Research Report

Lateralization of tonal and intonational pitch processing: An MEG study

R. Fournier[a],*, C. Gussenhoven[a], O. Jensen[b], P. Hagoort[b,c]

[a]*Department of Linguistics, Radboud University Nijmegen, The Netherlands*
[b]*Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands*
[c]*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

ARTICLE INFO

ABSTRACT

An MEG experiment was carried out in order to compare the processing of lexical-tonal and intonational contrasts, based on the tonal dialect of Roermond (the Netherlands). A set of words with identical phoneme sequences but distinct pitch contours, which represented different lexical meanings or discourse meanings (statement vs. question), were presented to native speakers as well as to a control group of speakers of Standard Dutch, a non-tone language. The stimuli were arranged in a mismatch paradigm, under three experimental conditions: in the first condition (lexical), the pitch contour differences between standard and deviant stimuli reflected differences between lexical meanings; in the second condition (intonational), the stimuli differed in their discourse meaning; in the third condition (combined), they differed both in their lexical and discourse meaning. In all three conditions, native as well as non-native responses showed a clear MMNm (magnetic mismatch negativity) in a time window from 150 to 250 ms after the divergence point of standard and deviant pitch contours. In the lexical condition, a stronger response was found over the left temporal cortex of native as well as non-native speakers. In the intonational condition, the same activation pattern was observed in the control group, but not in the group of native speakers, who showed a right-hemisphere dominance instead. Finally, in the combined (lexical and intonational) condition, brain reactions appeared to represent the summation of the patterns found in the other two conditions. In sum, the lateralization of pitch processing is condition-dependent in the native group only, which suggests that language experience determines how processes should be distributed over both temporal cortices, according to the functions available in the grammar.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Although the number of studies on speech prosody has grown considerably in the last decades, there is no consensus on how its different components are apprehended by the human brain (see Baum and Pell (1999) for a review). First, it has not so far been established with certainty whether prosody should be considered a mosaic of parameters (f_0, duration and intensity) which may be processed in different areas of the brain (Van Lancker and Sidtis, 1992; Zatorre et al., 1992), or whether the parameters form an entity that is treated differently depending on its cognitive function (Van Lancker, 1980). This so-

---

* *Corresponding author*. Department of Linguistics, Radboud University Nijmegen, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands.
  E-mail address: r.fournier@let.ru.nl (R. Fournier).

called *functional lateralization hypothesis* holds that the processing of linguistically-relevant prosodic features is lateralized to the left hemisphere (LH), while emotional aspects of prosody are processed in the right hemisphere (RH), or at least, not predominantly in the LH (Starkstein et al., 1994).

The LH specialization for the linguistic use of prosodic cues has been identified, for instance, in a PET (positron emission tomography) study by Gandour et al. (2000) based on Thai. As a tone language, Thai may distinguish lexical items only by means of a difference in pitch contour. The authors were able to show that native speakers used their left frontal operculum (in the vicinity of Broca's area, well-known for its role in language processing) when discriminating pitch patterns in Thai words. By contrast, native speakers of English (a non-tone language), but also of Chinese (a tone language), who were not familiar with Thai, did not activate this area when exposed to the same stimuli and rather exhibited a RH dominance during the pitch discrimination task. These results suggest that the LH is activated only if the input can be interpreted in a linguistic way, within a known system of tonal contrasts that convey *lexical meaning*.

The general picture whereby lexical-tonal contrasts are more LH-lateralized than non-linguistic pitch processing, fails to address a number of issues. First, lexical meaning is not the only linguistic component in which pitch is involved. Using the same phonetic parameter on the intonational level, speakers are also able to express *discoursal meanings*, such as the distinction between questions and statements. Recent studies on German (Friederici and Alter, 2004; Meyer et al., 2004) as well as on Mandarin Chinese (Gandour et al., 2003a,b) point to an increased RH activity during the processing of intonation. However, the stimuli used in these studies were not of the same kind as those used in most experiments on lexical tones. Instead of words, intonation studies usually have recourse to phrases or sentences, which are typical domains for intonational contrasts. As Baum and Pell (1999) point out, there might be a difference in lateralization depending on the domain size of stimuli, such that larger domains tend to be assigned to the RH. A recent cross-language (Chinese and English) fMRI experiment by Gandour et al. (2004) shows that when using equally long stimuli in tone and intonation discrimination tasks, the general RH dominance is in fact less obvious. Regardless of the task, both language groups do exhibit an RH asymmetry, but this asymmetry is confined to some regions, while other regions show an LH dominance in the Chinese speakers only. The interpretation proposed is that the RH mediates the prosodic analysis of complex sounds, while the LH is responsible for language-related processing.

A second issue which is not captured by the lateralization view identified above concerns the timing of pitch processing. Purely topographic hypotheses and research findings significantly underspecify the brain activation patterns that can be assumed to exist. Once a fuller conception is formulated about the tasks that subjects are faced with in pitch-related experiments, more detailed hypotheses readily suggest themselves. For instance, regardless of the status of a pitch contrast (lexical, structural intonational, or emotional), subjects will need to process the acoustic information in the stimuli before their linguistic or emotional status can be ascertained. Given the brain imaging results, this might imply that all stimuli will need to be processed in the RH first, and that subsequently only those that embody linguistic contrasts are further dealt with in the LH. Hypotheses that take a processing perspective as their point of departure cannot of course be tested without taking temporal information into account.

In sum, we believe there to be a need for information about brain activation patterns that are temporally fine-grained enough to be able to trace differently timed activations for similar stimuli. In addition, we need to be assured that we present stimuli that unambiguously represent both lexical and intonational contrasts in otherwise equivalent conditions. This latter requirement implies that we need to adopt a stimulus set comparable to that used in Gandour et al. (2004), in which the stimuli that represented lexical contrasts were equally long as and segmentally comparable to those representing intonational contrasts. Perhaps more so than was the case in Gandour et al. (2004), we need to ensure that the language from which we take our stimuli encodes intonational differences phonologically, i.e. by means of different tonal representations, rather than paralinguistically, i.e. by means of pitch range variation. On the basis of the literature (Shen, 1990; Wu, 2000), it could be argued that the difference between questions and statements in Mandarin Chinese may not be phonological in the way that lexical differences are. For instance, the syllable *yu* pronounced with a falling tone (Tone 4) represents the word for 'pen', whereas the same syllable pronounced with a rising tone (Tone 2) represents the word for 'fish'. By contrast, regardless of whether it is said as a statement or as a question, Tone 2 is always rising. What distinguishes both modes is the higher pitch register used in questions.

The present study aims at providing an account on the processing of tonal and intonational contrasts in Roermond Dutch, a dialect that unambiguously encodes both contrast types phonologically. Like the related tonal dialects spoken in the Netherlands, Belgium (together also referred to as Limburgian) and Germany, the dialect has two lexical tones, called Accent 1 and Accent 2. It is not difficult to find minimal pairs, like *vaer*[1] 'feather' and *vaer*[2] 'ferry', or *knien*[1] 'rabbits' and *knien*[2] 'rabbit'. Since native listeners perceive a lexical and discoursal meaning in any (grammatically correct) pitch contour of a word spoken in isolation, the Roermond dialect allows us to use units of the same length to compare tonal and intonational processing. Unlike Standard Chinese, the Roermond lexical tones can have drastically different pitch contours depending on whether they are said with statement or with interrogative intonation. For instance, a syllable like [kni:n] said with a pitch fall signifies the plural form for 'rabbits' spoken as a statement. When the same syllable is spoken with a rise followed by a fall, it signifies the same plural form 'rabbits' said with question intonation, and when it is said with a fall followed by a rise, it represents the singular form for 'rabbit' as spoken with a statement intonation (Gussenhoven, 2000). In this way, every monosyllabic pitch contour represents a unique combination of discoursal and lexical meanings. That is, neither the lexical tone category (Accent 1 vs. Accent 2) nor the intonational categories (statement vs. question) have invariant pitch contours like 'rise' or 'rise–fall' which can be said to represent any one of those categories; rather, any legitimate pitch contour simultaneously

represents a lexical tone category and an intonational contour. There can be no doubt, therefore, that the intonational differences are phonological in the same way as the lexical differences are.

In our experiment, Roermond Dutch word stimuli were used in order to create lexical contrasts (e.g. $vaer^1$ 'feather' vs. $vaer^2$ 'ferry') as well as intonational contrasts (e.g. $vaer^1$ with statement intonation vs. $vaer^1$ with question intonation), or a combination of these two (e.g. $vaer^1$ 'feather' with statement intonation vs. $vaer^2$ 'ferry' with question intonation). The different experimental conditions were set up according to the auditory mismatch paradigm, in which sequences of frequent, (near-)identical sounds are interrupted by infrequent, deviant sounds. The idea behind this method is that the repeated presentation of a sound creates a pattern in the listener's sensory memory, against which every incoming sound is matched. When a deviant sound is presented, the detection of change is reflected by an increased activity in the listener's auditory cortex in a time window from 150 to 250 ms after stimulus onset. This brain response is referred to as mismatch negativity (MMN), an ERP component that has been used extensively in speech perception studies because, interestingly, its amplitude directly depends on the subject's language experience. When a deviant phoneme or word pertains to the subject's mother tongue, the MMN will be stronger than in the case of unfamiliar stimuli (Dehaene-Lambertz, 1997; Näätänen, 2001; Endrass et al., 2004). It was also shown (Phillips et al., 2000) that acoustic variation within a phoneme category triggers weaker and later MMN than the difference between phonemes.

A closer look at the spatial distribution of MMN revealed that the increase in the MMN amplitude for native (as compared to non-native) differences and for phonological (as compared to acoustic, i.e. subphonemic) differences is left-lateralized (Näätänen et al., 1997), reflecting the LH specialization for phonological processing at the segmental level. However, when elicited in other experimental conditions, MMN can also have similar or even higher amplitudes in the RH. For instance, Shtyrov et al. (1998) showed that in the presence of environmental noise, the MMN amplitude in reaction to phoneme contrasts decreases in the LH while it increases in the RH. In the musical domain, it was also shown (Fujioka et al., 2004) that LH is not the only host of MMN: no laterality effect was found in subjects who were presented with five-note piano melodies differing either in contour (rising vs. falling ending) or in interval (last note raised by one tone without changing the pitch contour). This bilateral reaction to pitch contour differences was verified for linguistic pitch as well, in an experiment which compared intonational (statement vs. question) and segmental contrasts (/a/ vs. /e/) in Japanese (Imaizumi et al., 1998). While both contrasts triggered LH activity, only the intonational distinction between falling and rising contours required an additional RH contribution.

In the light of these results, mismatch responses seem to provide a very convenient handle for a comparative study of the processing of tone and intonation. On the one hand, it can be used for highlighting possible task-related laterality effects, and on the other hand, its sensitivity to language background may help to discriminate between phonetic and phonological

processing. More specifically, with respect to our experimental goals, we intended to use MMN to verify a number of hypotheses. First, the functional difference between tonal and intonational contrasts should be reflected by the lateralization of MMN. We expected tonal processing to be more left-lateralized than intonational processing, since it is more directly concerned with lexical characteristics, which have been shown to be left-lateralized (Petersen et al., 1988; Démonet et al., 1992; Indefrey and Cutler, 2005). Second, we assumed that the brain reactions to the stimuli would depend on the subjects' language experience. In order to test this hypothesis, we presented the same stimuli to speakers of Standard Dutch, a non-tone language. Since these subjects had no knowledge of Roermond Dutch or any other tone language, any perceived differences in pitch contours could only be processed within an intonational or emotional framework, thus shifting the MMN to the RH more than they would have during a lexical task.

Since the comparison between contrast types as well as between language groups intended to cover both temporal and spatial aspects of the data, it was decided to record brain activity with the help of a whole-head MEG system. Magnetoencephalography (MEG), like EEG (electroencephalogram), is a non-invasive technique that can measure neuronal brain activity on a millisecond time scale and hence makes it possible to detect the expected mismatch negativity, or rather magnetic mismatch fields, the magnetic equivalent to MMN. In addition to this outstanding temporal resolution, MEG provides more fine-grained spatial information than EEG. With this technique, we intended to collect new data on tonal and intonational processing which could be analyzed against the background of earlier PET/fMRI-based as well as MMN/MMNm-based results.

2. Results

In order to characterize the brain reactions to different pitch-based linguistic contrasts, we analyzed ERFs triggered by standard and deviant (henceforth called *target*) word stimuli which differed from each other in terms of tone, intonation, or both (as explained in Section 4.2). Fig. 1a shows the ERFs for the axial gradiometers for the standard and target stimuli averaged over all conditions (lexical-tonal, intonational and combined) and subjects (speakers of the Roermond dialect and a control group of Dutch subjects without knowledge of the dialect) and aligned with respect to the time point at which standard and target pitch contours started to diverge (see Table 1). The baseline was calculated with respect to the data in a 200 ms time window prior to the stimulus (not prior the divergence point). Systematic deflections in the ERFs are observed over the left and right hemispheres. Note that the difference between standards and targets form two dipolar patterns with polarities consistent with the negative and positive deflections in the fields. The dipolar distributions of the fields (indicated by the circles) suggest a source in the left and the right hemispheres each situated between the negative and positive fields. The strongest deflections are also associated with the strongest differences between the standards and
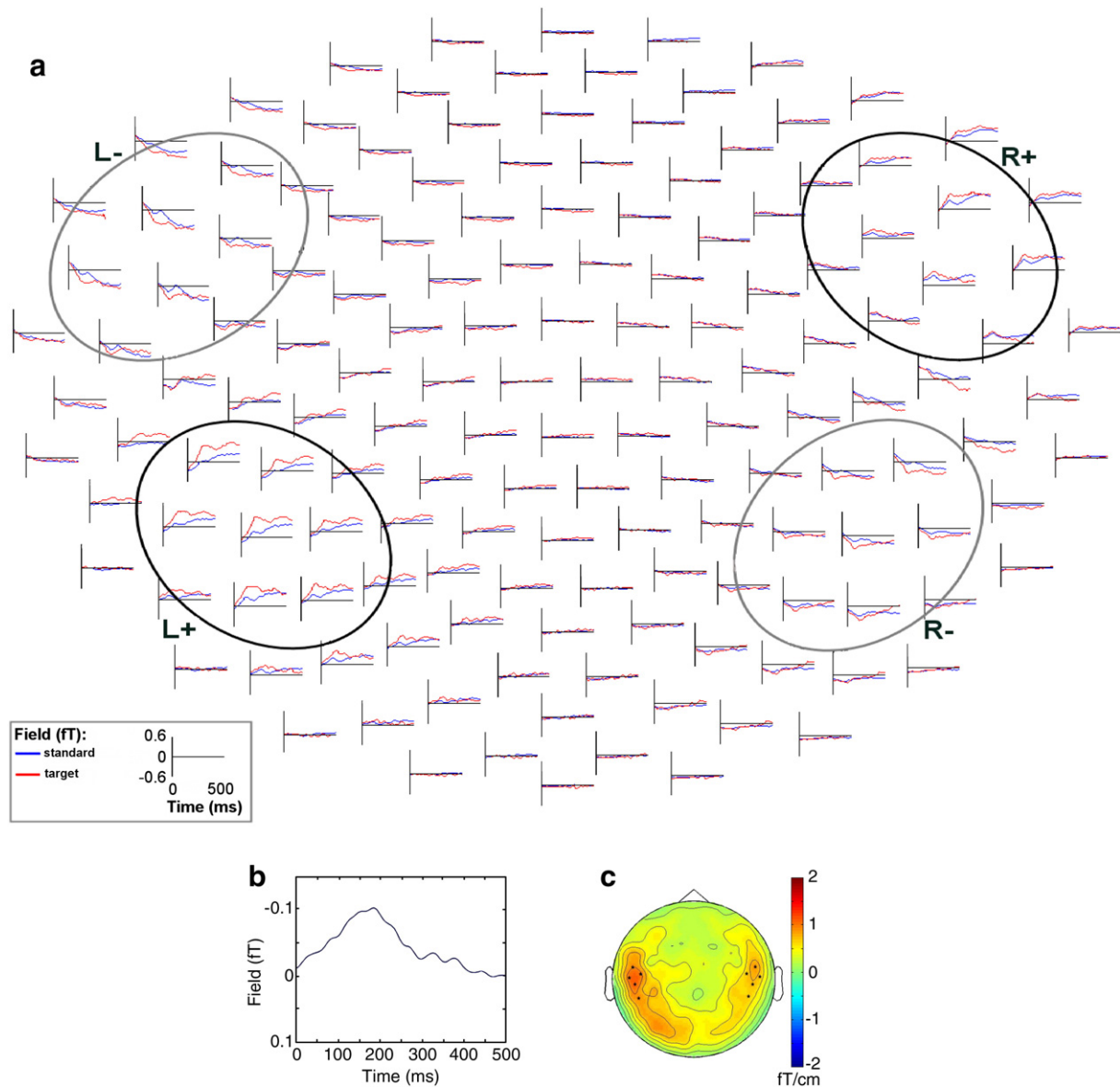
Fig. 1 – The event related fields (ERFs) from the axial gradient when comparing standard to targets. (a) The grand average of the ERFs for the standard (blue) and target (red) stimuli aligned to the divergence point. Each plot represents the field of a sensor arranged topographically according to the sensor position in the helmet. The two dipolar field patterns over the left and right hemispheres are indicated by the circles. Note the stronger deflection for targets compared to standards. (b) The grand average of the ERFs for the difference between standards and targets based on the sensors circled in (a), computed as the sum of the activity registered in the circles L– and R– minus the sum of the activity in the circles L+ and R+. A large difference is observed around 200 ms. (c) A topographic plot of the combined planar gradient of the ERFs for the difference between target and standard stimuli in the interval 0–500 ms after the divergence point. The sensors selected for the subsequent statistical analysis (MLT/MRT 13, 23, 24, 25, and 33) are marked with *.

targets. These distributions point to a left and right hemisphere source producing the ERFs and show that the magnitude of these sources is higher for targets than standards. Fig. 1b displays the difference between standards and targets, computed by combining equal-polarity fields of the 24 sensors circled in Fig. 1a. A large difference in the grand average of the waveforms in these 24 sensors was observed

around 200 ms. This latency, as well as the region over which the difference in activity occurs, corresponds to the magnetic equivalent of the mismatch negativity response (MMN), as reported for instance in Phillips (2001) and Näätänen (2001). Thus, in the rest of the study we used the interval 150–250 ms after the divergence point as the time interval of interest when comparing the different conditions.

Table 1 – Point at which the pitch contour of the standard stimuli and the target stimuli started to differ from each other, estimated for each experimental condition after word onset. A1/A2 refers to the two lexical tones (Accent 1 and Accent 2) and s/q to the two discourse meanings (s: statement; q: question).

Contrast	Standard ↔ target		Divergence point
Lexical	A1, s	A2, s	280 ms
	A1, q	A2, q	115 ms
Intonational	A1, s	A1, q	110 ms
	A2, s	A2, q	215 ms
Combined	A1, s	A2, q	150 ms
(lex. and int.)	A2, s	A1, q	130 ms

To further characterize the MMNm, we converted the ERFs to the combined planar gradient. The main advantage of this approach is that the largest planar gradient is observed directly above the source (Hämäläinen et al., 1993; Bastiaansen and Knösche, 2000). Fig. 1c shows the difference in the planar gradient for ERFs when subtracting standards from targets, also aligned in time to the divergence point. The sensors with the largest differences clustered over LH and RH temporal areas, again pointing to a source in each hemisphere. The left source was stronger than the right source. We selected five channels from the region of the largest MMNm, taking the intersection of the strongest signals in the three conditions and the two language groups. These channels, together with the five corresponding channels over the RH (marked in Fig. 1c), were used for all subsequent analyses, irrespective of subject and condition.

Fig. 2 shows the topographic plots of the difference between targets and standards in the three conditions (lexical, intonational and combined contrast) and two groups (subjects with or without knowledge of the Roermond dialect) in the 150–250 ms window for the combined planar gradient. It appears that the left hemisphere is dominant in all conditions and in both groups, except for the intonational contrasts in the

Roermond group. The difference between targets and standards is clearly larger in intonational and combined contrasts than in lexical contrasts. We also examined brain reactions in other time windows (not only 50–150 and 250–350 ms, but also 0–100, 100–200, 200–300 and 300–400 ms). However, in none of these additional time windows were significant effects observed. For this reason, we will only report on the differences observed in the 150–250 ms time window.

In order to put these observations to the test, we averaged the magnitudes of the combined planar gradients from the selected sensors (one average per hemisphere and per stimulus type: standard or target) and time window (150–250 ms). First of all, we checked whether the difference between targets and standards was significant in all conditions, including the lexical contrasts, by performing an ANOVA with the dependent variable "planar_gradient_magnitude" (as measured for standards as well as for targets) and the independent variables STIMULUS_TYPE (2 levels: standard and target), HEMISPHERE (2 levels) and LANGUAGE_GROUP (2 levels). The p-values found for the independent variable STIMULUS_TYPE (p_lexical = 0.001, p_intonational < 0.001, p_combined = 0.003) indicated that targets and standards were indeed significantly different. We then submitted the differences between standards and targets to a repeated-measure ANOVA using the within-subject factors CONTRAST (3 levels: lexical, intonational and combined) and HEMISPHERE (2 levels: left and right), and the between-subjects factor LANGUAGE_GROUP (2 levels: Roermond and control). Results were adjusted using the Greenhouse–Geisser correction. This omnibus ANOVA showed a significant main effect of CONTRAST ($F(1.5,20.9) = 5.17$, $p = 0.022$), as well as a significant interaction between LANGUAGE_GROUP, CONTRAST and HEMISPHERE ($F(1.7,23.8) = 4.67$, $p = 0.024$, for 3 contrasts). Further analyses per contrast revealed no significant effect or interaction of effects in the combined condition. We therefore excluded this condition from the analysis, focussing on the difference between lexical and intonational contrasts. A new ANOVA was performed, with the same factors as above but only two levels for CONTRAST, which again highlighted the effect of CONTRAST ($F(1,14) = 12.42$,
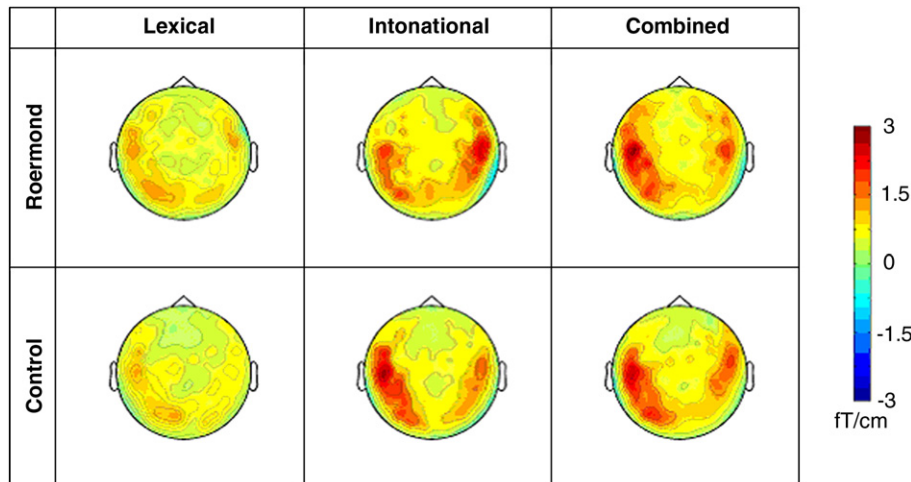

Fig. 2 – Topographic plots of the difference in magnetic fields between target and standard stimuli in the interval 150–250 ms after point of divergence, in the 3 conditions and 2 language groups. The maps represent the combined planar gradient.

p=0.003) and the LANGUAGE_GROUP*CONTRAST*HEMISPHERE interaction (*F*(1,14)=10.93, *p*=0.005).

A possible interpretation of this interaction, which would match the observations made on the topographic plots, is that language experience determines whether the hemispheric specialization of prosodic processing depends on the type of contrast. More precisely, the interaction CONTRAST*HEMISPHERE should be significant for the Roermond group but not for the control group, reflecting the particular Roermond lateralization pattern for intonational contrasts. This was tested with the help of separate ANOVAs per language group (with the factors CONTRAST and HEMISPHERE), and within each language group, separate ANOVAs per contrast (with the factor HEMISPHERE). In the Roermond group, the ANOVA using CONTRAST (2 levels) and HEMISPHERE highlighted the expected effect of the contrast on lateralization, with a significant CONTRAST*HEMISPHERE interaction (*F*(1,7)=9.93, *p*=0.016). In the control group, there was only a significant main effect for CONTRAST (*F*(1,7)=11.98, *p*=0.011), reflecting an overall stronger MMNm in intonational as compared to lexical contrasts (1.99 fT/m against 1.36 fT/m). The analyses per language group and per hemisphere revealed that this contrast effect was only significant in the LH for the control group (*F*(1,7)=12.64, *p*=0.009), and in the RH for the Roermond group (*F*(1,7)=12.45, *p*=0.010). Finally, the analyses per language group and per contrast, which allowed us to isolate the HEMISPHERE factor, showed lateralization effects in the Roermond group only (*F*(1,7)=6.32, *p*=0.040 in lexical contrasts; *F*(1,7)=6.26, *p*=0.041 in intonational contrasts).

In sum, the topographic plots as well as the statistical analysis of our data reveal clearer and more differentiated activation patterns in the Roermond group than in the control group. In particular, unlike the non-native speakers, Roermond subjects process intonational patterns predominantly in the right hemisphere. Fig. 3 gives an overview of the amplitudes found in the different conditions, per hemisphere and language group.

3.   Discussion

In our MEG study of lexical-tonal and intonational processing we found a clear MMNm in a time window from 150 to 250 ms
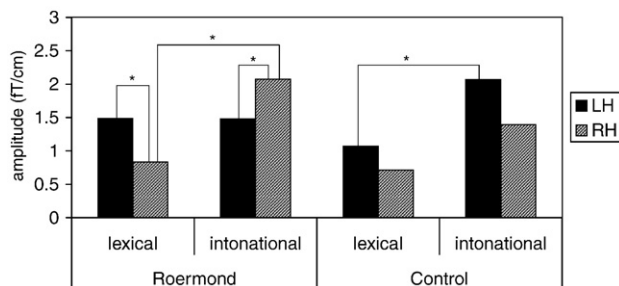

Fig. 3 – Differences between ERFs (150–250 ms) of targets and standards in lexical and intonational contrasts, with respect to language group and hemisphere. The values were obtained from the planar gradient of the sensors marked in Fig. 2c. Significant differences between values (*p*<0.05) are marked with *.

after the divergence point of standard and deviant pitch contours. The most remarkable finding was a clear difference in lateralization between the native Roermond speakers and a non-native control group. While non-native speakers showed a stronger MMNm over the left temporal cortex in all conditions, the native Roermond group showed a stronger response over the left temporal cortex for lexical contrasts, but a predominantly RH response for the intonational contrasts. This finding is suggestive of the possibility that the native Roermond speakers processed lexical contrasts predominantly in the LH, but intonational contrasts predominatly in the RH.

The differential distribution of brain activiation in the Roermond group compared to a more uniform pattern in the control group can be seen as an example of the functional plasticity of the brain (cf Hagoort et al., 2003). Temporal cortices are recruitable bilaterally for pitch processing, and lateralization patterns emerge during language acquisition according to the options available in the grammar. Our results cannot be explained on the basis of the phonetic features in which standards and deviants differed, since the same features (namely pitch, enhanced with duration) were used for encoding lexical and intonational contrasts. Moreover, if activation patterns solely depended on acoustic cues, they would be identical in native and non-native subjects, which was shown not to be the case. Our results must therefore be due to the linguistic functions that the phonetic features encoded.

The finding that Roermond subjects process lexical contrasts predominantly in the LH temporal cortex is consistent with earlier results on phonological and lexical processing. The shift towards the RH in reaction to intonational contrasts is similar to that reported by Imaizumi et al. (1998), who found that Japanese lexical segmental contrasts triggered LH activity without any significant RH activity, while bilateral activity was found for intonational contrasts. The Roermond RH dominance during intonational processing also agrees with earlier work (Gandour et al., 2003a,b; Friederici and Alter, 2004; Meyer et al., 2004). However, these earlier findings do not seem to be compatible with the general LH dominance found in the control group. Since tonal contrasts are absent from Standard Dutch (and therefore from the speech of the non-native participants), pitch contour differences are assumed to be interpreted as intonational or affective differences, which are both traditionally assumed to be right-lateralized (Pell, 2002; Buchanan et al., 2000). We assume that the greater complexity of the linguistic system that was acquired by the Roermond speakers, who learned a lexical tone distinction that combines in intricate ways with an intonational distinction, led to a different topography for the processing of pitch contrasts from that developed by the Standard Dutch subjects, who acquired a system with intonation contrasts only. Our findings support the view that prosodic processing may recruit a number of different regions in the brain, both within the same hemisphere and across the two hemispheres, and that there is no universal brain region for pitch processing (Baum and Pell, 1999). We must therefore not expect the processing of a linguistic function like intonation to be allocated to the same part of the brain by speakers of different languages. The group-specific right-hemisphere activity for the intonational contrasts may reflect the greater functional complexity of the

stimuli for the Roermond listeners, who needed to make decisions on the discourse type as well as on the lexical identity on each trial. If processing words involves a prior recognition of the intonation category rather than a prior recognition of the lexical category, any change in intonation will trigger a double processing path. It is these trials that are responsible for the between-group difference. One third of the trials involved a lexical shift within the same intonation contour. These trials are comparable to all the trials in the control group, who were only ever faced with a functionally one-dimensional decision, the intonational contrast. This assumption is thus entirely in line with Hickock and Poeppel's (2004) conclusion that phonetic aspects of speech prosody tend to show a right-hemisphere bias, while higher-level functions are processed in the left hemisphere. Also, even though we cannot make any direct comparisons, Tong et al. (2005) found that both English and Chinese listeners displayed widely dispersed regions of activity in fMRI measurements during discrimination tasks involving stress location and discourse category (question or statement) in Mandarin stimuli. They conclude that their data are compatible with Hickock and Poepel's view.

Our results confirm the conclusion reached by Gandour (2007) that studies of brain activation involving tone and intonation show 'a mosaic of multiple regional asymmetries that allows for different regions being differentially weighted depending on language experience, stimulus properties, and cognitive processes evoked by task demands'. It is to be expected that combining spatially as well as temporally sensitive registrations of stimuli in subjects with a greater variety of language backgrounds will not just confirm the functional plasticity of the brain, but reveal in more detail in what ways different regions in the brain collaborate in the processing of prosodic contrasts.

The sensitivity of MMNm to language background had led us to formulate the hypothesis that our stimuli would trigger a stronger MMNm in the native than in the non-native group. Such a difference could not be established in the present data, presumably because the stimuli were fully compatible with Standard Dutch phonological forms. Haas [ha:s] is the Standard Dutch word for 'hare', while vaer is well-formed phonologically, even if not representing a word. More importantly, the four pitch patterns are readily interpretable as Dutch intonation contours. The contour for the Accent 1 statement represents a neutral, declarative intonation, that for the Accent 1 interrogative represents an emphatic declarative. The falling–rising contour for the Accent 2 statement is interpretable either as a non-emphatic question or as a polite suggestion, while that for the Accent 2 question is an interrogative intonation. Because all four contours are phonologically (and semantically) distinct in Standard Dutch, the MMNm amplitudes were as strong as the ones found in the native group. We did find, however, significantly different MMNm amplitudes due to phonetic differences in the stimuli. As can be seen in Fig. 3, the Roermond interrogative contours have a wider pitch span than the declarative contours, making the difference between statements and questions always larger than the one between Accent 1 and Accent 2, where pitch spans are more similar. According to Näätänen (2001), an increase in acoustic differences between simple tones can cause an increase in MMNm

amplitude. It is reasonable to assume that this also holds for speech stimuli.

In conclusion, our MMNm study supports the view of function-driven, language-dependent pitch processing. It also highlights a clear discrepancy between linguistic and cerebral representations: What is considered universal amongst languages, such as the expression and recognition of discourse meanings by means of intonation, is not necessarily realized in an identical way in the human brain.

4. Experimental procedures

4.1. Subjects

Twelve right-handed native speakers of the Roermond dialect, between 21 and 34 years old (average age: 25.9 years), took part in the experiment. All subjects claimed to speak their dialect on a daily basis and obtained satisfactory scores in a short test that assessed their active and passive knowledge of Roermond Dutch. The test did not refer to any of the linguistic material that was to be investigated in the MEG experiment. Due to very high numbers of eye or muscle artifacts during the MEG recordings, four speakers were excluded from the analysis, which left us with 5 male and 3 female subjects. For the control group, 11 right-handed native speakers of Standard Dutch were recruited. They matched the Roermond subjects with respect to gender, age (average 26.2) and education (tertiary level), and they were not familiar with any Limburgian dialect or any other tone language. We selected 5 males and 3 females within the control group with the least MEG artifacts. Prior to the experiment, all subjects of the Roermond and control group gave written consent and filled a hand dominance questionnaire to attest their right-handedness. The subjects reported no history of neurological disorders or hearing impairments.

4.2. Word stimuli

The stimuli consisted of two sets of Roermond Dutch words recorded by a native speaker, alternately with statement and with question intonation. The first set contained three tonal minimal pairs, that is, word pairs with the same phoneme sequence but different tones (Accent 1 or 2), resulting in different lexical meanings. The second set contained seven words with distinct phoneme sequences, so that their meaning could be identified unambiguously at the segmental level. All recorded words were validated by 20 native speakers (who did not participate in the MEG experiment) in a perception test. All recorded words were validated by 20 native speakers (who did not participate in the MEG experiment) in a perception test. In the first part of the test, the speakers listened to the words presented in a random order and selected, on an answer sheet, the lexical meaning of each word. The second part of the test consisted in determining the discoursal meaning of the words ('statement' or 'question'). Since all lexical and discoursal meanings were identified correctly by at least 90% of the speakers, we assumed that the test words were representative of the linguistic forms we intended to include in the experiment.

These test words were: $haas^1$ (i.e. the syllable [ha:s] pronounced with Accent 1, meaning 'hare') and $haas^2$ (i.e. the syllable [ha:s] pronounced with Accent 2, meaning 'glove'), $graaf^1$ ('canal') and $graaf^2$ ('grave'), and finally, $vaer^1$ ('feather') and $vaer^2$ ('ferry'). Since each word was pronounced with two different intonations, it was also a member of an intonational minimal pair. For instance, $haas^1\_s$ (i.e. $haas$ pronounced with Accent 1 and as a statement) formed a tonal minimal pair with $haas^2\_s$, and an intonational minimal pair with $haas^1\_q$ (Accent 1, question). The use of minimal pairs allowed us to isolate lexical or intonational differences from phoneme-level differences, in order to facilitate the interpretation of brain responses. However, in natural speech, truly *minimal* pairs do not exist. There is always some acoustic variation amongst realizations of a word, even in the sections that are phonologically identical. We took this variation into account by using three versions of each expression, i.e. each word as said with either statement or question intonation. In this way, we ensured that any effects of the deviant stimuli would be attributable to the phonological difference with the preceding stimulus, and not to uncontrollable acoustic differences between one pronunciation and the next. This resulted in 36 one-word utterances that were members of tonal minimal pairs (3 versions of 6 words with 2 intonations). For the words which were not members of tonal minimal pairs, namely $diek^2$ ('dike'), $daak^2$ ('roof'), $lien^1$ ('line'), $pien^1$ ('pain'), $bank^2$ ('bank'), $bandj^2$ ('roof') and $huur^1$ ('rent'), only one token was used, one with a question intonation for $diek^2$, $lien^1$, $pien^1$ and $bandj^2$, and one with a statement intonation for $daak^2$, $bank^2$ and $huur^1$. These words were to be used in a detection task unrelated to the experimental question (see Section 4.3) and brain responses to these words were not analyzed. Another set of words for which the brain responses were not analyzed is the minimal pair $graaf^1/graaf^2$. We decided to ignore these results because of an erroneous labelling of one of the sound files containing a version of $graaf^2$. Therefore, the experimental results will be based on the minimal pairs $haas^1/haas^2$ and $vaer^1/vaer^2$.

Fig. 4 shows the f_0 contours of all instances of $haas$ (pronounced [ha:s]) and $vaer$ (pronounced [fe:əχ]) used in the experiment. Although the average word length is 586 ms, the information used for tonal or intonational identification only covers about 300 ms, corresponding to the voiced part of the signal (mostly the vowel [a:] or [e:] plus the transitions between [h] and [a:] and between [e:ə] and [χ]). There is some variation with respect to the temporal alignment of these voiced parts within the words, but they generally start around 100 ms after word onset. More variation can be found in the timing of the prosodic contrasts. For instance, the statement and question contours for Accent 1 begin to diverge at least 100 ms earlier than the statement and question contours for Accent 2. These timing differences were taken into consideration for the time-locked averaging of our MEG data by shifting the ERFs accordingly (see Data analysis). Two additional systematic differences may be observed in Fig. 4. First, the voiced part of the word is significantly longer for Accent 2 compared to Accent 1 (43 ms, $p < 0.001$), and for questions compared to statements (26 ms, $p = 0.014$; univariate ANOVA; factors: ACCENT and INTONATION). Second, as can be seen in Fig. 1, the difference between maximum and
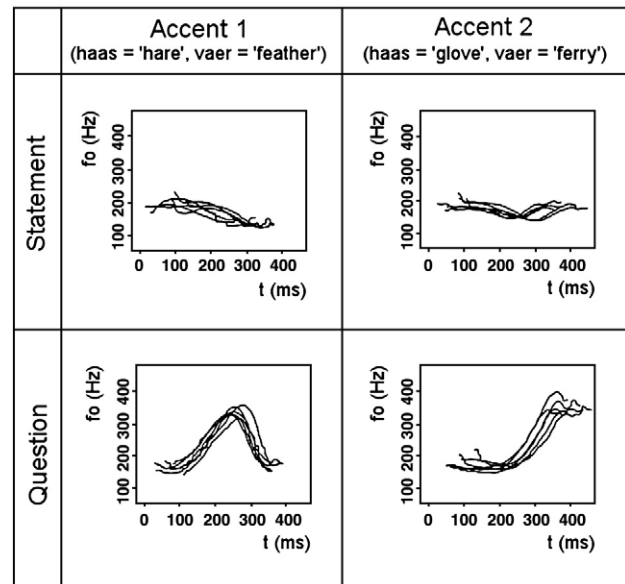

Fig. 4 – The f_0 contours of *haas* and *vaer* with Accents 1 and 2 (representing different lexical meanings) combined with statement or question intonation, as a function of time. Each cell shows the contours of 3 *haas* and 3 *vaer* words.

minimum f_0 within a word is smaller in statements than in questions. The possible effects of these various systematic differences on brain reactions were dealt with in Section 3. Other differences can be observed in Fig. 4, within each cell of the table. These differences reflect the inevitable variation between instances of the same token. In order to assess the size of the acoustic differences between different pronunciations of the same linguistic category, we computed Pearson's correlation coefficients for all possible pairs of f_0 contours within and between categories, based on 100 pitch values per contour, and found that the correlations within categories were always higher than correlations between categories. The highest correlation between categories was found for the lexical contrast (Accent 1 vs. Accent 2) with declarative intonation (average Pearson's coefficient: 0.19, stdev: 0.2), which is still lower than the lowest coefficient found within categories (0.75). This means that in all cases the differences between tokens of the same category are small compared to those among categories.

4.3. Stimulus presentation

The stimuli were presented in 36 blocks of 100 words implementing a mismatch negativity paradigm. Each 100-word block represented a lexical, an intonational or a combination of a lexical and an intonational contrast between *standard* (78%) and *target stimuli* (18%). Both standard and target stimuli belonged to the set of $haas^1/haas^2$ or $vaer^1/vaer^2$ words. A second type of deviant stimuli, the *novel stimuli* (4%), consisted of one-word utterances of one of the seven words which were not members of tonal minimal pairs ($diek^2$, $daak^2$ and the like). The words in each block were pseudo-randomized in such a way that target stimuli always occurred after at least three standard

stimuli, and that these three standard stimuli represented at least two different versions of the same word/intonation combination. We give here as an example the first 12 items in a block representing a lexical contrast, using $haas^1$/$haas^2$ pronounced as statements, with targets in bold print:

$haas^1$_s (v3) - $haas^1$_s (v1) - $haas^1$_s (v2) - $haas^2$_s **(v1)** - $haas^1$_s (v2) - $haas^1$_s (v1) - $lien^1$_q - $haas^1$_s (v3) - $haas^1$_s (v3) - $haas^1$_s (v1) - $haas^2$_s **(v3)** - $lien^1$_q - …

In this sequence, three different versions of $haas^1$_s (v1, v2 and v3) are used as standard stimuli, contrasting with different versions of the target stimuli $haas^2$_s. The word $lien^1$_q is used repeatedly as a novel stimulus. We constructed similar blocks with the words $haas^1$_q (standard), $haas^2$_q (target) and $pien^1$_d (novel), $haas^2$_s (standard), $haas^1$_s (target) and $lien^1$_q (novel), and finally $haas^2$_q (standard), $haas^1$_q (target) and $pien^1$_d (novel). The same was done using the $vaer^1$/$vaer^2$ and $graaf^1$/$graaf^2$ words. This resulted in 12 blocks for lexical contrasts, of which 8 were retained for the analysis (after exclusion of the $graaf$ blocks). Intonational and combined contrasts were modelled in the same way.

All blocks were randomized and presented binaurally to Roermond and control subjects. During presentation, the subjects were requested to focus on a fixation cross and to press a button when they heard a "completely different word", corresponding to the novel stimuli. This condition was not designed to answer our experimental question, but rather to keep the subjects alert. Such use of attended sound contrasts is somewhat unusual in MMN designs but as Phillips et al. (2000) point out, it does not prevent mismatch responses from occurring.

The instructions concerning the task were deliberately kept vague in order to avoid an explicit reference to the prosodic properties of the stimuli, but the correct execution of the task was supervised during a training session consisting of 10 words (30% target and 10% novel). Corrections only had to be made for two subjects of the Roermond group, who at first pressed the button after both target and novel stimuli. During the main session, all subjects showed the same behaviour with respect to the task, pressing the button for novels only, with almost no mistakes. The main session was divided into three groups of 12 blocks. Within each group, which lasted 20 min, blocks were separated by 3-second breaks, during which the fixation cross was replaced by a text announcing the next block. Longer breaks (1 min) were allowed between groups of 12 blocks. Words in the blocks were delivered every second. In total, the training and main session took about 65 min. The software Presentation (version 0.70, www.neuro-bs.com) was used for delivering the stimuli and recording the behavioural responses.

4.4. Data acquisition

Ongoing brain activity was recorded with a whole-head magnetoencephalography system with 151 axial gradiometers (CTF/VSM Systems, Port Coquitlam, British Columbia, Canada). Data were acquired with a 600 Hz sampling frequency after applying a 150 Hz low-pass filter. Head localization was done at the beginning and end of each recording session, using marker coils placed at the cardinal points of the head (nasion, left and right ear canal). The magnetic fields produced by these coils allowed us to measure the position of the subject's head with respect to the MEG sensor array. In addition to the MEG, the electrooculograms were recorded from electrodes placed above and below the left eye (vertical EOG) and at the outer canthus of each eye (horizontal EOG), for later eye artifact removal.

4.5. Data analysis

MEG data were processed using the FieldTrip toolbox developed at the F.C. Donders Centre for Cognitive Neuroimaging (http://www.ru.nl/fcdonders/fieldtrip/). Event-related fields (ERFs) were calculated from the data time-locked to the stimuli (1 s trials; 0.2 s baseline interval). Trials contaminated by eye movement or sensor jump artifacts were eliminated, leaving about 100 target stimulus trials and 340 standard stimulus trials for each subject and condition (lexical, intonational or combined contrast). The novel stimuli trials were not analyzed since they contained a motor response. Likewise, trials immediately following the novel stimulus trials were also excluded from the study.

In order to take into account possible differences in the timing of prosodic contrasts, we estimated, for each pair of contrasting pitch contours, the time point at which these contours started to differ acoustically. This was done by visually inspecting the plots of the relevant contour pairs. For instance, in order to determine the divergence point in a lexical contrast involving $haas^1$_s and $haas^2$_s, we compared all three versions of $haas^1$_s with the three versions of $haas^2$_s. The divergence points estimated in these comparisons (as measured from word onsets) were then averaged, giving a single value for the lexical contrasts involving statements. We repeated this operation for the lexical contrasts involving questions, the intonational contrasts involving Accent 1, and so on. Table 1 lists the resulting divergence points.

Subsequently, the ERFs were aligned in time according to the points of divergence, while retaining the baseline intervals time-locked to the onset of the words. The ERFs were low-pass filtered at 35 Hz. The analysis focussed on the difference waveform between standards and targets.

From the fields measured by axial gradiometers, we calculated the planar gradients of the MEG field distribution using a nearest-neighbour method yielding results compatible with the method described by Bastiaansen and Knösche (2000). The horizontal and vertical components of the estimated planar gradients approximate the signal measured by MEG systems with planar gradiometers. The *combined planar gradient* was then calculated from the root mean square of the horizontal and vertical planar field components (Helenius et al., 2002). The signals of the combined planar gradient fields are typically largest in magnitude directly above a given source (Hämäläinen et al., 1993). This is particularly advantageous when interpreting distributions of the magnitude of the ERFs at the sensor level.

REFERENCES

Bastiaansen, M.C., Knösche, T., 2000. Tangential derivative mapping of axial MEG applied to event-related

desynchronization research. Clin. Neurophysiol. 111, 1300–1305.

Baum, S., Pell, M., 1999. The neural bases of prosody: insights from lesion studies and neuroimaging. Aphasiology 13, 581–608.

Buchanan, T.W., Lutz, K., Mirzazade, S., Specht, K., Shah, N.J., Zilles, K., Jäncke, L., 2000. Recognition of emotional prosody and verbal components of spoken language: an fMRI study. Cogn. Brain Res. 9, 227–238.

Dehaene-Lambertz, G., 1997. Electrophysiological correlates of categorical phoneme perception in adults. NeuroReport 8, 919–924.

Démonet, J.F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J.-L., Wise, R., Rascol, A., Frackowiak, R., 1992. The anatomy of phonological and semantic processing in normal subjects. Brain 115, 1753–1768.

Endrass, T., Mohr, B., Pulvermüller, F., 2004. Enhanced mismatch negativity brain response after binaural word presentation. Eur. J. Neurosci. 19, 1653–1660.

Friederici, A.D., Alter, K., 2004. Lateralization of auditory language functions: a dynamic dual pathway model. Brain Lang. 89, 267–276.

Fujioka, T., Trainor, L.J., Ross, B., Kakigi, R., Pantev, C., 2004. Musical Training Enhances Automatic Encoding of Melodic Contour and Interval Structure. J. Cogn. Neurosci. 16, 1010–1021.

Gandour, J., 2007. Neural substrates underlying the perception of linguistic prosody. In: Gussenhoven, C., Riad, T. (Eds.), Tones and Tunes. : Experimental Studies in Word and Sentence Prosody, vol. 2. Mouton de Gruyter, Berlin/New York, pp. 3–25.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., Hutchins, G.D., 2000. A cross-linguistic PET study of tone perception. J. Cogn. Neurosci. 12, 207–222.

Gandour, J., Wong, D., Dzemidzic, M., Lowe, M., Tong, Y., Li, X., 2003a. A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. Hum. Brain Mapp. 18, 149–157.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Li, X., Satthamnuwong, N., Lurito, J., 2003b. Temporal integration of speech prosody is shaped by language experience: an fMRI study. Brain Lang. 84, 318–336.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., Lowe, M., 2004. Hemispheric roles in the perception of speech prosody. Neuroimage 23, 344–357.

Gussenhoven, C., 2000. The lexical tone contrast in Roermond Dutch in Optimality Theory. In: Horne, M. (Ed.), Intonation: Theory and experiment. Kluwer, Amsterdam, pp. 129–167.

Hagoort, P., Wassenaar, M., Brown, C., 2003. Real-time semantic compensation in patients with agrammatic comprehension: electrophysiological evidence for multiple-route plasticity. Proc. Natl. Acad. Sci. U. S. A. 100, 4340–4345.

Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography — theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev. Mod. Phys. 65, 413–497.

Helenius, P., Salmelin, R., Service, E., Connolly, J.F., Leinonen, S., Lyytinen, H., 2002. Cortical Activation during Spoken-Word Segmentation in Nonreading-Impaired and Dyslexic Adults. J. Neurosci. 22, 2936–2944.

Hickock, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92, 67–99.

Imaizumi, S., Mori, K., Kiritani, S., Hosoi, H., Tonoike, M., 1998. Task-dependent laterality for cue decoding during spoken language processing. NeuroReport 9, 899–903.

Indefrey, P., Cutler, A., 2005. Prelexical and lexical processing in listening. In: Gazzaniga, M.S. (Ed.), The new cognitive neurosciences III. MIT Press, Cambridge, pp. 759–774.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A.D., von Cramon, D.Y., 2004. Task-dependent laterality for cue decoding during spoken language processing. Brain Lang. 9, 277–289.

Näätänen, R., Lehtoskoskl, A., Lennes, M., Cheour, M., Huotilainen, M., Ilvonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., Alho, K., 1997. Language-specific phoneme representations revealed by electric and magnetic brain responses. Nature 385, 432–434.

Näätänen, R., 2001. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). Psychophysiology 38, 1–21.

Pell, M., 2002. Surveying Emotional Prosody in the Brain. In: Bel, B., Marlien, I. (Eds.), Proceedings of Speech Prosody 2002, Aix-en-Provence. Laboratoire Parole et Langage, Aix-en-Provence, pp. 77–82.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., Raichle, M.E., 1988. Positron emission tomographic studies of the cortical anatomy of single-word processing. Nature 331, 585–589.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., McGinnis, M., Roberts, T., 2000. Auditory cortex accesses phonological categories: an MEG mismatch study. J. Cogn. Neurosci. 12, 1038–1055.

Phillips, C., 2001. Levels of representation in the electrophysiology of speech perception. Cogn. Sci. 25, 711–731.

Shen, X.S., 1990. Prosody of Mandarin Chinese. University of California Press.

Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R.J., Näätänen, R., 1998. Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: magnetic mismatch negativity study. Neurosci. Lett. 251, 141–144.

Starkstein, S.E., Federoff, J.P., Price, R.C., Leiguarda, R.C., Robinson, R.G., 1994. Neuropsychological and neuroradiological correlates of emotional prosody comprehension. Neurology 44, 515–522.

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.

Van Lancker, D., 1980. Cerebral lateralization of pitch cues in the linguistic signal. Int. J. Hum. Commun. 13, 101–109.

Van Lancker, D., Sidtis, J.J., 1992. The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: all errors are not created equal. J. Speech Hear Res. 35, 963–970.

Wu, Z., 2000. From traditional Chinese phonology to modern speech technology: Realiziation of tone and intonation in spoken Chinese. Proceedings of the International Conference on Spoken Language Processing (ICSLP 2000), vol. 1, pp. B1–B12.

Zatorre, R.J., Evans, A.C., Meyer, E., Gjedde, A., 1992. Lateralization of phonetic and pitch discrimination in speech processing. Science 256, 846–849.