

The Neural Correlate of Speech Rhythm as Evidenced by Metrical Speech Processing

Eveline Geiser¹, Tino Zaehle¹, Lutz Jancke¹, and Martin Meyer^{1,2}

Abstract

■ The present study investigates the neural correlates of rhythm processing in speech perception. German pseudosentences spoken with an exaggerated (isochronous) or a conversational (nonisochronous) rhythm were compared in an auditory functional magnetic resonance imaging experiment. The subjects had to perform either a rhythm task (explicit rhythm processing) or a prosody task (implicit rhythm processing). The study revealed bilateral activation in the supplementary motor area (SMA), extending into the cingulate gyrus, and in the insulae, extending into the right basal ganglia (neostriatum), as well as activity in the right inferior frontal gyrus (IFG) related to the performance of the rhythm task. A direct contrast between isochronous and nonisochronous sentences revealed differences in lateralization of activation for isochronous processing as a function of the

explicit and implicit tasks. Explicit processing revealed activation in the right posterior superior temporal gyrus (pSTG), the right supramarginal gyrus, and the right parietal operculum. Implicit processing showed activation in the left supramarginal gyrus, the left pSTG, and the left parietal operculum. The present results indicate a function of the SMA and the insula beyond motor timing and speak for a role of these brain areas in the perception of acoustically marked temporal intervals. Secondly, the data speak for a specific task-related function of the right IFG in the processing of accent patterns. Finally, the data sustain the assumption that the right secondary auditory cortex is involved in the explicit perception of auditory suprasegmental cues and, moreover, that activity in the right secondary auditory cortex can be modulated by top-down processing mechanisms. ■

INTRODUCTION

The flow of speech is commonly perceived as having a particular rhythmic pattern. This rhythm is a pivotal structuring element of speech which is crucially involved in segmentation processes subserving language comprehension and even acquisition.

Recent evidence for this crucial function of speech rhythm has been found in electrophysiological research. Speech segmentation or the parsing of a sentence into words has been shown to rely on global patterns of prosodic phrasing which includes rhythmic and durational grouping as well as tonal pitch (Frazier, Carlson, & Clifton, 2006). At the word level, research has demonstrated that rhythmic patterning of German utterances facilitates spoken word identification as evidenced by an electrophysiological priming study (Friedrich, Kotz, Friederici, & Gunter, 2004). Depending on the rhythmic organization of one's native language, different segmentation strategies are observed. For French-speaking adults, the syllable appears to be the unit of segmentation (Segui, Djupoux, & Mehler, 1990), whereas English-speaking adults are guided by information about typical

word-stress patterns (Cutler & Norris, 1988). With respect to the segmentation strategy, the German language is comparable to the English language. The view that the rhythmic properties of language shape listeners' speech processing strategies has led to hypotheses of how infants develop efficient speech segmentation procedures in the acquisition of their mother tongue. There is evidence that infants' sensitivity to linguistic rhythm allows them to discriminate (nonnative) languages from different rhythmic classes (Nazzi & Ramus, 2003), suggesting a pivotal role of rhythm in language acquisition. Similar discrimination capabilities have even been reported from primate studies investigating tamarins (Tincoff et al., 2005).

The intuitive notion that spoken languages have characteristic underlying rhythmic patterns has resulted in a considerable amount of linguistic research aimed at assessing the exact acoustic characteristics of speech rhythm. This research was driven by the hypothesis of isochrony as an underlying principle of linguistic speech rhythm. Thus, a distinction is often made between stress-timed languages in which periodicity is assumed to be based on the regular occurrence of stressed syllables (e.g., English, German, Dutch) and syllable-timed languages, such as French, Italian, or Spanish, in which regularity is taken to be the outcome of all syllables being roughly of the same length (Pike, 1945).

¹University Zurich, Switzerland, ²University Hospital of Zurich, Switzerland

However, the phonological evidence for isochrony as the organizational principle of linguistic rhythm is still not unequivocal. As has been shown, the duration of interstress intervals in English is proportional to the number of syllables they contain, but can vary in duration influenced by the specific types of syllables they are comprised of as well as by the position of the interval within the utterance (Lehiste, 1977; Bolinger, 1965). Moreover, new statistical analyses of the speech signal, for instance, of consonantal and vocalic intervals (Ramus, Nespors, & Mehler, 1999) or of the auditory prominence of phonetic events (Lee & Todd, 2004), have also been inconclusive with respect to the identification of the acoustic characteristics of speech rhythm. Rhythmic differences between languages, context dependent variability of speech rhythm, and interindividual differences make identifying the acoustic as well as the phonological elements constituting speech rhythm experimentally difficult. Whereas linguists classify speech rhythm together with speech melody as prosodic aspects of speech using the term unspecifically for suprasegmental speech characteristics extending over more than one sound segment or phoneme, psycholinguistic researchers investigate speech rhythm as a concept in its own right. Such studies have included the field of language acquisition (Nazzi & Ramus, 2003), subsequent language learning (Curtin, 2005), speech segmentation (McQueen, Otake, & Cutler, 2001), and speech typologies (Ramus et al., 1999).

With respect to the neural correlate of speech rhythm, reliable evidence is still sparse. Because speech rhythm must be defined as a conglomerate of parameters with suprasegmental cues such as syllable duration, syllable stress, or pause being most important, it comes as no surprise that neuropsychological research has focused initially on these phonetic cues. In particular, specific event-related brain potentials, the “closure positive shift” (CPS) or the P350, have been observed to be modulated by the processing of speech accent or speech pauses (Friedrich, Kotz, Friederici, & Alter, 2004; Steinhauer, Alter, & Friederici, 1999). Furthermore, a recently published study has reported an increased N400 for sentences comprising a syllable lengthening in the last word (Magne et al., 2007). However, to our knowledge, the perception of speech rhythm as a composition of the mentioned suprasegmental characteristics and expanding over a longer period of time, such as the duration of a sentence, has never been the subject of investigation in functional brain research.

From the point of view of neuropsychological research, speech rhythm is of particular interest related to the question of functional hemispheric lateralization. Although the preponderance of the left hemisphere for speech processing is one of the best-evidenced facts in cognitive neuroscience, there is no doubt that the right hemisphere also contributes to speech perception (Jung-Beeman, 2005). However, the exact function of

the right hemisphere in speech processing has yet to be elucidated. The auditory association cortex, which is situated in the posterior supratemporal plane, has mostly been associated with a variety of auditory functions (Warren, Wise, & Warren, 2005), among others with speech melody processing (Homae, Watanabe, Nakano, Asakawa, & Taga, 2006; Gandour et al., 2004; Meyer, Steinhauer, Alter, Friederici, & von Cramon, 2004; Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002). Thus, one might hypothesize that speech rhythm processing, being classified as a prosodic aspect of speech, might also be related to a right temporal lobe function, more specifically the posterior superior temporal gyrus (pSTG). Support for this reasoning is provided by a current model of functional lateralization in language processing—the “asymmetric sampling in time” (AST) hypothesis (Poeppel, 2003). This model assumes that auditory fields in the two hemispheres prefer different temporal integration windows. Basically, the AST suggests that perception of rapidly changing cues (~ 40 Hz) preferentially drives the left hemisphere, whereas the right auditory cortex is better adept at processing of slowly changing acoustic cues (~ 4 Hz) available in spoken language. It is further proposed that the posterior portion of the auditory association cortex is the candidate region that accommodates this temporal processing. Speech rhythm involves the integration of rhythmic speech elements throughout a longer period such as the duration of a sentence. Thus, we hypothesize that speech rhythm perception should be considered a suprasegmental process that is preferentially supported by brain structures of the right hemisphere.

While investigating lateralization effects in language processing, a second line of research needs to be considered. It has been demonstrated by auditory functional imaging studies that the pattern of hemodynamic responses in frontal (Hsieh, Gandour, Wong, & Hutchins, 2001), but most interestingly, also in temporal cortical areas (Brechmann & Scheich, 2005; Tervaniemi & Hugdahl, 2003; Plante, Creusere, & Sabin, 2002), depends not only on the particular stimulus class but also on the task performed. Interestingly, this task-dependent modulation usually refers to a lateralization shift in auditory cortical areas involved in speech processing (Scheich, Brechmann, Brosch, Budinger, & Ohl, 2007; Noesselt, Shah, & Jancke, 2003; Tervaniemi & Hugdahl, 2003). Authors reporting task-dependent modulation effects have associated this effect with top-down processing, an interpretation that is further sustained by anatomical and functional plastic changes observed due to environmental and training influences (Luders, Gaser, Jancke, & Schlaug, 2004; Josse, Mazoyer, Crivello, & Tzourio-Mazoyer, 2003; Tervaniemi & Hugdahl, 2003). Assuming that lateralization in the temporal cortex might be influenced by task effects, we consider it very important to assess speech rhythm processing under different processing conditions.

The goal of the present study was to identify specific brain structures involved in the performance of a speech rhythm task. Furthermore, the study aimed to address the question of brain activation involved in speech rhythm processing in two different processing conditions, namely, explicit and implicit processing. Based on the aforementioned studies, we hypothesize involvement of the right peri-sylvian cortex in explicit speech rhythm processing.

METHODS

We investigated the processing of speech rhythm by using natural pseudosentences spoken with different rhythmic patterns. Hemodynamic responses were measured during explicit and implicit auditory processing.

Participants

Twenty-five subjects participated in this study. Twenty-four subjects, equally balanced over experimental groups, were analyzed (8 women, mean age = 27.4, *SD* = 9.3). One subject had to be excluded from the analysis due to a performance rate of below 63.8% correct answers. All participants were right-handed according to the Annett Handedness Questionnaire (Annett, 1992). They were native speakers of Swiss German with no history of neurological, major medical, psychiatric, or hearing disorders. Furthermore, subjects gave written consent in accordance with procedures approved by the local ethics committee and were paid for participation.

Stimuli

The stimulus material comprised a total of 144 German pseudosentences with natural speech rhythm manipulations. Pseudosentences follow the correct German syntax and phonology containing phonotactically legal pseudowords instead of content words (Friederici, Meyer, & von Cramon, 2000).

As mentioned in the Introduction, rhythmic distinctions can be found between languages. German is considered a stress-timed language that means it is the characteristic temporal distribution of accented syllables which constitutes its speech rhythm. This temporal distribution of accented syllables, where accents are marked by modulation of pitch or intensity, is not exactly isochronous in the spoken language, rather it is modulated by syllable lengthening and pauses. Notwithstanding, the German language shows a strong tendency toward isochrony (Noel Aziz Hanna, 2003; Völtz, 1991). This is especially evident in verse where speech becomes even more isochronous. We can conclude that spoken German can be described on a continuum be-

tween isochrony and nonisochrony, whereas verse is often isochronous. For the purpose of this study, we decided to use sentences with a normal conversational speech rhythm, later referred to as “nonisochronously” spoken sentences as well as “isochronously” spoken sentences.

The stimuli were controlled for syntactic differences across experimental conditions. The latter is supposed to represent an exaggerated form of German speech rhythm. Additionally, the “isochronously” spoken sentences followed a regular meter (i.e., iambs, trochees, dactyls) and “nonisochronously” spoken sentences followed an irregular meter (i.e., iambs or trochees with a dactyl interposed between two metrical feet).

Examples:

“Der Speiter pongt den spiten Galtung” = „isochronous“
 x | X x | X x | X x | X x |
 „Der Jüfele knelt den fürten Pflaster“ = „nonisochronous“
 x | X x x | X x | X x | X x |

The trained female speaker was instructed to say the “nonisochronous” sentences with a normal conversational speech rhythm. Before recording the “isochronous” sentences, the speaker listened to a metronome beat of approximately the average tempo of the “nonisochronous” sentences. To ensure that the “isochronous” condition stayed an ecological condition, the speaker did not have a constant reference to the metronome while speaking. Instead, she was instructed to follow the original beat as precisely as possible without making the utterance unnatural. Additionally, pseudosentences were recorded with the intonation contour of either a question or a statement balanced over the two rhythmic conditions (Table 1).

Additionally, a low-level auditory control condition consisting of isochronous syllables (e.g., “da de di do du”; *n* = 36) as well as a total of 40 null events as baseline condition were randomly included in the time course of the experiment.

All stimulus items were normalized in amplitude to 70% of the loudest signal in a stimulus item. Then, all stimulus items were analyzed by means of the PRAAT speech editor (Boersma & Weenink, 2000) for mean

Table 1. Experimental Conditions

Sentence Type	Processing Modus	
	Explicit (<i>n</i> = 12)	Implicit (<i>n</i> = 12)
Isochronous (<i>n</i> = 72)	question statement	question statement
Nonisochronous (<i>n</i> = 72)	question statement	question statement

duration, mean intensity (dB), and intensity on a root-means-square based measure. Stimuli were balanced with respect to mean duration and intensity on the root-means-square based measure over the experimental condition. The isochronous stimuli showed a mean intensity of 68.7 dB, whereas the “nonisochronous” sentences showed a mean intensity of a 67.4 dB.

Experimental Groups/Task

Subjects were randomly assigned to two different experimental groups, each group having to perform a different categorization task on the same set of stimuli (Table 1). The “explicit processing” group ($n = 12$) had to judge whether they had heard an “isochronous” or “non-isochronous” pseudosentence (rhythm task). The “implicit processing” group ($n = 12$) had to judge whether they had heard a statement or question (prosody task). Subjects indicated their response by pressing a button with the index and middle fingers, respectively, of their right hand. Subjects of the implicit processing group were naive with respect to the rhythmic manipulation and did not detect a rhythmic difference between the stimuli throughout the experiment as confirmed by a short debriefing after the experiment. No feedback was given during the experiment.

Procedure

In a short training session conducted prior to the fMRI experiment, subjects were made familiar with the task. We used an MR-compatible piezoelectric auditory stimulation system incorporated into standard Philips headphones for binaural stimulus delivery. Stimuli were presented in pseudorandom order binaurally. A total of 220 trials (144 pseudosentences, 36 low-level auditory control items, 40 null events) were presented in two runs. Before each stimulus presentation, a fixation cross was presented for 500 msec.

fMRI Design

We implemented a clustered sparse temporal acquisition technique that combines the principles of a sparse temporal acquisition with a clustered acquisition of three consecutive volume scans per trial (Schmidt et al., 2007; Zaehle et al., 2007). We collected three consecu-

tive volumes in order to cover the peak of the event-related hemodynamic signal (Figure 1).

Data Acquisition

Data were collected using a Philips Intera 3-T whole-body MR unit (Philips Medical System Best, The Netherlands) equipped with an eight-channeled Philips SENSE head coil. Functional time series were collected from 16 transverse slices covering the entire perisylvian cortex with a spatial resolution of $2.7 \times 2.7 \times 4$ mm using a Sensitivity Encoded (SENSE) (Pruessmann, Weiger, Scheidegger, & Boesiger, 1999) single-shot, gradient-echo planar sequence (acquisition matrix 80×80 voxels, SENSE acceleration factor $R = 2$, FOV = 220 mm, TE = 35 msec and interslice gap 2 mm). Three volumes were acquired per trial with each a Tacq = 1000 msec, $\theta = 68^\circ$ (decay sampling) and 12 sec inter-cluster interval. Furthermore, we collected a standard 3-D T1-weighted scan for anatomical reference with $1 \times 1 \times 1.5$ mm spatial resolution (180 axial slices, acquisition matrix 224×224 voxels, TE = 2.3 msec, TR = 20 msec, $\theta = 20^\circ$).

Data Analysis

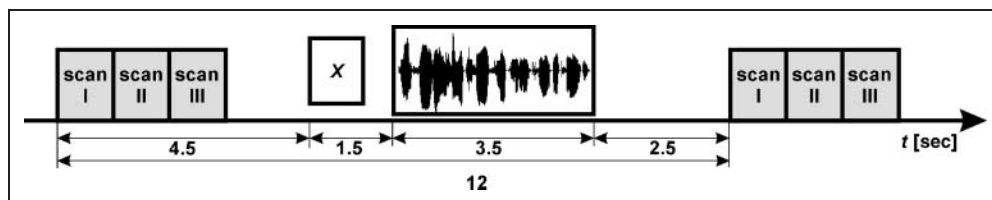
Behavioral Data

During the experiment, the behavioral performance of the two experimental groups was measured. Data were corrected for outliers (>2 std above or below mean value). Behavioral measures were aggregated by participants and conditions. As a measure for accuracy of discrimination, the mean percentage of correct answers over all experimental conditions (without low-level auditory control condition) was calculated. An independent-sample t test was performed to identify group differences.

fMRI Data

To account for different T1 saturation effects in subsequent volumes, we subjected the three volume scans (TPI–III) collected during each cluster to three separate analyses during preprocessing and individual statistics. The functional imaging data processing was carried out using MATLAB 6.5 (Mathworks, Natick, MA, USA) and the SPM99 software package (Wellcome Department of Cognitive Neurology, London, UK; www.fil.ion.ucl.ac.uk/spm/).

Figure 1. Acquisition scheme. The figure demonstrates the three time points of acquisition (TPI, TPII, TPIII) and the stimulus presentation in a single trial.



Functional data were realigned to the first volume, corrected for motion artifacts, and normalized into standard stereotactic space (voxel size $2 \times 2 \times 2$ mm, template provided by the Montreal Neurological Institute; Evans et al., 1992). For spatial smoothing, we applied an isotropic Gaussian kernel of 8 mm full-width-at-half-maximum. Low-frequency drifts were removed using a temporal high-pass filter (mean cutoff of 307 sec).

Statistical evaluation was based on a least-square estimation using the general linear model for serially autocorrelated observations, performed separately on each voxel (Friston et al., 1995). Single trials were treated as epochs and modeled by means of a boxcar function. Conditions were compared by calculating contrasts between conditions of interest for each participant and time point of acquisition. Direct contrasts between isochronous and nonisochronous sentences as well as between experimental conditions and the low-level auditory control condition were calculated. We limited the presentation of results to TPII as this acquisition time is supposed to reflect the amplitude peak of the hemodynamic response as it has been done formerly (Meyer, Baumann, Marchina, & Jancke, 2007; Bunzeck, Wuestenberg, Lutz, Heinze, & Jancke, 2005). Contrast images were submitted to a second-level group analysis. Random effects analysis consisted of between- and within-group comparison. The between-group analysis consisted of a two-sampled t test using “all sentences > auditory control” contrasts. The within-group analysis consisted of a paired t test using “isochronous sentences > auditory control” and “nonisochronous sentences > auditory control” contrasts. Results were thresholded at $T = 3.50$ for the two-sampled t test and $T = 4.02$ for the paired t tests ($p = .001$, uncorrected for multiple comparisons). Only clusters of significant size ($p < .05$, corrected for multiple comparison with a spatial extent of $k = 45$ for the two-sampled t test and $k = 27$ for the paired t test) were reported (Worsley et al., 1996).

RESULTS

Behavioral Results

Due to technical problems, the performance data of four subjects, two of each group, could not be recorded. The implicit processing group (prosody task, $n = 10$) performed better than the explicit processing group (rhythm task, $n = 10$). 80.25% of the answers were correct for the explicit processing group and 98.6% for the implicit processing group. A two-sampled t test performed to identify group differences revealed a significant difference in response accuracy [$t(18) = -5.28, p < .000$].

fMRI Results

The present article addressed two main issues. First, we were interested in the neurofunctional correlate related

to the performance of the speech rhythm task. Second, we aimed to analyze the brain areas involved in the processing of exaggerated German speech rhythm in an explicit and an implicit processing condition.

Performance of the Speech Rhythm Task

Between-group comparison between explicit and implicit processing of all sentences (Figure 2, Table 2) revealed activation in the SMA for the explicit processing group compared to the implicit processing group. This activation was found bilaterally, although the peak activation was in the right hemisphere. The cluster extended into the medial part of the superior frontal gyrus and the right cingulate gyrus. Furthermore, this contrast revealed bilateral activation in the left and right insula encroaching onto the neostriatum, namely, the putamen in the right hemisphere. In the right hemisphere, we found an additional activation in the inferior frontal gyrus (IFG, pars opercularis).

Explicit Processing of Isochronous Speech Rhythm

To reveal the effect of speech rhythm, we compared isochronously spoken sentences to nonisochronously spoken sentences in the explicit processing condition

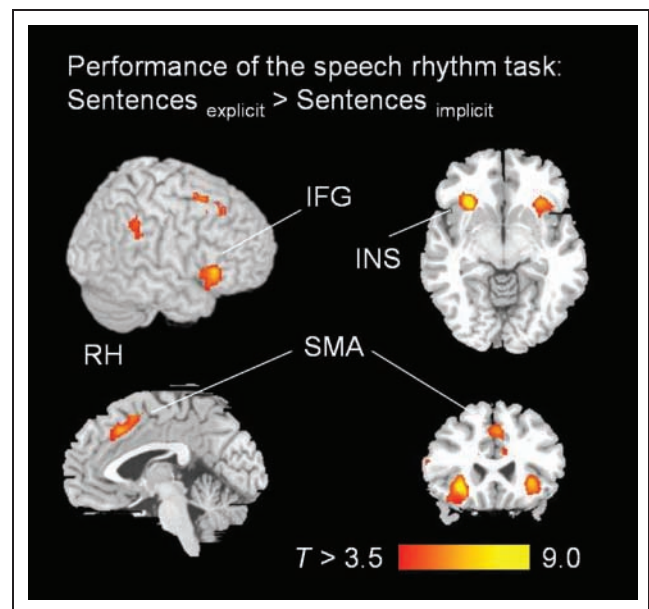


Figure 2. Brain areas that showed significantly greater activation during the performance of the rhythm task (explicit) as compared to the implicit processing condition. The figure shows resulting T -maps of the two-sampled t test using sentences versus auditory control contrast for the explicit as compared to the implicit processing condition (up right $z = -7$, down left $x = 2$, down right $y = 25$). The top-left image shows the activations projected on the cortical surface. Activation is thresholded at $T = 3.50$ ($p < .001$, uncorrected) and cluster corrected for multiple comparisons at $p < .05$ (Worsley et al., 1996). Figures are displayed in neurological convention.

Table 2. Performance of the Speech Rhythm Task

Anatomical Description	Left Hemisphere				Right Hemisphere			
	<i>p</i> *	<i>T</i>	Cluster Size	<i>x,y,z [mm]</i>	<i>p</i> *	<i>T</i>	Cluster Size	<i>x,y,z [mm]</i>
INS extending into NS					.000	7.57	212	27 21 -9
IFG (pars opercularis)					.000	6.89	168	51 12 12
SMG					.000	6.38	257	42 -42 36
INS extending into IFG (pars orbitalis/triangularis)	.000	6.3	176	-30 24 -3				
SMA and cingulate gyrus extending into SFG	.000	6.12	249	0 21 42				
IPL	.041	4.89	46	-39 -42 42				

Two-sampled *t* test: (all sentences > auditory control)_{expl} > (all sentences > auditory control)_{impl}.

Coordinates are given according to MNI space.

INS = insula; NS = neostriatum (caudate/head, putamen); IFG = inferior frontal gyrus; IPL = inferior parietal lobe; SFG = superior frontal gyrus; SMA = supplementary motor area; SMG = supplementary motor area.

**p* Values corrected for entire volume on cluster level ($k = 45$).

(Figure 3, Table 3) in a direct contrast. We observed statistically significant activation in right superior temporal gyrus extending into the supramarginal gyrus and the parietal operculum.

Implicit Processing of Isochronous Speech Rhythm

The contrast between isochronously spoken and non-isochronously spoken sentences in the implicit processing condition (Figure 3, Table 4) revealed activation in the left hemisphere, namely, the supramarginal gyrus and the superior temporal gyrus extending into the parietal operculum for the isochronously spoken sentences compared to nonisochronously spoken sentences.

Taken together, direct contrast between isochronous and nonisochronous sentences revealed a substantial effect observed for the processing of isochronously spoken speech, which was differentially lateralized in the explicit as compared to the implicit processing condition. Furthermore, the performance of the speech rhythm task relied on activation of the SMA as well as the insula bilaterally and the right IFG.

DISCUSSION

The main goal of the study was to identify brain areas related to speech rhythm processing, that is, the performance of a speech rhythm task. A further objective was

Figure 3. Brain areas showing significantly greater activation during the processing of isochronous compared to nonisochronous sentences in (A) the explicit processing condition (right $x = 48$) and (B) the implicit processing condition (right $z = 18$). For both the explicit and implicit processing conditions, the activation in the sagittal view is projected on the lateral convexity of cortical surface. Each cluster is thresholded at $T = 4.02$ ($p < .001$, uncorrected) and cluster corrected for multiple comparisons at $p < .05$ (Worsley et al., 1996). Figures are displayed in neurological convention.

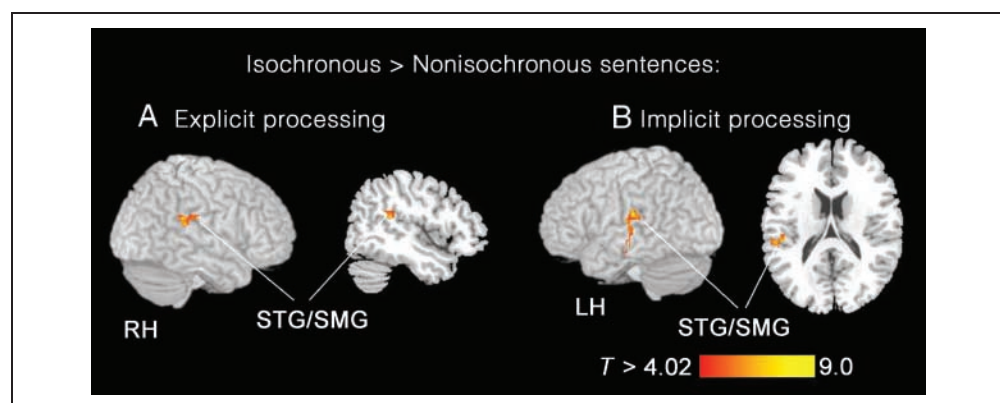


Table 3. Explicit Isochronous Speech Rhythm Processing

<i>Anatomical Description</i>	<i>Left Hemisphere</i>				<i>Right Hemisphere</i>			
	<i>p*</i>	<i>T</i>	<i>Cluster Size</i>	<i>x,y,z [mm]</i>	<i>p*</i>	<i>T</i>	<i>Cluster Size</i>	<i>x,y,z [mm]</i>
STG extending into SMG/PaOp	.002	6.72	49					48 -39 18

Paired *t* test: (Isochronous sentences > auditory control) > (Nonisochronous sentences > auditory control).

Coordinates are given according to MNI space.

STG = superior temporal gyrus; SMG = supplementary motor area; PaOp = parietal operculum.

**p* Values corrected for entire volume on cluster level ($k = 27$).

to determine possible differences between explicit and implicit rhythm perception. In the following section we will first discuss the brain activation evoked by the performance of the speech rhythm task. Then we will discuss the differential effects found for isochronous processing as compared to nonisochronous processing in both the explicit and, subsequently, the implicit processing conditions.

Performance of the Speech Rhythm Task

A clear effect related to the explicit performance of the rhythm task was observed in the *mesial premotor cortex (pre-SMA/SMA-proper)*, in contrast to the implicit processing group. We suggest that this finding gives rise to the notion of a general function of the mesial premotor cortex, which is related to timing processes in auditory perception as described in the following section. Considering other recent findings, this result speaks for a multimodal function of the SMA in temporal perception.

The SMA has frequently been reported to be involved in motor processes such as finger movement or speech production (Riecker et al., 2005). In the context of motor research, the SMA is divided into pre-SMA, associated with motor preparation, and SMA-proper, associated with movement execution, encoding, and retrieval of motor sequences (Lau, Rogers, & Passingham, 2006; Cunnington, Windischberger, Robinson, & Moser, 2005; Kansaku et al., 2005; Koeneke, Lutz, Wustenberg, & Jancke, 2004; Lewis, Wing, Pope, Praamstra, & Miall, 2004; Lutz, Specht, Shah, & Jancke, 2000). However, there is evidence that this involvement of the SMA-proper as well as the pre-SMA in motor processes might be due to a specific function in the timing of motor

actions, a function which might not only be restricted to motor processing. For example, both areas have been found to show an increasing relation between complexity, that is, temporal complexity, of motor sequences and activation (Riecker, Kassubek, Gröschel, Grodd, & Ackermann, 2006; Bengtsson, Ehrsson, Forssberg, & Ullen, 2005; Lewis et al., 2004; Lutz et al., 2000).

The assumption of a temporal processing function of the SMA is sustained by evidence from perception research in the visual domain, which described a correlation between attention to time and activity in the SMA (Coull, 2004; Nobre & O'Reilly, 2004; Lewis & Miall, 2003). A similar finding has been reported in the sensorimotor domain showing an involvement of the SMA in time estimation after a specific stimulation (Macar, Anton, Bonnet, & Vidal, 2004). Interpreting the abovementioned studies in the direction of a multimodal involvement of the SMA in time perception (Pastor, Macaluso, Day, & Frackowiak, 2006; Schubotz, Friederici, & von Cramon, 2000), our study poses an enhancement of this assumption in that we used acoustic stimuli and found a very specific activation in the SMA related to temporal perception. Further research will specify the different processes underlying timing perception. Furthermore, our results found in the contrast between sentences and low-level auditory control condition clearly speak for a task-driven effect of sequential timing perception.

One might argue that the two tasks are not of equal difficulty as suggested by the performance rate and that the SMA activity might be induced by this difference. A recently published study argues that SMA in all timing tasks is not related to timing but rather to supplementary activities (Livesey, Wall, & Smith, 2007).

Table 4. Implicit Isochronous Speech Rhythm Processing

<i>Anatomical Description</i>	<i>Left Hemisphere</i>				<i>Right Hemisphere</i>			
	<i>p*</i>	<i>T</i>	<i>Cluster Size</i>	<i>x,y,z [mm]</i>	<i>p*</i>	<i>T</i>	<i>Cluster Size</i>	<i>x,y,z [mm]</i>
SMG extending into STG/PaOp	.000	7.45	75	-57 -30 21				

Paired *t* test: (Isochronous sentences > auditory control) > (Nonisochronous sentences > auditory control).

STG = superior temporal gyrus; SMG = supplementary motor area; PaOp = parietal operculum.

**p* Values corrected for entire volume on cluster level ($k = 27$).

On the other hand, studies that aimed to disentangle task difficulty and temporal processing associated activity in the pre-SMA and SMA-proper specifically with temporal processing (Coull, 2004; Macar et al., 2004). Moreover, studies investigating task difficulty in a pitch discrimination task did not report an effect in SMA (Reiterer et al., 2005). In our opinion, this is an indication that task difficulty is not significantly involved in our data.

An alternative interpretation of the SMA activation could be given in terms of strategy mechanisms. It is possible that the SMA activation relates to spontaneous and intuitive synchronization of rhythm perception and body movements. That is, subjects intuitively “feel” the rhythm in their body while listening to the specific sentences. This assumption is sustained by reports showing that auditory cues are very effective in facilitating accurate rhythmic body movements (Patel, Iversen, Chen, & Repp, 2005) as well as by comments made by subjects after the performance of the experiment. This sensation could be mediated by motor cortices such as the SMA. However, spontaneous motor synchronization mechanisms must be a topic of further investigations.

Furthermore, our results show an extensive activation in the *insula of both hemispheres* and in the *right IFG*, specifically the *right pars opercularis*. Besides its function in motor (Jantzen, Steinberg, & Kelso, 2005) and speech production processes (Bohland & Guenther, 2006; Ackermann & Riecker, 2004; Blank, Scott, Murphy, Warburton, & Wise, 2002), specifically in motor timing, the insula has repeatedly been found in various auditory processes (Mutschler et al., 2007; Giraud et al., 2004; Bamiou, Musiek, & Luxon, 2003). Of note are auditory temporal processing (Lewis & Miall, 2003; Ackermann et al., 2001; Lewis, Beauchamp, & DeYoe, 2000), speech melody perception (Meyer et al., 2002), and auditory feedback and pitch control (Zarate & Zatorre, 2005) which have recently been reported. Moreover, functional lateralization in the insulae has been suggested on the basis of temporal acoustic characteristics (Ackermann et al., 2001) as well as on the function of pitch-related stimuli (Wong, Parsons, Martinez, & Diehl, 2004). Very recently, Bamiou et al. (2006) gave strong support to the notion of a crucial role of the insula in temporal processing, in particular, temporal resolution and sequencing. Thus, there is an evolving view that the insula is crucial in auditory timing perception. Our processing task involves the perception of the specific accent patterns, that is, the sequencing of syllables. Therefore, our finding of an insular involvement in the sequencing task is not surprising and in line with the abovementioned hypothesis about insular functions.

Of particular note is the activation in the right insula extending into the *basal ganglia (putamen)*. Lesion and imaging studies revealed an involvement of the basal ganglia in the processing of suprasegmental speech

cues, namely, emotional speech melody in speech production (Pell, Cheang, & Leonard, 2006) and speech melody perception (Meyer, Steinhauser, Alter, Friederici, & von Cramon, 2004; Kotz et al., 2003). Furthermore, lesion studies have found that patients with a basal ganglia insult also show symptoms of impaired prosodic functions (Van Lancker Sidtis, Pachana, Cummings, & Sidtis, 2006). A very recent finding has indicated a role of the basal ganglia in auditory beat perception (Grahn & Brett, 2007). These data expand the function of the basal ganglia over the commonly reported functions in motor processing. Moreover, our data strongly suggest that the involvement of the basal ganglia in the processing of auditory suprasegmental cues might not be limited to melodic aspects of speech but be necessary for the perception of rhythmic aspects of speech as well.

We want to put an additional emphasis on the activity observed in the right IFG. The right IFG has repeatedly been reported in addition to left frontal brain areas in speech processing and has been associated with increasing sentence complexity (Michael, Keller, Carpenter, & Just, 2001), syntax processing (Poldrack et al., 2001; Meyer, Friederici, & von Cramon, 2000; Ni et al., 2000), or the performance of specific speech tasks (Chou et al., 2006; Meyer et al., 2000). However, there is also a considerable amount of research which has associated the right IFG with specific acoustic perception (i.e., pitch processing). Activation in the right IFG has been found during pitch processing in both nonlinguistic (Liu et al., 2006; Hsieh et al., 2001; Humphries, Willard, Buchsbaum, & Hickok, 2001) and speech contexts (Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006; Gandour et al., 2004; Meyer et al., 2002). Pitch processing plays a crucial role in speech processing of tonal languages, that is, languages where the pitch of a word carries a specific semantic content (Gandour et al., 2003). In nontone languages, pitch becomes important in the context of speech melody perception (Meyer et al., 2002). In German language, the perception of accent patterns includes both pitch and intensity patterns. Thus, in our study, the activation in the right IFG must be interpreted more generally in relation to the integration of accent patterns. Furthermore, in our specific task, subjects processed the stimuli focusing on the accent pattern independently of a linguistic context. Thus, we strongly believe that our data provide evidence that the right IFG can be selectively activated in tasks requiring higher suprasegmental cue processing.

Explicit and Implicit Processing of Isochronous Speech Rhythm

We analyzed the effect of speech rhythm, isochronous versus nonisochronous sentences, in both the explicit and the implicit processing conditions. This analysis is based on recent findings of differentially lateralized

effects in auditory processing depending on the context of the stimulus perception (Brechmann & Scheich, 2005) as well as on findings reporting a top-down modulation on auditory processing (Tervaniemi & Hugdahl, 2003).

In the explicit processing condition, we measure task-driven processing of isochronous speech. Interestingly, this process leads to temporal activation in the right hemisphere, namely, activation in the *posterior part of the right STG* (specifically planum temporale, PT) extending into the supramarginal gyrus, for the isochronous as compared to the nonisochronous sentences. Most theories of the functional role of the PT now assume that the PT in the left and the right hemispheres is involved in a parameter specific segregation of auditory sounds leading to differentially lateralized processes of the two hemispheres. Two different parameter-specific concepts of the PT are found in the literature, however, they are widely considered to be two sides of the same coin. One concept posits that the posterior part of the STG is sensitive to spectrotemporal auditory processing and associates the right STG with spectral processing and the left STG with temporal auditory processing (Meyer et al., 2005; Zaehle, Wustenberg, Meyer, & Jancke, 2004; Griffiths & Warren, 2002; Menon et al., 2002; Zatorre & Belin, 2001). A complimentary model, the AST, specifies the potential functions of the pSTG slightly differently (Poeppel, 2003). This model proposes that the left non primary auditory cortex is preferentially driven by rapidly changing acoustic cues, whereas the right non primary auditory cortex is more amenable to slowly changing acoustic cues due to the specific temporal integration window. The perception of suprasegmentally modulated speech rhythm that unfolds in the time range of syllables (150–200 msec) should, therefore, more likely be supported by the right posterior temporal plane. Thus, we state that our findings can be explained by the AST of functional lateralization. However, our results have a second highly important implication. We find the right hemisphere to be involved in the processing of isochronous speech rhythm in the explicit processing condition only. The same contrast in the implicit processing condition does not show right hemispheric activation. Therefore, our results additionally provide strong evidence for attention modulated activation in the auditory cortex. More precisely, we believe that this gives strong evidence for a top-down modulation on lateralization effects in the auditory cortex (Brechmann & Scheich, 2005; Tervaniemi & Hugdahl, 2003).

One might reason why we did not find activation in the anterior superior temporal cortex of the right hemisphere in response to speech rhythm perception. Such activation has been found in melody perception and production (Brown, Martinez, Hodges, Fox, & Parsons, 2004; Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002), as well as in the perception of nonspeech vocal sounds (Belin, Zatorre, & Ahad, 2002). We argue that speech rhythm relies strongly on linguistic information,

which has not been found to involve the right anterior superior temporal brain areas.

In the implicit processing condition, we measure unattended processing of speech rhythm. Furthermore, as confirmed after the completion of the experiment, subjects were not aware of the rhythmic manipulation within the sentences. Surprisingly, the contrast between isochronous and nonisochronous sentences revealed significant activation in the *left supramarginal gyrus* extending into the *left posterior superior temporal gyrus (STG)*, more specifically, the PT and extending into the left parietal operculum for the isochronous condition as compared to the nonisochronous condition. The extent of activation observed in the left pSTG encroached onto the upper-posterior part of the Sylvian fissure and the supramarginal gyrus. The extension of activity over these cortical areas is not surprising. It has been shown that subjects with a larger left PT tend to show a larger leftward functional asymmetry of several peri-sylvian areas, namely, the inferior parietal lobule outside the supramarginal gyrus, Heschl's gyrus, the Rolandic operculum, and the temporal pole (Josse et al., 2003). This finding suggests possible functional common ground between these cortex areas. Hickok, Buchsbaum, Humphries, and Muftuler (2003) first described activation in the parietal operculum and in the adjacent posterior part of the Sylvian fissure, at the parietal-temporal boundary for both speech listening tasks and covert rehearsal of spoken language. The authors suggested that this area might be involved in auditory-motor activity similar to the visual-motor response properties in the dorsal visual stream (Rizzolatti, Fogassi, & Gallese, 1997). In the context of our study, it is important to take notice of another line of research reporting that musical meter, as the whole-numbered (i.e., isochronous) sequence of accented acoustic events, is the characteristic feature of music which relates to body movements (Lerdahl & Jackendoff, 1983). Very recently, Chen, Zatorre, and Penhune (2006) investigated the neural correlate of a synchronization process of body movements using the meter of an auditory cue. The authors found the left PT, the right pSTG, and the bilateral dorsal premotor cortex to covariate with metric salience, and thus, to facilitate the synchronization of motor actions to auditory cues. Therefore, we suggest that listening to exaggerated speech rhythm in the implicit processing condition leads to involuntary auditory-motor interactions due to the salience of isochronously spoken sentences.

Conclusion

In the context of an fMRI study on speech rhythm perception, we found the SMA, right IFG, bilateral insulae, as well as the right basal ganglia, to be involved in sequential timing perception related to the performance of a rhythmic categorization task on speech

stimuli. These data provide evidence for an involvement of the SMA in the perception of time intervals across modalities and for a specific task-related function of the right IFG in the processing of accent patterns in speech.

Furthermore, we observed a clear effect of speech rhythm, which was lateralized to the right auditory cortex in an explicit, task-driven processing condition. This observation can be explained by the fact that suprasegmental speech rhythm perception occurs over a long temporal integration window. Interestingly, this right lateralized activation did not occur in an implicit processing condition. This finding indicates that right auditory association cortex activity can be modulated by top-down processing and can be selectively recruited depending on task demands. The same contrast in an implicit, stimulus-driven processing condition evoked differentially lateralized brain activation in the left hemisphere. This finding is assumed to reflect an involuntary auditory-motor interaction effect.

Acknowledgments

This work was supported by the Swiss National Foundation (SNF-grant 3200B0-105877 conferred to M. Meyer). We thank the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig for assistance in stimulus recording. We specially thank Dr. Dorothea Weniger from the University Hospital of Zurich and Elsa Sägeser for helpful comments on previous versions of this manuscript. Furthermore, we thank Dr. Conny F. Schmidt from the Institute for Biomedical Technology, ETH and University of Zurich, Switzerland, for assistance in the data acquisition.

Reprint requests should be sent to Eveline Geiser, Department of Neuropsychology, Institute of Psychology, University of Zurich, Binzmühlestrasse 14/25, CH-8050 Zurich, Switzerland, or via e-mail: e.geiser@psychologie.unizh.ch.

REFERENCES

- Ackermann, H., & Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: A review and a hypothesis. *Brain and Language*, *89*, 320–328.
- Ackermann, H., Riecker, A., Mathiak, K., Erb, M., Grodd, W., & Wildgruber, D. (2001). Rate-dependent activation of a prefrontal-insular-cerebellar network during passive listening to trains of click stimuli: An fMRI study. *NeuroReport*, *12*, 4087–4092.
- Annett, M. (1992). 5 Tests of hand skill. *Cortex*, *28*, 583–600.
- Bamiou, D. E., Musiek, F. E., & Luxon, L. M. (2003). The insula (Island of Reil) and its role in auditory processing. Literature review. *Brain Research, Brain Research Reviews*, *42*, 143–154.
- Bamiou, D. E., Musiek, F. E., Stow, I., Stevens, J., Cipolotti, L., Brown, M. M., et al. (2006). Auditory temporal processing deficits in patients with insular stroke. *Neurology*, *67*, 614–619.
- Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Research, Cognitive Brain Research*, *13*, 17–26.
- Bengtsson, S. L., Ehrsson, H. H., Forssberg, H., & Ullen, F. (2005). Effector-independent voluntary timing: Behavioural and neuroimaging evidence. *European Journal of Neuroscience*, *22*, 3255–3265.
- Blank, S. C., Scott, S. K., Murphy, K., Warburton, E., & Wise, R. J. (2002). Speech production: Wernicke, Broca and beyond. *Brain*, *125*, 1829–1838.
- Boersma, P., & Weenink, D. (2000). PRAAT: Doing phonetics by computer. [Computer software]. Institute of Phonetic Sciences, University of Amsterdam.
- Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage*, *32*, 821–841.
- Bolinger, D. (1965). *Pitch accent and sentence rhythm, forms of English: Accent, morpheme, order*. Cambridge: Harvard University Press.
- Brechmann, A., & Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cerebral Cortex*, *15*, 578–587.
- Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., & Parsons, L. M. (2004). The song system of the human brain. *Brain Research, Cognitive Brain Research*, *20*, 363–375.
- Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H. J., & Jancke, L. (2005). Scanning silence: Mental imagery of complex sounds. *Neuroimage*, *26*, 1119–1127.
- Chen, J. L., Zatorre, R. J., & Penhune, V. B. (2006). Interactions between auditory and dorsal premotor cortex during synchronization to musical rhythms. *Neuroimage*, *32*, 1771–1778.
- Chou, T. L., Booth, J. R., Burman, D. D., Bitan, T., Bigio, J. D., Lu, D., et al. (2006). Developmental changes in the neural correlates of semantic processing. *Neuroimage*, *29*, 1141–1149.
- Coull, J. T. (2004). fMRI studies of temporal attention: Allocating attention within, or towards, time. *Brain Research, Cognitive Brain Research*, *21*, 216–226.
- Cunnington, R., Windischberger, C., Robinson, S., & Moser, E. (2005). The selection of intended actions and the observation of others' actions: A time-resolved fMRI study. *Neuroimage*, *29*, 1294–1302.
- Curtin, S., Mintz, T. H., & Christiansen, M. H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition*, *96*, 233–262.
- Cutler, E. A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. In G. T. M. Altmann (Ed.), *Psycholinguistics: Critical concepts* (pp. 157–177). London: Routledge.
- Evans, A. C., Marrett, S., Neelin, P., Collins, L., Worsley, K., Dai, W., et al. (1992). Anatomical mapping of functional activation in stereotactic coordinate space. *Neuroimage*, *1*, 43–53.
- Frazier, L., Carlson, K., & Clifton, C., Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *10*, 244–249.
- Friederici, A. D., Meyer, M., & von Cramon, D. Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language*, *75*, 289–300.
- Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Alter, K. (2004). Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. *Brain Research, Cognitive Brain Research*, *20*, 300–308.
- Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Gunter, T. C. (2004). ERPs reflect lexical identification in word fragment priming. *Journal of Cognitive Neuroscience*, *16*, 541–552.
- Friston, K. J., Holmes, A. P., Worsley, K. P., Poline, J. B., Frith, C. D., & Frackowiak, R. S. (1995). Statistical parameter maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*, 189–210.
- Gandour, J., Tong, Y., Wong, D., Talavage, T., Dziedzic, M., Xu, Y., et al. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage*, *23*, 344–357.

- Gandour, J., Xu, Y., Wong, D., Dziedzic, M., Lowe, M., Li, X., et al. (2003). Neural correlates of segmental and tonal information in speech perception. *Human Brain Mapping, 20*, 185–200.
- Giraud, A. L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M. O., Preibisch, C., et al. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex, 14*, 247–255.
- Grahn, J. A., & Brett, M. (2007). Rhythm and beat perception in motor areas of the brain. *Journal of Cognitive Neuroscience, 19*, 893–906.
- Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences, 25*, 348–353.
- Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience, 15*, 673–682.
- Homae, F., Watanabe, H., Nakano, T., Asakawa, K., & Taga, G. (2006). The right hemisphere of sleeping infant perceives sentential prosody. *Neuroscience Research, 54*, 276–280.
- Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language, 76*, 227–252.
- Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: An fMRI study. *NeuroReport, 12*, 1749–1752.
- Jantzen, K. J., Steinberg, F. L., & Kelso, J. A. (2005). Functional MRI reveals the existence of modality and coordination-dependent timing networks. *Neuroimage, 25*, 1031–1042.
- Josse, G., Mazoyer, B., Crivello, F., & Tzourio-Mazoyer, N. (2003). Left planum temporale: An anatomical marker of left hemispheric specialization for language comprehension. *Brain Research, Cognitive Brain Research, 18*, 1–14.
- Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences, 9*, 512–518.
- Kansaku, K., Muraki, S., Umeyama, S., Nishimori, Y., Kochiyama, T., Yamane, S., et al. (2005). Cortical activity in multiple motor areas during sequential finger movements: An application of independent component analysis. *Neuroimage, 28*, 669–681.
- Koecke, S., Lutz, K., Wustenberg, T., & Jancke, L. (2004). Bimanual versus unimanual coordination: What makes the difference? *Neuroimage, 22*, 1336–1350.
- Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language, 86*, 366–376.
- Lau, H., Rogers, R. D., & Passingham, R. E. (2006). Dissociating response selection and conflict in the medial frontal surface. *Neuroimage, 29*, 446–451.
- Lee, C. S., & Todd, N. P. (2004). Towards an auditory account of speech rhythm: Application of a model of the auditory “primal sketch” to two multi-language corpora. *Cognition, 93*, 225–254.
- Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5*, 253–263.
- Lerdahl, F., & Jackendoff, R. (1983). An overview of hierarchical structure in music. *Music Perception, 1*, 229–252.
- Lewis, J. W., Beauchamp, M. S., & DeYoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex, 10*, 873–888.
- Lewis, P. A., & Miall, R. C. (2003). Brain activation patterns during measurement of sub- and supra-second intervals. *Neuropsychologia, 41*, 1583–1592.
- Lewis, P. A., Wing, A. M., Pope, P. A., Praamstra, P., & Miall, R. C. (2004). Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping. *Neuropsychologia, 42*, 1301–1312.
- Liu, L., Peng, D., Ding, G., Jin, Z., Zhang, L., Li, K., et al. (2006). Dissociation in the neural basis underlying Chinese tone and vowel production. *Neuroimage, 29*, 515–523.
- Livesey, A. C., Wall, M. B., & Smith, A. T. (2007). Time perception: Manipulation of task difficulty dissociates clock functions from other cognitive demands. *Neuropsychologia, 45*, 321–331.
- Luders, E., Gaser, C., Jancke, L., & Schlaug, G. (2004). A voxel-based approach to gray matter asymmetries. *Neuroimage, 22*, 656–664.
- Lutz, K., Specht, K., Shah, N. J., & Jancke, L. (2000). Tapping movements according to regular and irregular visual timing signals investigated with fMRI. *NeuroReport, 11*, 1301–1306.
- Macar, F., Anton, J. L., Bonnet, M., & Vidal, F. (2004). Timing functions of the supplementary motor area: An event-related fMRI study. *Brain Research, Cognitive Brain Research, 21*, 206–215.
- Magne, C., Astésano, C., Aramaki, M., Ystad, S., Kronalnd-Martinet, R., & Besson, M. (2007). Influence of syllabic lengthening on semantic processing in spoken French: Behavioral and electrophysiological evidence. *Cerebral Cortex, 17*, 2659–2668.
- McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language, 45*, 103–132.
- Menon, V., Levitin, D. J., Smith, B. K., Lembke, A., Krasnow, B. D., Glazer, D., et al. (2002). Neural correlates of timbre change in harmonic sounds. *Neuroimage, 17*, 1742–1754.
- Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping, 17*, 73–88.
- Meyer, M., Baumann, S., Marchina, S., & Jancke, L. (2007). Haemodynamic responses in human multisensory and auditory association cortex to purely visual stimulation. *BMC Neuroscience, 8*. doi:10.1186/1471-2002-8-14.
- Meyer, M., Friederici, A. D., & von Cramon, D. Y. (2000). Neurocognition of auditory sentence comprehension: Event related fMRI reveals sensitivity to syntactic violations and task demands. *Brain Research, Cognitive Brain Research, 9*, 19–33.
- Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language, 89*, 277–289.
- Meyer, M., Zaehle, T., Gountouna, V. E., Barron, A., Jancke, L., & Turk, A. (2005). Spectro-temporal processing during speech perception involves left posterior auditory cortex. *NeuroReport, 16*, 1985–1989.
- Michael, E. B., Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). fMRI investigation of sentence comprehension by eye and by ear: Modality fingerprints on cognitive processes. *Human Brain Mapping, 13*, 239–252.
- Mutschler, I., Schulze-Bonhage, A., Glauche, V., Demandt, E., Speck, O., & Ball, T. (2007). A rapid sound–action association effect in human insular cortex. *PLOS ONE, 2*, e259. doi:10.1371/journal.pone.0000259.

- Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication*, *41*, 233–243.
- Ni, W., Constable, R. T., Mencl, W. E., Pugh, K. R., Fulbright, R. K., Shaywitz, S. E., et al. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *Journal of Cognitive Neuroscience*, *12*, 120–133.
- Nobre, A. C., & O'Reilly, J. (2004). Time is of the essence. *Trends in Cognitive Sciences*, *8*, 387–389.
- Noel Aziz Hanna, P. (2003). *Sprachrhythmus in Metrik und Alltagssprache*. München: Wilhelm Fink Verlag.
- Noesselt, T., Shah, N. J., & Jancke, L. (2003). Top-down and bottom-up modulation of language related areas—An fMRI Study. *BMC Neuroscience*, *4*, 13.
- Pastor, M. A., Macaluso, E., Day, B. L., & Frackowiak, R. S. (2006). The neural basis of temporal auditory discrimination. *Neuroimage*, *30*, 512–520.
- Patel, A. D., Iversen, J. R., Chen, Y., & Repp, B. H. (2005). The influence of metricality and modality on synchronization with a beat. *Experimental Brain Research*, *163*, 226–238.
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, *36*, 767–776.
- Pell, M. D., Cheang, H. S., & Leonard, C. L. (2006). The impact of Parkinson's disease on vocal-prosodic communication from the perspective of listeners. *Brain and Language*, *97*, 123–134.
- Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.
- Plante, E., Creusere, M., & Sabin, C. (2002). Dissociating sentential prosody from sentence processing: Activation interacts with task demands. *Neuroimage*, *17*, 401–410.
- Poeppl, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as “asymmetric sampling in time”. *Speech Communication*, *41*, 245–255.
- Poldrack, R. A., Temple, E., Protopapas, A., Nagarajan, S., Tallal, P., Merzenich, M., et al. (2001). Relations between the neural bases of dynamic auditory processing and phonological processing: Evidence from fMRI. *Journal of Cognitive Neuroscience*, *13*, 687–697.
- Pruessmann, K. P., Weiger, M., Scheidegger, M. B., & Boesiger, P. (1999). SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, *42*, 952–962.
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292.
- Reiterer, S. M., Erb, M., Droll, C. D., Anders, S., Ethofer, T., Grodd, W., et al. (2005). Impact of task difficulty on lateralization of pitch and duration discrimination. *NeuroReport*, *16*, 239–242.
- Riecker, A., Kassubek, J., Gröschel, K., Grodd, W., & Ackermann, H. (2006). The cerebral control of speech tempo: Opposite relationship between speaking rate and BOLD signal changes at striatal and cerebellar structures. *Neuroimage*, *29*, 46–53.
- Riecker, A., Mathiak, K., Wildgruber, D., Erb, M., Hertrich, I., Grodd, W., et al. (2005). fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology*, *64*, 700–706.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (1997). Parietal cortex: From sight to action. *Current Opinion in Neurobiology*, *7*, 562–567.
- Scheich, H., Brechmann, A., Brosch, M., Budinger, E., & Ohl, F. W. (2007). The cognitive auditory cortex: Task-specificity of stimulus representations. *Hearing Research*, *229*, 213–224.
- Schmidt, C. F., Zaehle, T., Meyer, M., Geiser, E., Boesiger, P., & Jancke, L. (2007). Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task. *Human Brain Mapping*, doi:10.1002/hbm.20372.
- Schubotz, R. I., Friederici, A. D., & von Cramon, D. Y. (2000). Time perception and motor timing: A common cortical and subcortical basis revealed by fMRI. *Neuroimage*, *11*, 1–12.
- Segui, J., Djupoux, E., & Mehler, J. (1990). The role of the syllable in speech segmentation, phoneme identification and lexical access. In G. Altmann & R. Sillcock (Eds.), *Cognitive models of speech processing* (pp. 263–280). Cambridge: MIT Press.
- Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*, 191–196.
- Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research, Brain Research Reviews*, *43*, 231–246.
- Tincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., & Mehler, J. (2005). The role of speech rhythm in language discrimination: Further tests with a non-human primate. *Developmental Science*, *8*, 26–35.
- Van Lancker Sidtis, D., Pachana, N., Cummings, J. L., & Sidtis, J. J. (2006). Dysprosodic speech following basal ganglia insult: Toward a conceptual framework for the study of the cerebral representation of prosody. *Brain and Language*, *97*, 135–153.
- Völtz, M. (1991). Das Rhythmusphänomen. *Zeitschrift für Sprachwissenschaft*, *10*, 284–296.
- Warren, J. E., Wise, R. J., & Warren, J. D. (2005). Sounds do-able: Auditory-motor transformations and the posterior temporal plane. *Trends in Neurosciences*, *28*, 636–643.
- Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research*, *156*, 249–268.
- Wong, P. C., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *Journal of Neuroscience*, *24*, 9153–9160.
- Worsley, K. P., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, *4*, 58–73.
- Zaehle, T., Schmidt, C. F., Meyer, M., Baumann, S., Baltes, C., Boesiger, P., et al. (2007). Comparison of “silent” clustered and sparse temporal fMRI acquisitions in tonal and speech perception tasks. *Neuroimage*, *37*, 1195–1204.
- Zaehle, T., Wustenberg, T., Meyer, M., & Jancke, L. (2004). Evidence for rapid auditory perception as the foundation of speech processing: A sparse temporal sampling fMRI study. *European Journal of Neuroscience*, *20*, 2447–2456.
- Zarate, J. M., & Zatorre, R. J. (2005). Neural substrates governing audiovisual integration for vocal pitch regulation in singing. *Annals of the New York Academy of Sciences*, *1060*, 404–408.
- Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*, 946–953.