisted in measuring maximal tongue displacement and peak velocity within the syllable.
The displacement data presented in Figure 3a show that, for this speaker, the tongue displacement in the /la/ syllable is larger in focused than in unfocused conditions, when the syllable belongs to a word that is not utterance initial. This means that the tongue movement is larger under focus in these sentence positions. When the word is utterance-initial, the tongue displacement is not larger when the syllable is focused. This is probably due to the fact that word-initial syllables in utterance-initial position are often articulated with more amplitude, even in unfocused contexts (see e.g. Fougeron & Keating, 1997; Tabain, 2003).

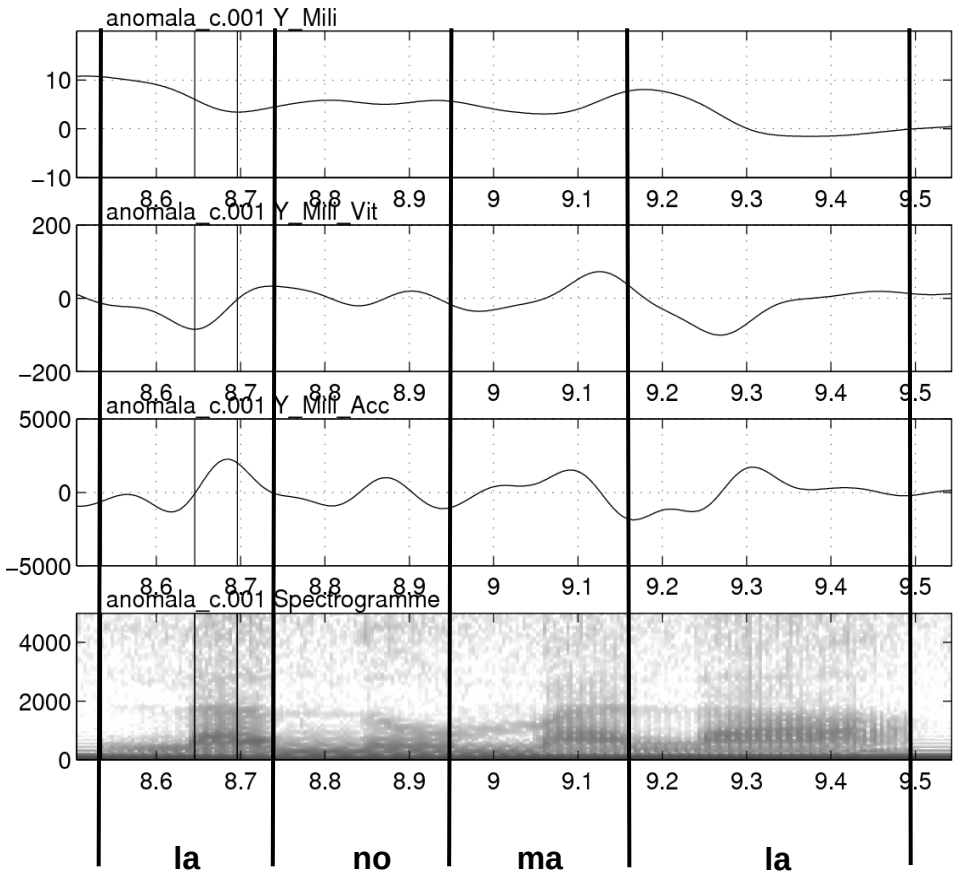
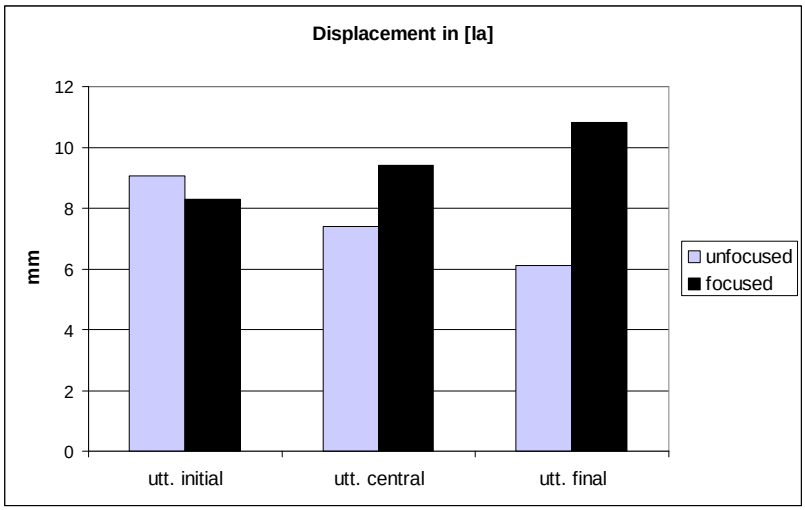


Figure 2: Articulatory labels for the first syllable of [lanomala] in 'Elle annihilait l'anomala en l'éloignant'. Top: tongue-middle vertical position (in mm); Middle: velocity (mm/s); Bottom: acceleration (mm/s^2). The dark vertical lines are acoustic landmarks for syllables [la], [no], [ma] and [la]. The disconnected lines are articulatory labels within the first syllable [la].

The results of the duration data, presented in Figure 3b, show that, for this speaker, the duration of the /la/ syllable is longer in focused than in unfocused conditions, when the syllable is not utterance initial. For utterance initial syllables, there is no increase in duration when the syllable is focused. This is probably due to the fact that word-initial syllables (particularly the onsets) in utterance-initial position are often lengthened, even in unfocused conditions (see Tabain, 2003).

a.



b.

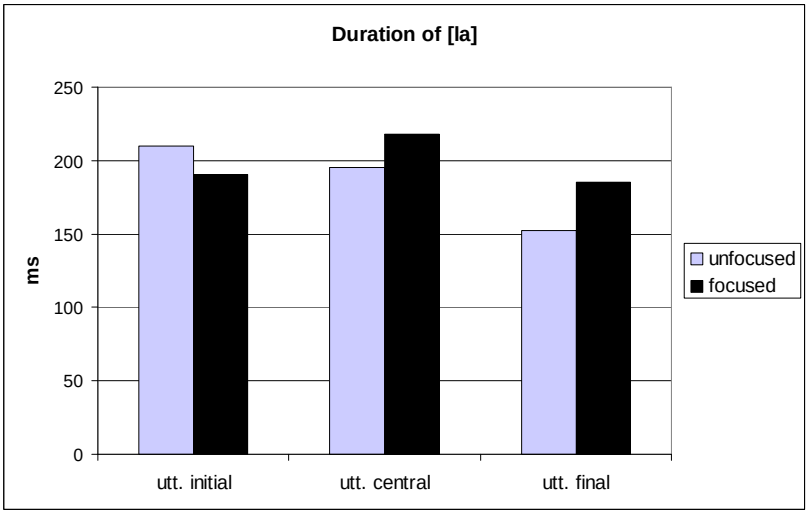


Figure 3: a) Tongue vertical displacement (mm) for /la/ in unfocused (light bars) and focused (dark bars) conditions, as a function of the position of the phrase in the utterance. **b)** Duration of /la/ in unfocused and focused conditions.

2.2.2. LIP CUES

In a second set of studies, movements of the lips were tracked during the production of vocal pointing (prosodic contrastive focus) using different measurement techniques.

a. The video study

The results summarised below are detailed in Dohen et al. (2004) and Dohen & Lœvenbruck (2005). The aim was to examine the possible articulatory visible correlates of contrastive focus in French. Data were collected for two speakers (A & B) for reiterant (A) and real speech (A &

B) using a very accurate lip-tracking device (descriptions can be found in Audouy, 2000). It appeared that contrastive focus was characterized by an increase (hyper-articulation) in inter-lip area and in inter-lip area peak velocity for the focal phrase. The amount of hyper-articulation was however highly speaker-dependent. Speaker A hyperarticulated much more than speaker B. Due to corpus constraints, protrusion could be analysed for speaker B only. It was also hyper-articulated and to a greater extent than inter-lip area. In addition, it appeared that speaker B hypo-articulated the post-focal sequence (reduced inter-lip area, inter-lip area velocity and protrusion) while speaker A barely did. Durational measurements were also conducted since duration can be a visual cue as well. These measurements showed that the focused syllables were significantly lengthened for both speakers, the first phoneme of the focused phrase being even more significantly lengthened. For speaker A, it was also observed that the last syllable of a phrase was significantly lengthened and hyper-articulated when the following phrase was focused. This was related to an anticipation strategy. We concluded from these observations that there is a global tendency towards hyperarticulating the focused phrase but that other visible cues are produced and that they appear to be speaker dependent. This is why, in order to be able to identify potential generic strategies, it seemed important to extend this study to a greater number of speakers.

Visual only perception tests were also conducted using the videos of speakers A and B (Dohen & Lœvenbruck, 2005). These tests showed that contrastive focus could be perceived through the visual modality alone and that the visual cues used for perception corresponded at least in part to those identified in the production studies. For both speakers, a few stimuli were well perceived even though the visible correlates described above (i.e. focal hyper-articulation and lengthening, with or without post-focal hypo-articulation) were not present. This suggested that other more subtle facial correlates may intervene. Studies on other languages have indeed shown that other facial movements such as eyebrow movements (Krahmer et al., 2002; Grandström & House, 2005) or head movements (Hadar et al., 1983; Cerrato & Skhiri, 2003; Munhall et al., 2004) or both (Graf et al., 2002) could intervene. Cavé et al. (1996) also showed that, in French, F0 variations and eyebrow movements could be

linked. This is why it also seemed necessary to enlarge the set of facial movements measured by the use of a complementary technique.

b. The 3D motion capture (Optotrak) study

The aim of this second study, detailed in Dohen et al. (2006), was to further explore the articulatory visible correlates of contrastive focus in French, by extending the preliminary study in two directions. The first extension consisted in including a greater number of speakers in order to identify possible generic strategies. The preliminary study showed that one speaker's way to single out the focused phrase was to locally increase inter-lip area and its derivative as well as duration. The other speaker's behaviour was spread across the entire utterance. It consisted in creating a contrast within the utterance, by slightly hyper-articulating and lengthening the focused phrase as well as hypo-articulating the post-focal sequence.

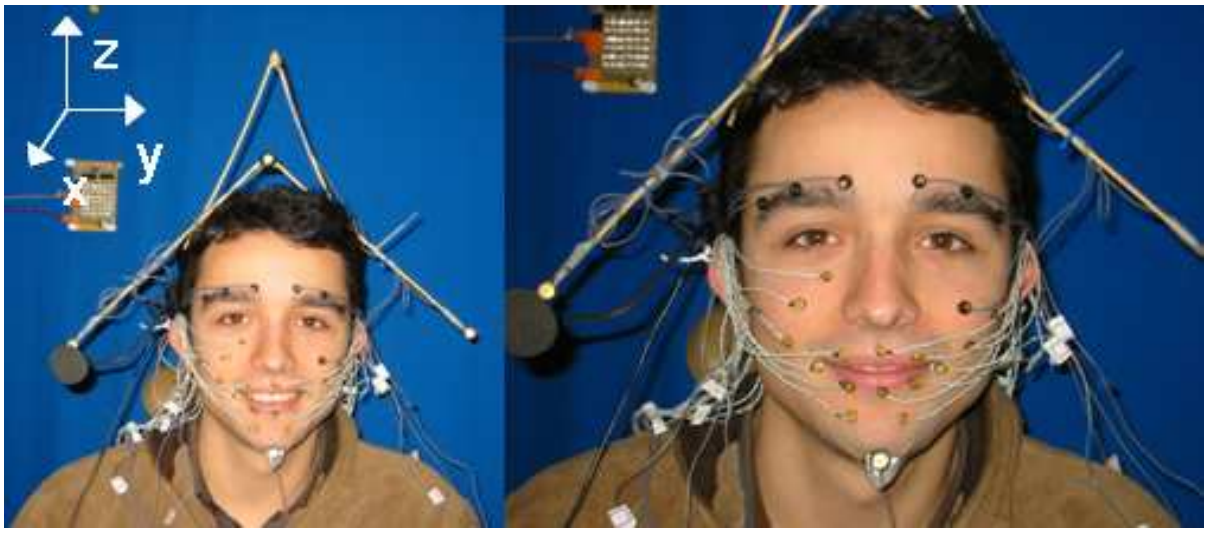


Figure 4: Optotrak measurement device: experimental setup. The 24 IREDs glued to the speaker's face are visible, as well as the 4 additional IREDs attached to a head rig for head motion correction. The x, y, z axes used for the articulatory measurements are shown. See Appendix A.2.

The first aim of this second study was thus to determine whether the two behaviours observed are shared by other speakers or if more variability is at stake. The second addition to the preliminary study consisted in extending the set of facial measurements by the use of a complementary

recording technique. Results of the perceptual tests in the preliminary study suggested that other facial correlates than lip parameters may play a role in focus perception. The second study made use of an optical technique that allowed for the monitoring the positions of several markers on the face. Details on the experimental paradigm, the recording procedure and the data processing can be found in Appendix A.2. Five native speakers of French (B, C, D, E and F) were recorded using a 3D optical tracking system: Optotrak (IRED tracking system using markers glued to the speaker's face), which is less accurate on lip contours than the system used in *a.* but provides more facial data. Figure 4 gives an idea of the experimental setup used.

Articulatory and durational analysis

Figure 5 provides a visual overview of the results. Table 1 groups detailed intra- and inter-utterance contrasts (percent changes for each speaker and each parameter).

Speaker B – This speaker significantly lengthened and hyper-articulated (except lip spreading) the focused phrase. He also hypo-articulated the post-focal phrase(s). The strongest contrasts were measured for lip protrusion and duration and the smallest contrasts for jaw movements.

Speaker C – This speaker significantly lengthened and hyper-articulated the focused phrase. A slight hypo-articulation of the post-focal phrase(s) was also observed but only for jaw movements and lip opening. This speaker anticipated focus on the preceding phrase (slight lengthening and hyper-articulation for jaw movements and lip opening limited to the directly preceding phrase). The strongest contrasts were measured for lip protrusion and duration.

Speaker D – This speaker significantly lengthened and hyper-articulated (except lip spreading) the focused phrase. A post-focal hypo-articulation was measured for lip opening and lip protrusion. This speaker anticipated focus by hyper-articulating (only lip opening) the directly preceding phrase. The strongest contrasts were measured for lip protrusion and the smallest contrasts for lip spreading.

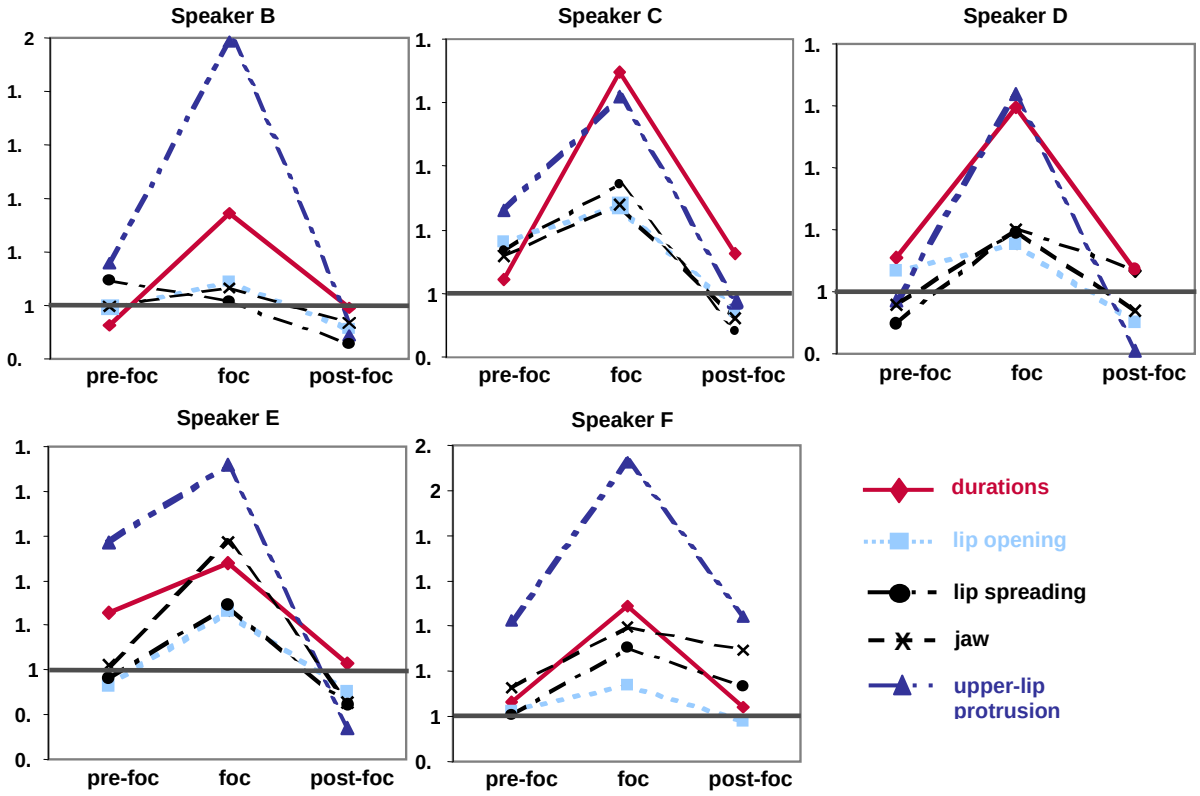


Figure 5: Durational and articulatory measurements for all five speakers: normalised values corresponding to the pre-focal (pre-foc), focal (foc) and post-focal (post-foc) sequences (the dark horizontal lines correspond to the neutral case, i.e. 1).

Speaker E – This speaker significantly lengthened and hyper-articulated the focused phrase. He also hypo-articulated the post-focal phrase(s) and anticipates focus by hyper-articulating (only lip protrusion) the directly preceding phrase. The strongest contrasts were measured for lip protrusion and the smallest contrasts for lip opening and lip spreading.
Speaker F – This speaker significantly lengthened and hyper-articulated the focused phrase. He also anticipated focus by hyper-articulating (only lip protrusion) the directly preceding phrase. The strongest contrasts were measured for lip protrusion and the smallest for lip opening.

Analysis of the other facial data
Eyebrow movement (raising) – There appeared to be a link between eyebrow raising and the production of prosodic contrastive focus only for three out of five speakers (B, C & E). This eyebrow raising was however not systematic and did not occur whenever focus is produced. Speaker B was the one for which the combined productions (synchronous eyebrow raising and focus production) were the most

frequent. The amplitudes of the movements are however very small (largest movement: 2mm). The other speakers either never raised their eyebrows, or did it randomly with no particular link to focus production.

Table 1: Results from the Optotrak study for each speaker (B, C, D, E and F) and each parameter: duration (Dur), vertical jaw movements (Jaw), lip opening (LO), lip spreading (LS) and upper lip protrusion (LP). The table provides the percent changes for the focused phrase compared to the rest of the utterance (intra-utterance contrast) and to the same phrase in the neutral version (inter-utterance contrast). The statistical significance threshold was fixed at p=0.01. 'ns' corresponds to a non-significant change.

				B	C	D	E	F
Intra-utterance contrast		**Dur**		+38.7%	+30.5%	+25.3%	+16.8%	+43.8%
		Jaw		+9.9%	+14.9%	+12.5%	+31.5%	+18.5%
		LO		+13.8%	+16.8%	+5.1%	+17.4%	+13.5%
		LS		ns	+17%	ns	+19.4%	+23.2%
		LP		+62.8%	+25%	+37.4%	+39.4%	+69.1%
Inter-utterance contrast		**Dur**	**foc**	+20.9%	+34.1%	+29.8%	+23.9%	+49%
			pre-foc	ns	+4.7%	+8.1%	+16.4%	+9%
			post-foc	ns	ns	ns	ns	ns
		Jaw	**foc**	+6.6%	+14%	+9.9%	+28.5%	+39.5%
			pre-foc	ns	+6%	ns	ns	ns
			post-foc	-6.5%	-8.7%	ns	-6.2%	+23.9%
		LO	**foc**	+8.7%	+13.9%	+7.6%	+13.3%	+13.6%
			pre-foc	ns	+6.3%	+5.7%	-4.1%	ns
			post-foc	-8.9%	-2.2%	-4.5%	-3.4%	ns
		LS	**foc**	ns	+17.3%	ns	+14.5%	+30%
			pre-foc	ns	ns	ns	ns	ns
			post-foc	-15.6%	ns	ns	-5.8%	ns
		LP	**foc**	+98.4%	+30.9%	+32%	+46%	+112.4%
			pre-foc	+25.1%	ns	ns	+24.4%	+53.7%
			post-foc	-8.1%	ns	-10%	-13%	+41.6%

Head movement – Speaker B was the only one for whom a correlation between head nods and focus production was observed. The correlation

was not systematic, however, and the amplitudes and temporal alignment of the movements were highly variable. The other speakers also moved their heads but the movements were not correlated with focus production.

Perception tests
Visual perception tests were carried out on a selection of recordings from four of the speakers. These showed that focus could be correctly detected visually in 66.4% of the cases (chance level 33.3%) and suggested that the articulatory and facial visual cues identified above are well perceived visually.

c. Summary of findings on the articulatory signalling of focus
The production study described above along with that described in the previous section show that there are potential visible articulatory correlates to prosodic contrastive focus in French. One of our main conclusions is the fact that focus affects the whole utterance and not only the specific focused phrase. A number of articulatory gestures are affected by focus. The way and the extent to which these articulatory gestures are affected depend on the speaker. However, two main strategies emerge, an absolute one and a differential one:
Absolute visual signalling strategy: the focused constituent is lengthened and hyper-articulated to a large extent (inter-lip area, protrusion and jaw movements). Previous studies (Dohen & Lœvenbruck, 2005) showed that the peak velocities were also increased which signals an increase of the underlying articulatory effort during the gestures (Nelson, 1983). The speakers using this strategy therefore concentrate their efforts on the hyper-articulation of the focused phrase. Some speakers also slightly anticipate focus.
Differential visual signalling strategy: in this case, the focused phrase is also lengthened and hyper-articulated but to a smaller extent. Focus is sometimes anticipated. Additionally, the post-focal sequence is hypo-articulated compared to the neutral case. A visible contrast is thus created within the utterance: the focal hyper-articulation is not very distinct but is reinforced by the post-focal hypo-articulation.
We further observed that the visible articulatory parameter that was the most hyper-articulated under focus was protrusion. This is consistent

with the finding that lip protrusion is the most visible lip feature (Benoît et al., 1994). It is a robust feature which is long-anticipated in the articulation of the preceding phonemes (Cathiard, 1994). We also found that there could be a link between prosodic contrastive focus and head (nod) and/or eyebrow (raising) movements. However this link is far from being systematic, particularly for the head movements. There are important inter- and intra-speaker variations in the movement amplitude and in its synchronisation with the acoustic signal.

d. Contribution of articulatory and facial visual cues in global perception

Combined auditory-visual perception studies detailed in Dohen & Lœvenbruck (under revision) analysed the potential contribution of visual information in global auditory-visual perception of prosodic contrastive focus in French. Whispered speech was used to 'naturally' degrade part of the acoustic prosodic cues (no fundamental frequency for whispered speech) and thereby the auditory only perception performances. This rendered the measurement of a potential visual advantage possible by avoiding a ceiling effect. The studies showed that the visual information available from articulatory and facial movements was combined to auditory information especially when auditory information was not sufficient to make a straight-forward perceptual judgment (whispered speech). It therefore appears that articulatory and facial cues to prosodic focus are not only produced but that these cues can be used in the perception process.

Altogether, these results suggest that speakers use reliable strategies to organise phonation and articulation adequately in order to convey prosodic pointing. On the perception side, the visual only and auditory-visual perception tests show that these phonatory and articulatory patterns are recovered by listeners/viewers and used for prosodic pointing detection. Our conjecture is that integrated (acoustic, articulatory, proprioceptive) representations may be needed in order to produce prosodic pointing adequately and that these representations may be formed via the activation of associative cerebral areas, such as temporal and/or parietal regions. The fact that manual and ocular pointing both recruit the (associative) posterior parietal cortex adds to the motivation to explore the cerebral correlates of prosodic pointing.

2.3. CEREBRAL CORRELATES OF VOCAL POINTING

2.3.1. PRODUCTION FMRI STUDY

The fMRI production study summarized below is described in more details in Lœvenbruck et al. (2005). The aim was to examine the cerebral correlates of vocal pointing in French, conveyed by prosodic focus and syntactic extraction. Sixteen healthy, male, right-handed native speakers of French were examined. The stimuli consisted of visually presented sentences in French. Three isosyllabic sentences were presented, one for each condition:
- baseline condition: "Madeleine m'amena"
(Madeleine brought me around).
- prosodic deixis condition: "MADELEINE m'amena"
(MADELEINE brought me around), to elicit contrastive focus on the agent.
- syntactic deixis condition: "C'est Mad'leine qui m'am'na"
(It's Mad'leine who brought me 'round).
The number of syllables in the sentence was maintained equal to 6, using schwa deletion. The methods are described in Appendix A.3.
The results of the fixed effect group and random effects analyses are reported here. Figure 6 represents the functional activations obtained for the main effects with the fixed effect analysis. The pattern of activations common to the two deixis conditions (each compared to the baseline) included Broca's region (BA 45, 47), the left insula and the premotor cortex (BA 6) bilaterally. Prosodic deixis additionally activated the left anterior cingulate gyrus (BA 24, 32), the left supramarginal gyrus (LSMG, BA 40) and the left postero-superior temporal gyrus (Wernicke's area, BA 22). The (prosodic deixis - syntactic deixis) contrast yielded significant activation in the left posterosuperior temporal gyrus and the LSMG. The results of the random effect analysis, using the same statistical significance threshold ($p < .001$ corrected), for the same contrasts did not provide significant activations. With a less stringent significance threshold however ($p<.05$ non corrected), the contrasts provide a similar pattern of activations as the one obtained with the fixed effect analysis.
This fMRI study shows activation of Broca's region in the deixis

conditions compared to the baseline. The Left Inferior Frontal Gyrus (LIFG) was therefore activated during verbal pointing at the agent of the action, through prosody or syntax. This activation is consistent with functional neuroimaging studies on complex syntactic processing (Caplan et al., 2000; Friederici, 2002; Just et al., 1996). These studies have shown the involvement of the LIFG in plausibility judgments about syntactically complex constructions, which require intricate tracking of thematic roles. Our findings are also in line with studies on the observation and mental imagery of action which show LIFG activation during action tracking (Binkofski et al., 2000; Grafton et al., 1996; Iacoboni et al., 1999; Rizzolatti et al., 2007). According to Rizzolatti and colleagues, the role of the LIFG in speech would have evolved from a 'basic mechanism originally not related to communication: the capacity to recognize actions' (Rizzolatti & Arbib, 1998). In addition, Dogil and colleagues' fMRI study on the production of prosodic features at the syllable and phrase levels has also revealed left IFG activation (Mayer et al., 2002; Dogil et al., 2002). Taken together, these observations support the claim that the role of the LIFG is that of an action-structure parser, which, in morphosyntactic encoding and decoding, handles the parsing of the predicate and its arguments, or the attentional monitoring of "who does what to whom".

The left insula was also found activated in both (deixis – baseline) contrasts. The involvement of the left precentral gyrus of the insula in articulatory planning during speech has already been shown (Dronkers, 1996). As described above, prosody has acoustic and articulatory correlates. The production of prosodic focus may require more accurate planning of the movements of the larynx, the tongue and the jaw, which could explain why the prosodic deixis condition yields significant activation of the left insula when compared with the baseline (same words to articulate, but a more stringent prosody). Similarly, the syntactic deixis condition (compared to the baseline) likely requires more accurate articulatory planning, given the larger number of consonant clusters involved (due to schwa deletion).

The activation of the LSMG and of Wernicke's area in the prosodic deixis condition alone suggests that, when deixis is already encoded by syntax, no additional recruitment of the inferior parietal lobule and Wernicke's area is necessary. As mentioned in section 2, several studies suggest that

the inferior parietal regions in both hemispheres function as sensory integrators which form representations necessary in the organisation of motor actions, such as linguistic or non-linguistic pointing at targets. The left hemisphere would have a linguistic predominance.

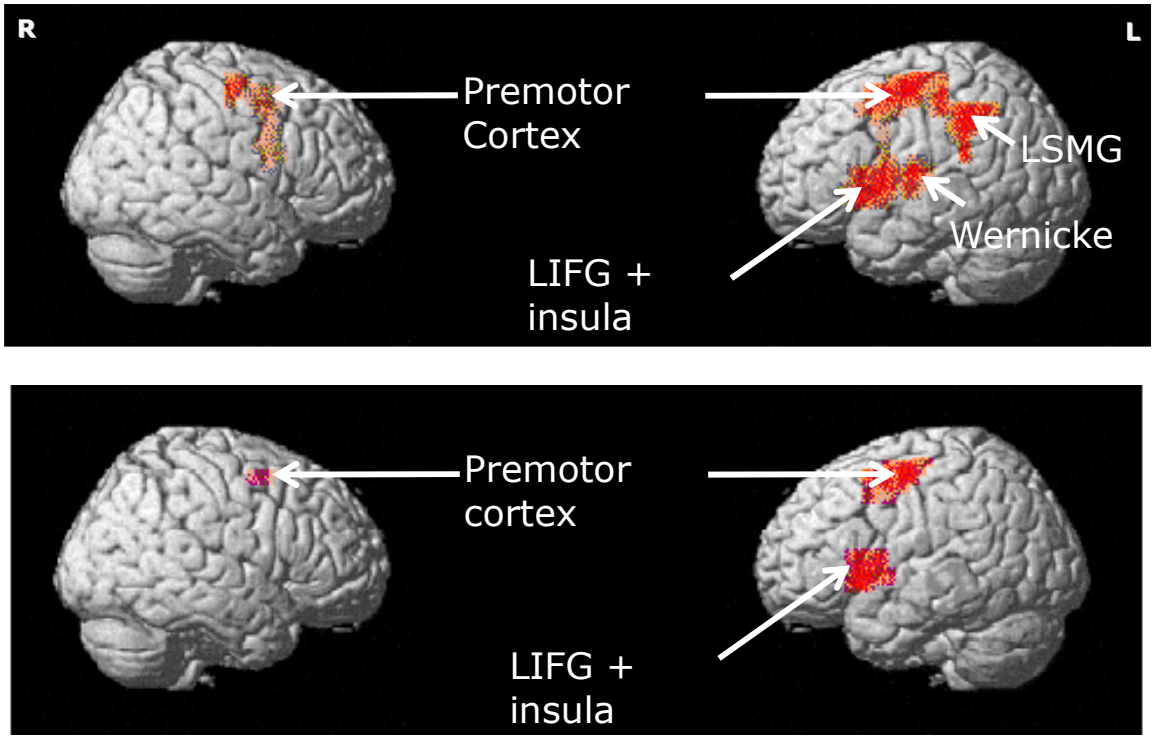


Figure 6: Activations in the prosodic deixis contrast (top) and syntactic deixis contrast (bottom). Both contrasts provide activations of frontal regions. Only the prosodic contrast shows temporal and parietal activations.

In speech, a left temporo-parieto-frontal network might be recruited in the organisation of verbal motor actions from auditory representations. Our results, with the activations of the LSMG, the LIFG and Wernicke's area in prosodic deixis, are in line with this hypothesis. Like visually-guided manual pointing, prosodic pointing may need multisensory representations to be formed via superior temporal and inferior parietal regions to organise articulation and phonation adequately. We therefore suggest that non-grammaticalized linguistic deixis recruits the temporo-parieto-frontal network and that grammaticalized deixis (syntactic deixis) is handled solely by the LIFG.

2.3.2. PERCEPTION FMRI STUDY

The production study described above suggests that associative brain areas are recruited for the production of vocal pointing. As demonstrated in section 2.2., the articulatory and phonatory patterns produced by speakers are also perceived by listeners/viewers. It can therefore be hypothesised that the cerebral network recruited for the production of vocal pointing is at least partly also recruited for its perception. This is what the study presented hereafter aims at testing.

Although, the acoustic and articulatory correlates of prosodic focus have been quite extensively studied (as recalled in section 2.), it remains unclear what neural processes underlie its perception. Meanwhile studies have shown that prosodic processing in general cannot be restricted to the right hemisphere (see Baum & Pell, 1999 for a review). Two studies have analysed the processing of prosodic focus (or closely related prosodic phenomena). The first one (Wildgruber et al., 2004) aimed at contrasting affective vs linguistic prosody. The linguistic prosodic task was an indirect informational focus detection task (find the most suitable answer to a specific question). For the linguistic prosodic task, the authors found bilateral activations of the primary and secondary auditory cortices, of the anterior insular cortex and of the frontal operculum (BA 6/44/47), right hemisphere activation of the dorso-lateral-frontal regions and left hemisphere activations of the inferior frontal cortex. The second study which examined the processing of prosodic focus (Tong et al., 2005) aimed at differentiating the processing of 'intonation' (question/affirmation discrimination) and that of contrast-ive stress. It additionally compared English and Chinese. For the processing of contrastive stress, the authors put forward bilateral activation of the intra-parietal sulcus (BA 40/7), right hemisphere activation of the medial frontal gyrus (BA 9/46) and left hemisphere activations of the supramarginal gyrus and the posterior medio-temporal gyrus (BA 21/20/37).

Moreover, as we have shown above (in section 2.2.), 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

(Dohen & Lœvenbruck, under revision). This finding emphasises the necessity to consider the perception of prosodic contrastive focus and speech prosody in general as multimodal. The perception fMRI study presented here aims at analysing the neural processing of prosodic focus from a multimodal point of view.
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. After seeing/hearing/seeing and hearing each stimulus, subjects were asked to tell whether they had perceived a correction (i.e. contrastive focus) or not. A detailed description of the stimuli and of the fMRI procedure (paradigm and data collection and analysis) can be found in Appendix A.4.

Behavioral results
Table 2 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 (the percentages of correct answers were well above chance, in all conditions).

Table 2: 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	2.0	89.6	6.5
A	97.4	3.0	69.9	12.8
V	86.4	7.5	88	8.6

Preliminary fMRI results
The analysis of the fMRI data is still underway and the results presented here are only preliminary. A preliminary analysis was conducted for the

focus vs no focus contrast for the auditory and auditory-visual modalities. It appeared that auditory alone (A) detection of prosodic focus involved the right associative auditory cortex and fusiform gyrus (BA 19) as well as the left middle frontal gyrus (BA 6/46) and inferior temporal gyrus (BA 37) and the cerebellum bilaterally. For the auditory-visual modality, we found bilateral activations of the middle and inferior frontal gyri (BA 40), the middle temporal gyrus (BA 21), the inferior parietal lobule (BA 40) and the fusiform gyrus (BA 37) as well as left activation of the supramarginal gyrus (BA 40).
It appears that for all modalities, prosodic focus detection or processing involves bilateral activations of associative brain areas. Auditory perception of prosodic focus (vs no focus) appears to be essentially processed in associative areas right superior temporal gyrus and left inferior temporal gyrus (BA 37). Multimodal (AV) perception of prosodic focus involves bilateral activations of temporal and parietal associative areas as well as inferior and middle frontal regions. This illustrates the underlying necessity of associating various types of information to detect focus (especially auditory and articulatory) and supports the assumption that the articulatory and phonatory patterns produced by the speaker are integrated in perception.
Similar activations of a complex neural network in the processing of focus have been observed in two recent ERP studies (Bornkessel et al., 2003; Magne et al., 2005). The implication of the left parietal lobe in the auditory-visual perception of prosodic pointing is interesting, in the light of our fMRI study of the production of prosodic pointing, and of the studies on manual and ocular pointing.

3. MULTIMODAL POINTING

The aim of this multimodal fMRI study was to investigate the cerebral correlates of pointing in several modalities (Lœvenbruck et al., 2007). To test the hypothesis that multimodal pointing could involve a continuum of cerebral regions, we designed an fMRI paradigm including four conditions: 1) index finger pointing, 2) eye pointing, 3) prosodic focus, 4) syntactic extraction.

Subjects and material

Fifteen healthy right-handed volunteers, aged 18-55 years, native speakers of French were examined. Stimuli consisted of two types (1 and 2) of images consisting of a girl (Lise) and a boy (Jules) next to each other and alternatively located on the right and on the left side of the screen. Type 1 images showed the girl holding a book, while the boy did not; Type 2 images showed the reverse. In the middle of all images, a fixation cross was displayed. A blank screen with a mid-centered black cross preceded each stimulus. The tasks consisted of verbal and non-verbal pointing to a character. Control conditions were included. The same question: "Est-ce que Lise tient le livre?" (Does Lise hold the book?) was used for all tasks. Subjects were instructed: (a) to confirm the question when type 1 images were presented (control), (b) to point to Jules when type 2 images were presented (contrastive pointing). The tasks were the following: During *prosodic pointing*, subjects uttered «JULES tient le livre» (JULES holds the book), with contrastive focus on "Jules". During *syntactic pointing*, they uttered «C'est Jules qui tient le livre» (It is Jules who holds the book), with syntactic extraction of "Jules". During *digital pointing*, they pointed with the right index finger to Jules. During *ocular pointing*, subjects looked in the direction of Jules.

The controls were the following: In prosodic and syntactic controls, subjects neutrally uttered: «Lise tient le livre » (Lise holds the book). In digital control, they performed a downward finger movement. In ocular control, they made a downward eye movement. The detailed methods are presented in Appendix A.5.

Preliminary fMRI results

Patterns of activation were examined for 15 subjects for each of the following contrasts:

- digital pointing tasks vs digital control
- ocular pointing tasks vs ocular control
- prosodic pointing tasks vs prosodic control
- syntactic pointing tasks vs syntactic control

Preliminary results of the random effect group analysis are reported. Figure 7 represents functional activations during each pointing condition vs its control.

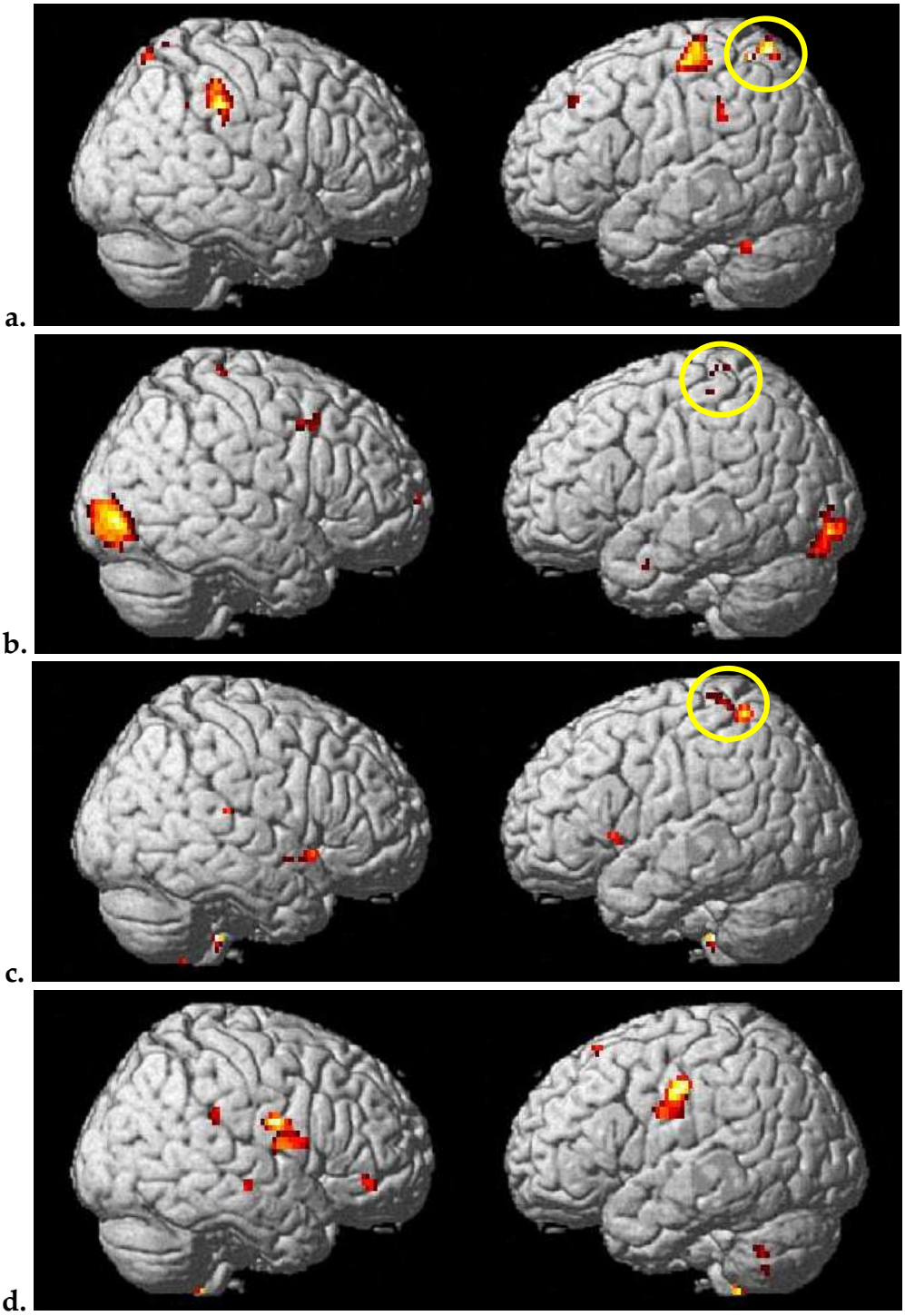


Figure 7: Activations during each pointing condition vs its control, rendered on a sagittal template, in a. digital- b. ocular- c. prosodic- d. syntactic-modes. The left hemisphere is on the right, the right one on the left

Digital pointing (vs its control) yielded bilateral activation of the superior parietal lobule (BA 7), with a left hemisphere dominance. There was bilateral activation of the supramarginal gyrus (BA 40), with a right

hemisphere dominance. There was left frontal cortex activation (BA 4 ,6). *The ocular condition* activated the left hemisphere predominantly in the post central gyrus (BA 3) and the frontal lobe bilaterally (BA 4, 6). The occipital gyrus (BA 17, 18) was activated bilaterally. *The prosodic condition* recruited the superior parietal lobule (BA 7) bilaterally, with a left dominance. The supramarginal gyrus activation did not reach significance, contrary to our previous finding on covert prosody (see section 2.3.1.). The left postcentral gyrus was activated (BA 2, 3). Perisylvian regions (BA 47, 13, 42) were activated bilaterally. *The syntactic condition* activated the right supramarginal gyrus (BA 40) and did not recruit the left parietal lobule. Right perisylvian regions (BA 44, 47, 21) were activated, as well as the premotor cortex (BA 6) bilaterally.

To summarise, this preliminary analysis revealed that the left superior parietal lobule (BA 7 or its neighbour BA 3) was activated in all three digital, ocular and prosodic pointing but not in syntactic pointing. These results indicate that pointing in different modalities may recruit the left superior parietal lobule, ocular pointing being more anterior than prosodic pointing, itself more anterior than digital pointing.

These results are in accordance with the results from our behavioural studies which suggest that speakers may use multisensory represent-tations in order to produce prosodic pointing adequately, just like they do to produce a manual or ocular gesture. These representations may be formed via the activation of associative parietal areas. The lack of parietal activation in syntactic pointing could be due to the absence of on-line multisensory construction in syntactic pointing, compared with prosodic pointing. Syntactic pointing uses a grammaticalized "frozen" construction and may not need integrated representations to be formed. This fMRI study therefore suggests that the left superior parietal lobule could be a potential substrate for pointing with the finger, the eye, and the larynx.

4. CONCLUSION

We have shown that pointing is a critical device to explore the potential

common neural networks at play in gesture and language. The role of manual (or digital) pointing in language acquisition strongly suggests indeed that vocal pointing and pointing in other modalities may well be grounded in a common cerebral network.

We have argued from several neuroimaging studies on manual and ocular pointing that these two modalities seem to recruit a network including the left posterior parietal and frontal cortices.

Behavioural studies in our laboratory have shown that prosodic vocal pointing is a sophisticated behaviour that requires accurate laryngeal and articulatory control. Perception studies have revealed that the potential articulatory visible cues to prosodic vocal pointing are in fact visually perceived. In other words, viewers can decipher these sophisticated signals. We have claimed that speakers may need multisensory representations to produce prosodic pointing adequately and that these representations may involve the activation of associative parietal regions, close to the regions involved in manual and ocular pointings.

The results of a first study on covert vocal pointing, including prosodic pointing (i.e. focus) and syntactic pointing (i.e. syntactic extraction) have been presented. It was shown that covert prosodic pointing does recruit left parietal regions, whereas syntactic pointing mainly involves frontal region. The involvement of the left parietal lobe was confirmed by a recent fMRI study of the perception of prosodic pointing. Finally we presented the preliminary results of a new study of multimodal pointing (digital, ocular, overt prosodic and syntactic pointing). They seem to reveal a common pattern of left parietal activation in ocular, digital and prosodic pointing. A grammaticalization process has been suggested to explain the lack of parietal activation in syntactic pointing. Altogether these fMRI studies are in line with our conjecture that linguistic on-line pointing (prosodic focus) is grounded in the same cerebral network as gestural (manual and ocular) pointings. More analyses are underway to consolidate these results. If they come out as robust, then they will contribute to the debate on the gestural origin of language.

ACKNOWLEDGEMENTS

Many researchers were involved in the different works presented here.

In alphabetical order, they are C. Abry, M. Baciu, A. Callan, D. E. Callan, F. Carota, M.–A. Cathiard, H. Hill, L. Lamalle, P. Perrier, C. Pichat, C. Savariaux, J.-L. Schwartz, C. Segebarth. We thank them all sincerely.

REFERENCES

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.

Astésano C. (2001). *Rythme et accentuation en français: invariance et variabilité stylistique*, PhD Thesis, L'Harmattan édition et diffusion.

Astésano C., Magne C., Morel M., Coquillon A., Espesser R., Besson M. & Lacheret A. (2004). Marquage acoustique du focus contrastif non codé syntaxiquement en français. *Proceedings of the XXVièmes Journées d'Etudes sur la Parole*, 126–129. Fez, Morocco.

Astésano C., Bard E. G. & Turk A.. (2007). Structural influence on initial accent placement in French. *Language and Speech*, 50 (3), 423-446.

Audouy, M. (2000). *Traitement d'images vidéo pour la capture des mouvements labiaux.* Final engineering report, Institut National Polytechnique de Grenoble.

Bates, E., Camaioni, L. & Volterra V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly*. 21: 205-226.

Bates E., Benigni L., Bretherton I., Camaioni L., Volterra V. (1979). *The emergence of symbols: cognition and communication in infancy.* New York: Academic Press.

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

Bellugi U., Poizner H. & Klima E. S. (1989). Language, modality and the brain. *Trends in Neurosciences*, 12 (10), 380-388.

Benoît C., Mohamadi T. & Kandel S. (1994). Effects of Phonetic Context on Audio-Visual Intelligibility of French. *Journal of Speech and Hearing Research*, 37, 1195-1203.

Berthoud A.-C. (1990). Deixis, thématisation et détermination. *La deixis. Colloque en Sorbonne.* M.-A. Morel & L. Danon-Boileau (dir.), PUF, 527-542.

Binkofski F., Amunts K., Stephan K. M., Posse S., Schormann T., Freund H.-J., Zilles K. & Seitz R. (2000). Broca's region subserves imagery of motion: a combined cytoarchitectonic and fMRI study. *Hum. Brain Mapp.*, 11, 273-285.

Bornkessel I., Schlesewsky M., & Friederici A. D. (2003). Contextual information modulates initial processes of syntactic integration: The role of inter- versus

intrasentential predictions. *Journal of Experimental Psychology: Learning Memory and Cognition*, 29, 269-298.

Brown H. D. (1971). Children's comprehension of relativized English sentences. *Child Development,* 42, 1923–36.

Bruner, J. (1975). The ontogenesis of speech acts. *Journal of Child Language*, 2, 1– 19.

Butcher C. & Goldin-Meadow S. (2000). Gesture and the transition from one- to two-word speech: when hand and mouth come together. In D. McNeill (Ed.), *Language and gesture*. Cambridge: Cambridge University Press: 235-257.

Butterworth G. (2003). Pointing is the royal road to language. *Pointing, where language, culture, and cognition meet.* Kita S. (ed.). Lawrence Erlbaum Associates Publishers, Mahwah, NJ, 9-34.

Capirci O., Iverson J. M., Pizzuto E., & Volterra V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child language*, 23, 645- 673.

Caplan D., Alpert N., Waters G. & Olivieri A. (2000). Activation of Broca's area by syntactic processing under conditions of concurrent articulation. *Hum. Brain Mapp.*, 9, 65-71.

Caselli M.C. (1990). Communicative gestures and first words. In V. Volterra & C.J. Erting (Eds.), *From gesture to language in hearing and deaf children*. New York: Springer-Verlag, pp. 56-67.

Cathiard M.-A. (1994). *La perception visuelle de l'anticipation des gestes vocaliques : cohérence des évènements audibles et visibles dans le flux de la parole*, Unpublished doctoral dissertation, Université Pierre Mendès France, Grenoble.

Cavé, C.; Guaïtella, I.; Bertrand, R.; Santi, S.; Harlay, F.; Espesser, R. (1996). About the Relationship between Eyebrow movements and F0 Variation. *Proceedings of ICSLP 96*, 4, 2175-2179.

Cerrato, L.; Skhiri, M. (2003). Analysis and measurement of communicative gestures in human dialogues. *Proceedings of AVSP 2003*, 251-256.

Cheney D. L. & Seyfarth R. M. (1981). Selective forces affecting the predactor alarm calls of vervet monkeys. *Behaviour*, 76, 25-61.

Chomsky N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky N. (1966). *Cartesian linguistics*. New York: Harper Row.

Connaghan K. P., Moore C. A., Reilly K. J., Almand K. B., & Steeve R. W. (2001). Acoustic and physiologic correlates of stress production across systems, Poster presented at the *American Speech-Language-Hearing Association*, New Orleans, L. A.

Connolly J. D., Goodale M. A., Desouza J. F., Menon R. S., Vilis T. (2000). A comparison of frontoparietal fMRI activation during anti-saccades and antipointing. *J Neurophysiol*., 84, 1645–1655.

Corballis, M. C. (1991). *The lopsided ape: Evolution of the generative mind*. New York: Oxford University Press; 1991.

Dahan D. & Bernard J.-M. (1996). Interspeaker Variability in Emphatic Accent Production in French. *Language and Speech,* 39(4), 341-374.

Delais-Roussarie E., Rialland A., Doetjes J. & Marandin J.-M. (2002). The Prosody of Post Focus Sequences in French. In *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, 239-242.

Di Cristo A. (1998). Intonation in French. In D. Hirst; A. Di Cristo (Eds.), *Intonation Systems: a Survey of Twenty Languages*, 195-218. Cambridge University Press.

Di Cristo A. & Jankowski L. (1999). Prosodic Organisation and Phrasing after Focus in French. *Proceedings of ICPhS 1999*, San Francisco, USA, 1565-1568.

Diessel H. & Tomasello M. (2000). The development of relative clauses in spontaneous child speech. *Cognitive Linguistics*, 11, 131-151.

Dogil G., Ackermann H., Grodd W., Haider H., Kamp H., Mayer J., Riecker A., Wildgruber D. (2002). The Speaking Brain. *Journal of Neurolinguistics*, 15 (1), 59-90.

Dohen M., Lœvenbruck H., Cathiard M.-A. & Schwartz J.-L., 2004. Visual perception of contrastive focus in reiterant French speech. *Speech Communication*, 44, 155-172.

Dohen M. & Lœvenbruck H. (2004). Pre-focal Rephrasing, Focal Enhancement and Post-focal Deaccentuation in French. In *Proceedings of the 8th International Conference on Spoken Language Processing* (ICSLP 04), Jeju island (Korea), 4-8 October 2004, Vol. 1, 785-788.

Dohen, M.; Lœvenbruck, H., (2005). Audiovisual Production and Perception of Contrastive Focus in French: a multispeaker study. *Proceedings of Interspeech 2005*, 2413-2416.

Dohen M., Lœvenbruck H. & Hill H. (2006). Visual correlates of prosodic contrastive focus in French: description and inter-speaker variability. *Proceedings of Speech Prosody 2006*, Dresden, Germany, May 2-5 2006, Vol. I, 221-224.

Dohen M. & Lœvenbruck H. (under revision). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech*.

Dronkers N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, 384, 159-161.

Edwards M. G. & Humphreys G. W. (1999). Pointing and grasping in unilateral visual neglect: effect of on-line visual feedback in grasping. *Neuropsychologia*, 37 (8), 959-973.

Erickson D. (submitted). Effects of contrastive emphasis on jaw opening.

Fillmore C. J. (1997). *Lectures on deixis.* Stanford: Center for the Study of Language and Information.

Fougeron C. & Keating P. A. (1997) Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustic Society of America*, 101, 3728-3740.

Friederici A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6 (2), 78-84.

Goffman L. & Malin C. (1999). Metrical Effects on Speech Movements in Children and Adults. *Journal of Speech, Language, and Hearing Research*, 42 (4), 1003-1015.

Goldin-Meadow S. & Butcher C. (2003). Pointing toward two-word speech in young children. In *Pointing: Where language, culture, and cognition mee.t* S. Kita (Ed.), 85-107. Mahwah, NJ: Earlbaum Associates.

Grafton S. T., Arbib M. A., Fadiga L. & Rizzolatti G. (1996). Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Exp. Brain Res*., 112 (1), 103-11.

Graf, H.P.; Cosatto, E.; Strom, V.; Huang, F.J. (2002). Visual Prosody: Facial Movements Accompanying Speech. *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition* (FGR'02), 381-386.

Granström, B.; House, D. (2005). Audiovisual representation of prosody in expressive speech communication. *Speech Communication*, 46, 473-484.

Hadar U., Steiner T. J., Grant E. C., Rose F. C. (1983). Head movement correlates of juncture and stress at sentence level. *Language and Speech,* 26, 117–129.

Hagler D. J. Jr, Riecke L., Sereno M. I. (2007). Parietal and superior frontal visuospatial maps activated by pointing and saccades. *NeuroImage*, 35 (4), 1562-1577.

Hewes G. W. (1981). Pointing and language. In T. Myers, J. Laver & J. Anderson (eds.). *The cognitive representation of speech*, 263-269. Amsterdam: North-Holland.

Hickok G. & Poeppel D. (2001). Towards a functional neuroanatomy of speech perception. T*rends In Cognitive Sciences*, 4 (4), 131-138.

Higginbotham, D. (1992). Reference and control. In: R.K. Larson, S. Iatridou, U. Lahiri & J. Higginbotham (eds.). *Control and grammar*. Dordrecht: Kluwer, 79-108.

Hornby P. A. & Hass W. A. (1970). Use of contrastive stress by preschool children. *Journal of Speech and Hearing Research*, 13, 395-399.

Iacoboni M., Woods R. P, Brass M., Bekkering H., Mazziotta J. C. & Rizzolatti G. (1999). Cortical mechanisms of human imitation. *Science*, 286, 2526-2528.

Iacoboni M., Koski L. M., Brass M., Bekkering H., Woods R. P., Dubeau M.-C., Mazziotta J; C. & Rizzolatti G. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. Proc. *Natl. Acad. Sci. USA*, 98 (24), 13995-13999.

Jackendoff R. (2002). Foundations of Language - Brain, Meaning, Grammar, Evolution. Oxford University Press.

Jisa H. & Kern S. (1998). Relative clauses in French children's narrative texts. *Journal of Child Language,* 25, 623-652.

de Jong K., Beckman M. E. & Fletcher J. (1991). The articulatory kinematics of final lengthening. *J. Acoust. Soc. Am.*, 89, 369-382.

Jun S.-A. & Fougeron C. (2000). A Phonological Model of French Intonation. In A. Botinis (Ed.), *Intonation: Analysis, modelling and technology,* 209-242. Dordrecht: Kluwer Academic Publishers.

Just M. A., Carpenter P. A., Keller T. A., Eddy W. F. & Thulborn K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, 274, 114-116.

Kendon A. (1996). An Agenda for Gesture Studies. *Semiotic Review of Books*, 7(3), 8-12.

Kendon A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kertzman C., Schwarz U., Zeffiro T. A. & Hallett M. (1997). The role of posterior parietal cortex in visually guided reaching movements in humans. *Exp. Brain Res.*, 114 (1), 170-183.

Kidd E. & Bavin E. L. (2002). English-speaking children's comprehension of relative clauses: Evidence for general-cognitive and language-specific constraints on development. *Journal of Psycholinguistic Research,* 31, 599–617.

Kita S. (2003). *Pointing: where language, culture, and cognition meet*. Lawrence Erlbaum Associates Publishers, Mahwah, NJ.

Klima E. S. & Bellugi U. (1988). *The signs of language*. Harvard University Press.

Konopscynsky G. (1986). *Du prélangage au langage : acquisition de la structuration prosodique.* PhD Thesis, University of Strasbourg.

Krahmer, E.; Ruttkay, Z.; Swerts, M.; Wesselink, W. (2002). Pitch, Eyebrows and the Perception of Focus. *Proceedings of Speech Prosody 2002*, 443-446.

Lacquaniti F., Perani D., Guigon E., Bettinardi V., Carrozzo M., Grassi F., Rossetti Y. & Fazio F. (1997). Visuomotor transformations for reaching to memorized targets: a PET study. *Neuroimage*, 5 (2), 129-146.

Leavens D. A., Hopkins W. D. & Bard K. A. (2005). Understanding the Point of Chimpanzee Pointing. Epigenesis and Ecological Validity. *Current Directions in Psychological Science*, 14, 185-189.

Liberman A. M. (1982). On Finding that Speech Is Special. *American Psychologist*, 37, 2, 148-167.

Lœvenbruck H. (1996). *Pistes pour le contrôle d'un robot parlant capable de réduction vocalique.* Unpublished Doctoral Dissertation in Cognitive Sciences, INPG, Grenoble.

Lœvenbruck H. (1999). An investigation of articulatory correlates of the Accentual Phrase in French. *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1, 667-670, San Francisco, USA.

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.

Lœvenbruck H., Vilain C., Carota F., Baciu M., Abry C., Lamalle L., Pichat C. & Segebarth C. (2007). Cerebral correlates of multimodal pointing: An fmri study of prosodic focus, syntactic extraction, digital- and ocular- pointing. *Proceedings of the XVIth International Congress of Phonetic Sciences*, 6-10 August 2007, Saarbruecken.

Magne C., Astesano C., Lacheret-Dujour A, Morel M., Alter K., Besson M. (2005). On-Line processing of "pop-out" words in Spoken French dialogues. *Journal of Cognitive Neurosciences*, 17(5), 740-756.

Macwhinney B., & Ple´h C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition,* 29, 95–141.

Mayer J., Wildgruber D., Riecker A., Dogil G., Ackermann H. & Grodd W. (2002). Prosody production and comprehension: converging evidence from fMRI studies. *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, 11-13 April 2002, 487-490.

McNeill, D. (1992). *Hand and Mind*. Chicago: University of Chicago Press.

Mckee C., Mcdaniel D. & Snedeker J. (1998). Relatives children say. *Journal of Psycholinguistic Research,* 27, 573–96.

Ménard L., Lœvenbruck H. & Savariaux C. (2006). Articulatory and acoustic correlates of contrastive focus in French: a developmental study, in Harrington, J. & Tabain, M. (eds), *Speech Production: Models, Phonetic Processes and Techniques, Psychology Press* : New York, 227-251.

Morford M. & Goldin-Meadow S. (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *Journal of Child Language*, 19 (3), 559-580.

Munhall K.G., Jones J. A., Callan D. E., Kuratate T., Vatikiotis-Bateson E. (2004). Visual Prosody and Speech Intelligibility – Head Movement Improves Auditory Speech Perception. *Psychological Science*, 15(2), 133-137.

Nelson, W.L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics,* 46(2), 135-147.

Oldfield R.C. (1971). The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia*, 9(97), 113.

Pollick A. S. & de Waal F. B. M. (2007). Ape gestures and language evolution. *P. N. A. S.*, 104, 19,8184-8189.

Povinelli D. J. & Davis D. R. (1994). Differences between chimpanzees (Pan troglodytes) and humans (Homo sapiens) in the resting state of the index finger: Implications for pointing. *Journal of Comparative Psychology.* 108, 134–139.

Povinelli D.J., Bering J. M., & Giambrone S. (2003). Chimpanzee 'pointing': Another error of the argument by analogy? In *Pointing: Where language culture and cognition meet*. S. Kita (Ed.), 35-68, Lawrence Erlbaum Associates.

Rizzolatti G., Fogassi L., Gallese V. (1997). Parietal cortex: from sight to action. *Curr. Opin. Neurobiol*., 7 (4), 562-567.

Rolfe L. (1996). Theoretical stages in the prehistory of grammar. In A. Lock & C. R. Peters (eds.), *Handbook of human symbolic evolution*, 776-792. Oxford: Clarendon.

Rossi M. (1999). La focalisation. In *L'intonation, le système du français: description et modélisation*, Chap. II-6, 116-128. Ophrys.

Séguinot, A. (1976). L'accent d'insistance en français standard. *Studia Phonetica*, 12, 1-58.

Sheldon A. (1974). The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior*, 13, 272–81.

Simon O., Mangin J. F., Cohen L., Le Bihan D. & Dehaene S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron*, 33, 475–487.

Tabain M. (2003). Effects of prosodic boundary on /aC/ sequences: Articulatory results. *J. Acoust. Soc. Am*., 113 (5), 2834-2849.

Tavakolian S. (1981). The conjoined clause analysis of relative clauses. *Language acquisition and linguistic theory*, S. Tavakolian (ed.), 167–87. Cambridge, MA: MIT Press.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Journal of Child Language*, 34, 1-20.

Tong Y., Gandour J., Talavage T., Wong D., Dzemidzic M., Xu Y., Li X. & Lowe M. (2005). Neural circuitry underlying sentence-level linguistic prosody. *NeuroImage*, 28, 417-428.

Touati P. (1987). Structures prosodiques du suédois et du français. *Lund Working Papers* 21, Lund University Press.

de la Tour G. (1635). *Le Tricheur à l'as de carreau* (The cheat with the ace of diamonds), Musée du Louvre, Paris.

Vallar G., Guariglia C., Nico D. & Pizzamiglio L. (1997). Motor deficits and optokinetic stimulation in patients with left hemineglect. *Neurology*, 49 (5), 1364-1370.

Volterra V., Caselli M. C., Capirci O. & Pizzuto E. (2005). Gesture and the emergence and development of language. In *Beyond Nature-Nurture. Essays in Honor of Elizabeth Bates.* M. Tomasello & D. Slobin (eds.), Mahwah, NJ: Lawrence Erlbaum Associates. 3-40.

Wechsler, S. & Wayan, A. (1998). Syntactic ergativity in Balinese: an argument structure based theory. *Natural Language and Linguistic Theory* 16, 387-441.

Wildgruber D., Hertrich I., Riecker A., Erb M., Anders S., Grodd W. & Ackerman H. (2004). Distinct Frontal Regions Subserve Evaluation of Linguistic and Emotional Aspects of Speech Intonation. *Cerebral Cortex*, 14(12), 1384-1389.

Wilkins D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). *Pointing: where language, culture, and cognition meet*. Kita S. (ed.). Lawrence Erlbaum Associates Publishers, Mahwah, NJ, 171-216.

Wurmbrand, S. (2001). *Infinitives: restructuring and clause structure*. Berlin: Mouton de Gruyter.

APPENDICES

A.1. METHODS FOR THE EMA STUDY

Recordings

Simultaneous acoustic and articulatory recordings were collected using EMA (Carstens AG100), for a female native speaker of French. Five pellets were glued midsagittally to the apex, middle and dorsum of the tongue, and to the lower and upper incisors. The EMA data were sampled at 500 Hz, the acoustic data at 16000 Hz. The data were corrected for rotation, low-pass filtered and normalized by the reference pellet to correct for head movements. Velocity and acceleration traces were obtained using a finite difference method.

Corpus

The corpus consisted of several read sentences, containing 4-syllable target words, which constituted an Accentual Phrase. Position of the target word in the sentence was varied (initial, central, final), as in the example below for which the target word is "l'illuminée" (the crank):

'L'illuminée a allumé néanmoins le monument.' (Initial).

The crank has nevertheless lighted the monument.

'Il a humilié l'illuminée en l'éloignant.' (Central).

He humiliated the crank by moving her away.

'L'aumonier a néanmoins éloigné l'illuminée.' (Final).

The chaplain has nevertheless moved the crank away.

The corpus was designed so that analyses were made easier. To facilitate F0 tracking, only sonorants were used, and to obtain clear tongue movements, syllables containing vowels /i/ and /a/, which correspond to large tongue displacements in the horizontal and vertical dimensions, were selected. Several target words were pronounced. Results are presented for the target word: "*l'anomala*" (/la-no-ma-la/, beetle), which entails clear tongue movements on the initial and final syllables. As described in Jun & Fougeron (2000), in the contrastive focus rendition, initial and final syllables are potential spots for the F0 peak to occur.

The sentences were pronounced under two conditions: a neutral (broad focus) condition and a contrastive focus condition. Three speaking rates were produced: slow, normal, fast. Contrastive emphasis or focus was elicited on the target word as follows. Before each recording, a sentence was played to the subject, where the target word had been replaced by a wrong 4-syllable word. The subjects had to correct the sentence, placing contrastive focus on the target word.

Articulatory data analysis

Among the 8 traces (horizontal and vertical positions of the tongue -apex, -middle and -dorsum and of the lower incisor), the vertical position of the tongue-middle showed the most variation. It was thus chosen as the representative articulator. For each syllable, the tongue-middle vertical position at the time the vowel was fully reached, the peak velocity of the movement from the consonant to the vowel, and the duration of the syllable were measured, using hand-labelled events. The spectrogram of the acoustic signal, as well as velocity and acceleration traces, were used to mark these events. As shown in figure 2 for [**la**-no-ma-la], the beginning and end of the syllable were marked using the spectrogram (and listening to the signal) and provided the duration of the syllable. The tongue-middle minimum vertical position for /a/ was measured using zero-crossing of the velocity trace. The peak velocity was measured using zero-crossing of the acceleration trace.

A.2. METHODS FOR THE OPTOTRAK STUDY

For more details the reader can refer to Dohen et al. (2006).

Corpus
The corpus used for this study consisted of 13 subject-verb-object (SVO) structured sentences with CV syllables and, whenever possible, sonorants. Below is an example of one of the sentences used:
Lou ramena Manu. *Lou gave a lift back to Manu.*

Recordings
Four focus conditions were elicited: subject-, verb- and object-focus (narrow focus) and a neutral version (broad focus). In order to trigger contrastive focus, the speakers had to perform a correction task. They heard a prompt in which two speakers were talking and they were then asked to correct a phrase (S, V or O) which had been mispronounced. Two repetitions of each utterance were recorded.
The Optotrak system consists of three infrared (IR) cameras used to track the motion of infrared emitting diodes (IREDs) which in this case were glued to the speaker's face. The 3D coordinates of each IRED were automatically detected. For this experiment, we used two Optotraks in order to compensate for missing data. A total of 24 IREDs were glued to the speakers' faces. An additional 4 IREDs were attached to a head rig and were used to correct for head motion. IRED positions were sampled at 60 Hz and low-pass filtered. The acoustic signals were recorded simultaneously and sampled at 22 kHz.

Acoustic validation
The data recorded were first acoustically validated i.e. it was checked check whether focus had actually been produced acoustically. For all the speakers, we checked that the focused utterances displayed a typical focused intonation and that focus was well perceived auditorily (informal auditory perception tests).

Articulatory measurements
In our previous studies (see *a.*), two articulatory features had been analyzed, namely inter-lip area and protrusion. These parameters best represented the high segmental variability of speech and were the most relevant parameters to isolate supra-segmental originating variations. However, it is not possible to compute inter-lip area accurately from Optotrak data. This is why, in this study, we separately analyzed lip opening (difference between the vertical coordinates of the upper and lower middle lip markers) and lip spreading (difference between the horizontal coordinates of the two lip corner markers). Jaw vertical movements were analyzed using the chin marker. Upper lip protrusion was computed as well.

Facial movements
Based on other studies of the facial movements accompanying speech and more specifically prosody, we decided to limit our study to the head and eyebrow movements. Cavé et al. (1996) showed that eyebrow movements accompanying prosody were mainly raising movements. Therefore we decided to study the raising of both the left and the right eyebrows (middle eyebrow markers). As for head movements, the three rotations and translations were available. Since Munhall et al. (2004) and Graf et al. (2002) had found that the main movements related to prosody were nods, we analyzed the rotation of the rigid body around the horizontal axis.

Data shaping
In order to be able to isolate and compare supra-segmental variations for different segmental constituents, we used a normalization technique. After normalization, a value of 1 corresponds to no variation of the considered parameter compared to the neutral version, a value above 1 corresponds to an increase and a value below 1 to a decrease. For both articulatory and facial movement parameters, we analyzed the inter- and intra-utterance contrasts related to focus (inter: comparison of a phrase in its focused and neutral versions; intra: comparison of a focused phrase with the other phrases of the same utterance).

A.3. METHODS FOR THE FMRI PRODUCTION STUDY

Stimuli
Each sentence was presented for 3 seconds at the beginning of the corresponding condition. Then a fixation mark, alternating every 3 seconds between a '+' and a 'x' sign, appeared in the middle of the screen. This alternation aimed at triggering the silent (covert speech) repetition (14 times per condition) of the sentence presented.

Subjects' performance
Before entering into the magnet, the subjects were trained to execute the tasks, first in overt mode, then in covert mode. In addition, pre- and post-scan audio recordings were carried out to estimate the subjects' task performance during the fMRI scan. Subjects were prompted by exactly the same script as during the scans. They produced each of the sentences 4 times. For the post-scan recording, the instruction was to reproduce in overt speech the intonation patterns mentally produced during the scans. Overall, the subjects' performance, as measured by the audio recordings before and after the scans, indicated that their production varied neither in rhythm nor in intonation between recordings.

fMRI paradigm
Three functional scans were performed. A block paradigm was used. A scan comprised 8 epochs (two repetitions of each condition) of 42 seconds each. The order of presentation of the four conditions was varied across scans and across subjects.

Functional MR imaging was performed on a 1.5T imager (Philips NT) with echo-planar (EPI) acquisition. Twenty-five adjacent, axial, slices (5mm thickness each) were imaged 10 times during each epoch. The imaging volume was oriented parallel to the bi-commissural plane. An EPI MR pulse sequence was used. The MR parameters were: TR = 3700ms, TE = 45ms, pulse angle = 90°, acquisition matrix = 64x64, reconstruction matrix = 128x128, field-of-view = 256x256mm^2. Between the first and second functional scans, a high-resolution 3D anatomical scan was acquired.

fMRI data processing

Data analysis was performed using the SPM-99 software (Wellcome Department of Cognitive Neurology, London). First, motion correction was applied. All images within a functional scan were realigned by means of a rigid body transformation. Then, the anatomical volume was spatially normalized into a reference space using the Montreal Neurological Institute template. The normalization parameters were subsequently applied to the set of functional images. Finally, to conform to the assumption in SPM that the data are normally distributed, the functional images were spatially smoothed.

Contrasts between conditions were determined voxelwise using the General Linear Model. Statistical significance threshold for individual voxels was established at $p = 0.001$. Clusters of activated voxels were then identified, based on the intensity of the individual responses and the spatial extent of the clusters. Finally, a significance threshold of $p = 0.05$ (corrected for multiple comparisons) was applied for identification of the activated clusters.

A.4. METHODS FOR THE FMRI PERCEPTION STUDY

Subjects

Twelve healthy right handed volunteers (Edinburgh Inventory; Oldfield 1971), native speakers of French (4 women and 8 men), with normal or corrected-to-normal vision and normal neurological history were examined. The study was approved by the ATR Human Subject Review and subjects gave informed written consent for experimental procedures.

Stimuli

The stimuli consisted of sentences spoken by a native French speaker (women). 24 were pronounced in normal speech and 6 in whispered speech. The structure of the sentences was: subject (two-syllable first name) - two-syllable action verb - object (two-syllable noun preceded by a one-syllable singular determiner). All syllables were CV syllables (sonorant consonants). An example of one of the sentences used is given below:

Nico mangea le bonbon. *Nico ate the candy.*

The audiovisual stimuli were recorded at ex-ICP (now Speech & Cognition Department – GIPSA-lab) in a sound attenuated room. Two focus conditions were

recorded for each sentence: no-focus (neutral) and focus on the subject. A question/answer procedure was used to elicit prosodic focus as naturally as possible.
After the recordings, audio data was normalized regarding intensity. Video clips were re-centred for the head of the speaker to be in the middle of the image. Three types of stimuli were designed: audio only (A), visual only (V) and audiovisual (AV). The AV stimuli (avi) were directly extracted from the recordings. Both audio and video streams were then isolated to design the A (wav) and V (avi) stimuli. The duration of all the stimuli was 3s.
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-centred black cross that they were asked to fixate (fixation cross).

Procedures
The subjects were told that they would see and/or hear a sentence extracted from a dialogue of the following type:
A: Nico mangea le bonbon. *Nico ate the candy.*
B: Sarah mangea le bonbon ? *Sarah ate the candy?*
A: NICO mangea le bonbon. *NICO ate the candy.*
Two situations were possible: 1. B did not understand who performed the action (as in the example above) and A corrected him (focus case); 2. B understood well but was unsure and repeated the sentence in a question mode and A simply repeated the correct sentence in a neutral mode (no focus case). The subjects were asked to identify the situation: A corrected B or not. They were therefore indirectly asked to identify focus cases.
The stimuli (video and/or audio) were presented to the subjects inside the MR scanner using Neurobehavioral Systems' Presentation software. Audio was presented via MR-compatible headphones (Hitachi Advanced Systems' ceramic transducer headphones). Video was presented through a projector located outside the MR room and a mirror positioned inside the head coil just above the subject's eyes. The subjects responded via a MR-compatible two-button response box placed in their right hand. They were told beforehand that they had to provide an answer even if they were not sure and that they should respond only once the stimulus was finished (fixation cross back).
Prior to the experiment, subjects were trained with the experimental task. Once inside the MR-scanner, the subjects could train using the response box and audio volume was adjusted.

fMRI paradigm
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). For normal speech, one functional scan used the first 12 sentences available in all conditions and

the second the last 12 sentences. For whispered speech, the two sessions both consisted of the same 6 sentences replicated twice for all conditions. A total of 14 null events (NE, fixation cross) were added to the 72 stimuli to vary intertrial interval times. Trials were presented as events lasting 5.1s: stimulus (3s) + response delay (2.1s). NE also lasted 5.1s. The trial sequences were presented following a pseudo-random order. Total duration of a scan was approximately 7 min (5.1s × 86 trials).

fMRI data collection and processing

Blood Oxygenation Level-Dependent (BOLD) contrast was measured over the whole brain using a Shimadzu-Marconi Magnex Eclipse 1.5 Tesla Power Drive 250 imager at the ATR Brain Activity Imaging Center. T2*-weighted images were acquired using a gradient echo-planar sequence (repetition time: 2000ms; echo time: 48ms; flip angle: 80°; 5 initial dummy scans). A total of 20 axial sequential slices with a 3×3×6 mm voxel resolution were acquired. They covered the cortex and cerebellum. A high-resolution 3D anatomical MR scan of the volume examined functionally was acquired after the four functional MR scans (191 slice RF-FAST sequence: repetition time: 12ms; echo time: 4.5ms; flip angle: 20°; voxel size: 1x1x1 mm).

Functional data 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). All the functional images were pre-processed as follows. Movement artifact was corrected for by realigning the images. Differences in acquisition time between slices were accounted for and images were spatially normalized to a standard space. Finally, the images were smoothed using a 6×6 ×14 Gaussian kernel.

The haemodynamic response to the onset of each event was modeled using a delayed haemodynamic response function (HRF). Contrasts were determined voxelwise using a General Linear Model. Statistical significance threshold for individual voxels was established at $p<.01$ uncorrected. Clusters of activated voxels were then identified, based on the intensity of the individual responses and the spatial extent of the clusters (10 voxels). A random effect analysis was then conducted for the group of 12 subjects.

A.5. METHODS FOR THE MULTIMODAL FMRI STUDY

fMRI paradigm

A pseudorandom, event-related fMRI paradigm was employed. Four functional scans were acquired, one for each type of pointing. Each functional scan included a sequence of the following five 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) and a null event (NE, black fixation cross). 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. 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. NE condition also lasted 4.5 s. The trial sequences were presented in a pseudo-random optimized order. Total duration of a scan was approximately 9 min (4.5 s × 5 conditions × 24 trials).

Apparatus

Three types of behavioural responses were recorded: vocal production, eye movement and index finger movement. 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 tube was connected with a mask placed over the subjects' mouth. This apparatus reinforced the signal-to-noise ratio. Eye position was monitored using an ASL 504 eye-tracker (Applied Science Laboratories, Bedford, MA) coupled with the scanner. Right index finger responses were recorded using a digital camera placed out of the fMRI room, behind the window.

fMRI data acquisition and processing

A whole-body 3 Tesla MRI imager (Bruker) with gradient echo acquisition was used to measure blood oxygenation level-dependent 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 parallel to the bi-commissural plane, encompassing the whole brain and the cerebellum. A high-resolution 3D anatomical scan was obtained. Anatomical images were acquired using a sagittal MPRAGE 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 twice. Functional data analysis was performed using SPM2 software (Wellcome Department of Cognitive Neurology, London). Functional data were realigned to correct for head motion using a rigid body transformation. A spatial normalisation was applied. The functional images were spatially smoothed.

To examine cerebral activation during the crucial part of the trials (after the preparation phase), the haemodynamic response to the onset of each event was modelled with a delayed haemodynamic response function, empirically shifted to onset 3.5 s later. Contrasts between conditions were determined voxelwise using the General Linear Model. To perform a random effect analysis, the contrast images (pointing vs control) were calculated for each subject individually and were then entered into a one-sample t test with a significance threshold of $p<.005$ uncorrected.

Behavioural results

Monitoring the subjects' utterances included word accuracy and correct prosodic contour productions. Word recognition was possible in the audio signals from all

subjects but 2. F0 examination was possible on 9 out of 16 subjects. In the other cases, technical problems with the microphone or the mask drastically degraded the signal. Word accuracy and F0 patterns were correct in all available prosodic and syntactic conditions (and their controls). Eye-tracking data were analyzed with the eye-tracker software. Horizontal positions of the eyes were checked using a Matlab script. Subjects behaved according to the instructions. An overview of the video data on finger movements shows that all the subjects performed adequately.