Neural correlates of rhythm processing in speech perception

Eveline Geiser, Conny Schmidt, Lutz Jancke & Martin Meyer

Department of Neuropsychology
University Zurich, Switzerland
e.geiser@psychologie.unizh.ch

Abstract

The present study investigates the neural correlates of speech rhythm perception. Metric and non-metric German pseudo-sentences were compared in an auditory fMRI experiment. One group of subjects was to decide which type of sentence they had heard (explicit processing). A second group performed a prosody task on the same stimuli (implicit processing). As an active baseline condition isochronous syllables were presented. Group analysis revealed activation in the supplementary motor area (SMA) for the explicit processing group. This activation was not present in the implicit processing group. A direct contrast between the metric and the non-metric sentences for the implicit processing group revealed significant activation in the left planum temporale (PT) for the metric condition. Our results suggest that rhythm processing relies on neural correlates different from those related to speech melody processing. The implicit perception of unexpected speech rhythm relies on brain areas which have previously been associated with temporal auditory processing in the left hemisphere.

1. Introduction

In the realm of speech perception rhythm is a phenomenon that is of significant importance. Many behavioral studies have shown that attending to rhythmic aspects of speech facilitates speech perception [1-3] and even infants are capable of perceiving rhythmic differences in language [4]. To our knowledge, neural correlates of speech rhythm perception have not yet been investigated.

This is partially due to the fact that of all prosodic aspects of speech, rhythm has proved to be the most problematic to define. Up to now, the phonological or acoustic correlate of speech rhythm has not yet precisely been identified, even though researchers have investigated speech rhythm since the early 19th century [5]. The variability of speech rhythm in one language as a consequence of inter-individual and contextual differences account for the difficulties in experimentally identifying the elements constituting speech rhythm.

Researchers formerly defined speech rhythm as an isochronous recurrence of a specific type of speech unit [6;7], suggesting that phonological speech units, such as syllables, might be the basis of speech rhythm. Yet, so far, researchers have failed to identify the correlate of these recurrent speech units. For speakers of German language, Kohler [8;9] found that the recurrence of accented syllables in time is at least modulated proportionally to the number of syllables between accents in order to keep the speech tempo constant. Ramus et al. [10] have suggested a promising alternative to classifying languages in rhythmic categories, proposing measures for proportions of vocalic intervals (%V), standard deviation of vocalic intervals (Δ V) within the sentence, and standard deviation of consonantal intervals (Δ C) within the sentence. They suggest that these measures might be related to intuitive rhythm types of languages.
However, up to now, neither the acoustic nor the phonological correlate of the perceived speech rhythm was concordantly determined.

Currently, the concept of speech rhythm is used in research areas of language acquisition [11], subsequent language learning [12], speech segmentation [13] and speech typologies [10]. The term is commonly used unspecifically for suprasegmental speech characteristics such as syllable duration, syllable stress, or pause.

Assuming that the perception of speech rhythm is based on a conglomerate of suprasegmental speech elements such as accents, tone pitch, and pauses, the existing neuropsychological evidence was reviewed. Prosodic aspects of speech such as accents, intonation or pauses have earlier been associated with specific EKP components [1;2]. To our knowledge, no one has investigated the neural correlate of speech rhythm unfolding over a period of several seconds as is the case in spoken sentences. Hypothesing, that accents might specifically be the basis of rhythm in German language [9] we assume that the perception of speech rhythm has to rely partially on intensity and pitch processing, and thus on melodic aspects of speech.
However, evidence from related research fields, e.g. lesion studies in music research, shows that rhythm perception can be selectively impaired in perception and performance and thus seem to rely on a different neural network than music melody perception [14-17]. Imaging studies have identified differentially lateralized effects of melody processing and musical temporal processing in adults [18;19]. Thus, there is converging evidence that melody and rhythm in music perception are probably based on different neural correlates. This observation gives rise to the notion, that both processes in speech perception might be functionally independent as well.

To summarize, even though the concept of speech rhythm is considered to be relevant for speech perception and used in various domains of language research, the acoustic or phonological correlate of speech rhythm has not yet been concordantly identified. The rhythmic elements of German speech rhythm specifically are assumed to rely on prosodic aspects such as accents, which can acoustically be defined as changes in pitch and intensity. The goal of our study was to provide complementary evidence on the nature of speech rhythm by investigating the perception of rhythm on sentence level rather than the perception of its constituent elements.

We hypothesized to find anatomically distinct brain areas related to speech rhythm processing as compared to speech melody processing.

German pseudo-sentences spoken with different rhythmic patterns were used. We measured cortical activity during explicit (task induced) and implicit (stimulus induced) auditory processing.

2. Methods

2.1. Materials

2.1.1. Subjects

25 paid volunteers (16 male, mean age 28.4 years) participated in this study. One subject had to be excluded from the analysis due to a performance rate of about 50 percent correct answers. Subjects were native speakers of German (Swiss). None of them had a history of neurological, major medical, psychiatric or hearing disorders. All participants were right handed according to the Annett handedness scale [20]. After information about the fMRI procedure was given, written informed consent was obtained from all participants.

2.1.2. Stimuli

228 German pseudo-sentences were presented throughout the experiment. Experimental conditions were constructed according to a 2*2-design including four experimental conditions (metric vs. non-metric and question vs. statement). 36 stimuli were assigned to each of the 4 experimental conditions.
The "metric" sentences followed a regular meter (i.e. jambs, trochees, dactyls) and were spoken in such a way that stressed syllables followed each other isochronously. "Nonmetric" sentences followed an irregular meter (i.e. a dactyl interposed between two jambs or trochees) and were spoken with a normal conversational speech rhythm.

Examples of a „metric"sentence:
"Der Speiter pongt den spiten Galtung"

Example of a „non-metric" sentence:
„ Der Jüfele knelt den furten Pflaster"

Prosody / Rhythm	Statement	Question
Metric	36	36
Non-metric	36	36

Figure 1: *Experimental conditions*

Following the principle of cognitive subtraction, an active baseline condition consisting of isochronous syllables (e.g. "da de di do du") were presented. 36 stimuli were assigned to this baseline condition. Additionally, 48 empty trials were included in the experiment. The sentence material was constructed according to the phonotactical rules of the German language. Stimuli were spoken by a trained German speaker and controlled for syntax, stimulus length, and intensity on a root-mean-square based measure. The stimuli underwent analysis by means of the PRAAT speech editor [21].

2.1.3. Experimental Groups / Task

Subjects were assigned to two different experimental groups. The "rhythm group" (n=12) had to decide which rhythmic type of sentence (metric / non-metric) they had heard. The "prosody group" (n=12) had to decide which prosodic type of sentence (statement / question) they heard. Subjects had to indicate their response by pressing a button with their right hand. No feedback was given during the experiment.

2.2. Procedure

In a short training session conducted before the experiment, subjects learned to perform the given task. Stimuli were presented in pseudorandom order and presented binaurally through headphones in the scanner. The study employed a single-trial design [22]. The experiment was conducted in two sessions of 114 stimuli with a short break in between the sessions. Before stimulus presentation a fixation cross was presented for 500 ms.

2.2.1. fmri design

We implemented a clustered sparse temporal acquisition technique that combines the principles of a sparse temporal acquisition (STA) with a clustered acquisition of three consecutive volume scans per trial (Schmidt et al., submitted). The rationale of this acquisition technique is the presentation of the auditory stimulus in silence on the one hand, and on the other, the long inter-scan interval (inter stimulus interval of 12 s) allows both the functional response to the auditory stimulus and the response evoked by the scanner noise to decay prior to the next trial. Thus, this approach is capable of clearly separating the task-induced functional response from the scanner-noise induced functional response.

2.2.2. Data acquisition

Data were collected using a Philips Intera 3T whole body MR unit (Philips Medical System Best, The Netherlands) equipped with an eight-channeled Philips SENSE head coil. Functional time series were acquired in 16 transverse slices covering auditory cortex with a high spatial resolution of 2.7 x 2.7 x 4 mm using a Sensitivity Encoded (SENSE) [23] single-shot gradient-echo planar sequence (acquisition matrix 80 x 80, SENSE acceleration factor R = 2, FOV = 220 mm, TE = 35 ms and inter-slice gap 2 mm). With the CTA schema, three volumes were acquired per trial with each a Tacq=1000 ms, θ = 68o (decay sampling) and 12 s inter scan interval (ISI). Additionally, a standard 3D T1 weighted scan was obtained for anatomical reference.

2.2.3. Data analysis

The functional imaging data processing was carried out using MATLAB 6.5 (Mathworks Inc., Natick, MA, USA) and the software package SPM99 (Wellcome Department of Cognitive Neurology, London, UK, http://www.fil.ion.ucl.ac.uk/spm/). Functional data were realigned to the first volume, corrected for motion artifacts and (mean-adjusted by proportional scaling) normalized (non-linear spatial transformation with 7 x 8 x 7 basis functions) into standard stereotactic space (template provided by the Montreal Neurological Institute). For spatial smoothing we applied an isotropic Gaussian kernel

(8 mm full-width-at-half-maximum). We used a high-pass filter (default SPM cut-off frequency) to eliminate low-frequency signal changes.
Statistical evaluation was based on a least-square estimation using the general linear model for serially autocorrelated observations, performed separately on each voxel [24]. Single trials were treated as epochs and modeled by means of a box car function.
Conditions were compared by calculating direct contrasts between conditions for each participant and time point of acquisition (TA). Contrast images were submitted to a second level group analysis. The group analysis consisted of one-sampled t-tests across the specific contrast image of all participants in one experimental group. Group comparison consisted of a two sampled t-test across the specific contrasts of both groups. All resulting t-statistical parametrical maps were thresholded at T = 4.02 ($p<001$; uncorrected for multiple statistical comparison). Only clusters of significant size ($p < 0.05$, corrected for multiple comparisons) were reported [25]. Only activations obtained during the second TA are reported in this paper (see also [26]).

3. Results and Discussion

3.1.1. Behavioral results

During the experiment the behavioral performance of 20 subjects, reaction times for correctly answered trials and response accuracy were measured. Due to technical problems, the performance data of 4 subjects could not be recorded. 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 baseline) was calculated. 80.25% were correct for the rhythm group and 98.6% for the prosody group. Mean reaction times were 4451 ms for the rhythm group and 3985 ms for the prosody group. An independent sample t-test performed to identify group differences revealed significant difference in the response accuracy (F1=17.490; p = 0.001) and no significant differences in reaction time (F1 < 0.05; p=883).

3.1.2. fmri results

A comparison of the two sentence types relative to the active baseline condition in the rhythm task showed an activation in the supplementary motor area (SMA), in the inferior frontal gyrus (IFG) extending into the anterior insula bilaterally, and in the inferior parietal lobe of the right hemisphere. This activation was not observed in the prosody group ($p<0.05$, see Fig 2, Tab. 1). Thus, it seems obvious, that rhythm processing relies on brain areas different from speech melody processing. However, SMA is not one of the classic language perception areas. It has previously been observed in motor preparation, such as the preparation of rhythmic speech production [27] and has also been found for encoding as well as retrieval of motor sequences [28]. Thus it seems possible that auditory rhythmic categorization might rely on motor representation and that the activation in the SMA is task induced rather than stimulus induced. However, another line of research has attributed SMA to timing-related functions in attentional and perceptive processes in the visual domain [29]. Thus, our results suggest complementary, that speech rhythm processing is based on trans-modal brain areas related to attention.
A direct contrast between metric and non-metric sentences in the prosody task revealed activation in the anterior part of the left superior temporal gyrus (STG) and in the posterior part of the superior temporal gyrus (pSTG), namely the planum temporale (PT) extending into the Heschl gyrus (HG). The PT is a secondary auditory region which has been associated with the performance of various speech perception tasks, specifically the temporal processing of auditory signals [30]. The metrical condition compared to the non-metrical condition is assumed to violate the perception of "normal", conversational speech with respect to timing. Thus, these results suggest that processing temporally "deviant" speech depends on more extensive temporal processing than conversational speech even in implicit speech perception.

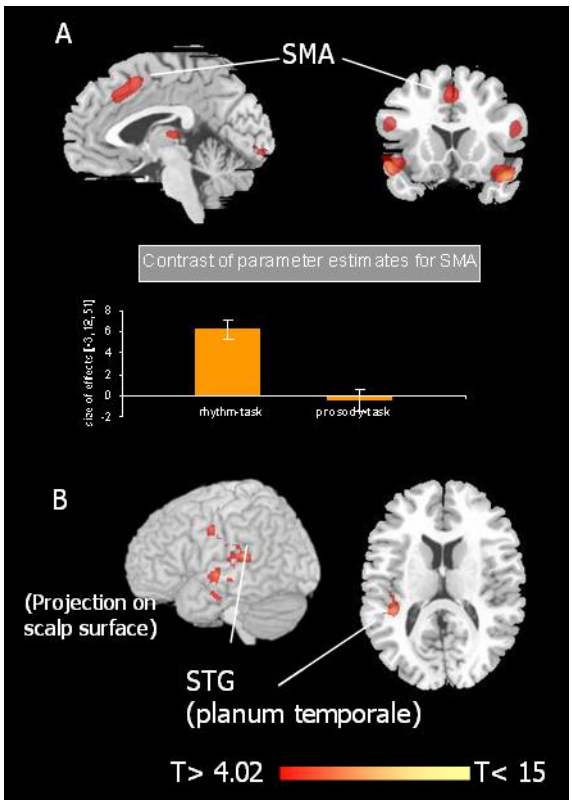

Figure 2: *(A) Comparison between sentences and active baseline condition in the rhythm group; different size of effects for SMA [-3, 12, 51] of the two experimental groups ($p< 0.5$). (B) Comparison between metric and non-metric sentences in the prosody group*

4. Conclusion

The present data suggest that the explicit processing of speech rhythm might rely on brain areas which are functionally distinct from areas related to speech melody processing, and furthermore, that implicit perception of speech rhythm relies on secondary auditory cortex areas of the left hemisphere. Further investigations are necessary to decide, if these findings allow a conclusion concerning the acoustic elements of speech rhythm perception.

References

[1] Steinhauer,K., Alter,K., Friederici,A.D., Brain potentials indicate immediate use of prosodic cues in natural speech processing, Nat. Neurosci., 2 (1999) 191-196.

[2] Friedrich,C.K., Kotz,S.A., Friederici,A.D., Alter,K., Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence, Brain Res. Cogn Brain Res., 20 (2004) 300-308.

[3] Cutler,A., Segmentation Problems, Rhythmic Solutions, Lingua, 92 (1994) 81-104.

[4] Nazzi,T., Ramus,F., Perception and acquisition of linguistic rhythm by infants, Speech Communication, 41 (2003) 233-243.

[5] Lloyd,A.J., Speech signals in telephony, London, 1940.

[6] Pike,K., The Intonation of Amercian English, University of Michigan Press, Ann Arbor, 1946.

[7] Abercrombie,D., Elements of General Phonetics, Edinburgh University Press, Edinburgh, 1967.

[8] Kohler,K. Rhytmus im Deutschen. 19, 89-105. 1982. Arbeitsberichte, Institut für PHonetik der Universität Kiel.

[9] Kohler,K. Stress-timing and speech Rate in German. A Production Model. 20, 7-53. 1983. Arbeitsberichte, INstitut für Phonetik der Universität Kiel.

[10] Ramus,F., Nespor,M., Mehler,J., Correlates of linguistic rhythm in the speech signal, Cognition, 73 (1999) 265-292.

[11] Nazzi,T., Ramus,F., Perception and acquisition of linguistic rhythm by infants, Speech Communication, 41 (2003) 233-243.

[12] Curtin,S.M.T.H.C.M.H., Stress changes the representational landscape: Evidence from word segmentation, Cognition, (2006).

[13] McQueen,J.M., Otake,T., Cutler,A., Rhythmic cues and possible-word constraints in Japanese speech segmentation, Journal of Memory and Language, 45 (2001) 103-132.

[14] Peretz,I., Kolinsky,R., Boundaries of separability between melody and rhythm in music discrimination: a neuropsychological perspective, Q. J. Exp. Psychol. A, 46 (1993) 301-325.

[15] Murayama,J., Kashiwagi,T., Kashiwagi,A., Mimura,M., Impaired pitch production and preserved rhythm production in a right brain-damaged patient with amusia, Brain Cogn, 56 (2004) 36-42.

[16] Liegeois-Chauvel,C., Peretz,I., Babai,M., Laguitton,V., Chauvel,P., Contribution of different cortical areas in the temporal lobes to music processing, Brain, 121 (Pt 10) (1998) 1853-1867.

[17] Patel,A.D., Peretz,I., Tramo,M., Labreque,R., Processing prosodic and musical patterns: a neuropsychological investigation, Brain Lang, 61 (1998) 123-144.

[18] Zatorre,R.J., Neural specializations for tonal processing, Biological Foundations of Music, 930 (2001) 193-210.

[19] Samson,S., Ehrle,N., Baulac,M., Cerebral substrates for musical temporal processes, Biological Foundations of Music, 930 (2001) 166-178.

[20] Annett,M., 5 Tests of Hand Skill, Cortex, 28 (1992) 583-600.

[21] Boersma,P., Weenink,D. PRAAT: Doing phonetics by computer. 2000. Instiute of phonetic Sciences, University of Amsterdam.

[22] D'Esposito,M., Zarahn,E., Aguirre,G.K., Event-related functional MRI: Implications for cognitive psychology, Psychological Bulletin, 125 (1999) 155-164.

[23] Pruessmann,K.P., Weiger,M., Scheidegger,M.B., Boesiger,P., SENSE: Sensitivity encoding for fast MRI, Magnetic Resonance in Medicine, 42 (1999) 952-962.

[24] Friston,K.J., Holmes,A.P., Worsley,K.P., Poline,J.B., Frith,C.D., Frackowiak,R.S., Statistical parameter maps in functional imaging: A general linear approach, Hum. Brain Mapp., 2 (1995) 189-210.

[25] Worsley,K.P., Marrett,S., Neelin,P., Vandal,A.C., Friston,K.J., Evans,A.C., A unified statistical approach for determining significant signals in images of cerebral activation , Hum. Brain Mapp., 4 (1996) 58-73.

[26] Glover,G.H., Deconvolution of impulse response in event-related BOLD fMRI, Neuroimage, 9 (1999) 416-429.

[27] Riecker,A., Wildgruber,D., Dogil,G., Grodd,W., Ackermann,H., Hemispheric lateralization effects of rhythm implementation during syllable repetitions: an fMRI study, Neuroimage., 16 (2002) 169-176.

[28] Heun,R., Freymann,N., Granath,D.O., Stracke,C.P., Jessen,F., Barkow,K., Reul,J., Differences of cerebral activation between superior and inferior learners during motor sequence encoding and retrieval, Psychiatry Res., 132 (2004) 19-32.

[29] Coull,J.T., fMRI studies of temporal attention: allocating attention within, or towards, time, Brain Res. Cogn Brain Res., 21 (2004) 216-226.

[30] Zaehle,T., Wustenberg,T., Meyer,M., Jancke,L., Evidence for rapid auditory perception as the foundation of speech processing: a sparse temporal sampling fMRI study, Eur. J. Neurosci., 20 (2004) 2447-2456.