PROSODY IN SPEECH PRODUCTION: A fMRI STUDY

Jörg Mayer[1], Grzegorz Dogil[1], Dirk Wildgruber[2], Axel Riecker[2],
Hermann Ackermann[2], Wolfgang Grodd[2]
[1]*Experimental Phonetics, University of Stuttgart;* [2]*Neuroradiology, University of Tübingen*

ABSTRACT

Clinical observations of distortions of production (and perception) of prosody implicate that distinct, non-overlapping neural circuits are responsible for distinct prosodic cues and functions [1]. These observations motivate a question whether similar evidence can be found in the neurologically intact brain. In our experiments we use new technology designed to reveal the function of active, healthy brain. The experiments are constructed to check the neuroanatomical basis of the PROSODY GENERATOR, a functional unit in the phonological system which integrates and controls the variation of prosodic parameters [2]. The results show that relatively small, non-overlapping, distinct perisilvian areas of both the right and the left hemisphere are involved in the generation of prosody. We found specific activity correlated with the FOCUS accent, the MODUS marker and the AFFECT characterization. The localization appears to correlate best with the address frame of the prosodic cues. These results can not be fully accommodated by any of the existing theories of prosodic representation of speech in the human brain.

1. INTRODUCTION

Prosody is a mode of communication which provides a parallel channel to speech. Prosodic features, unlike other linguistic features, are often produced without conscious intention and are open to forms of interpretation which rely on emotional, non-cognitive processes. The communicative content of many prosodic signals parallels that of stereotypic call vocalizations characteristic of communication systems of other species. It has been often argued that the neuroanatomical basis for these call vocalizations should be fundamentally different from the neuroanatomical basis of the symbolic aspects of human communication. But unlike calls of other species, prosodic organization of human communication is continuous and highly correlated with the semantic, syntactic, morphological and segmental organization of speech. Regardless of function, there exist only three prosodically active phonetic parameters: duration, intensity and pitch.

The variety of prosody functions and cues in language processing has led to multiple hypotheses concerning the neurolinguistic and neuroanatomical basis of prosody. At least four contradictory hypotheses have been particularly influential (cf. [1] for a critical summary).

(1) The RIGHT HEMISPHERE HYPOTHESIS contends that all aspects of prosody are independently processed by the right hemisphere and integrated with the linguistic information (which is processed by the left hemisphere) via interhemispheric connections (i.e. the fibres of the corpus callosum).

(2) The FUNCTIONAL LATERALIZATION HYPOTHESIS assumes that there is a continuum from linguistic to affective functions of prosody and processing shifts from the left hemisphere (more linguistically-based tasks) to the right hemisphere (more affectively-based tasks).

(3) The SUBCORTICAL PROCESSING HYPOTHESIS claims that prosodic functions are highly dependent on subcortical processing and are not lateralized to one or another hemisphere.

(4) The ACOUSTIC CUES HYPOTHESIS contends that duration, pitch (and possibly intensity) may be independently lateralized.

All these contradictory hypotheses find their support in the clinical observation of language and speech impaired subjects. There are problems with the interpretation of data from patients if it is used in isolation. This data may reflect neural reorganization or the development of compensatory strategies. It can not be simply assumed that the absence of function after a stroke means that the patient has normal cognition minus one part. Apart from that, the data provided by observation of patients are a product of a highly complex cognitive process which can be hardly further fractionalized. Modern cognitive theories question the assumption of a simple correspondence between complex tasks (like prosody) and large brain areas (like whole brain hemispheres).

In our research we assume a highly fractionalised and elaborated model of prosody generation (Fig. 1) and test its individual components with experiments designed to reveal the function of active, healthy brain. The experiments follow the methodological spirit of [2]. The main interest in the experiment described below concerns the role of the **address frames** in prosody generation.

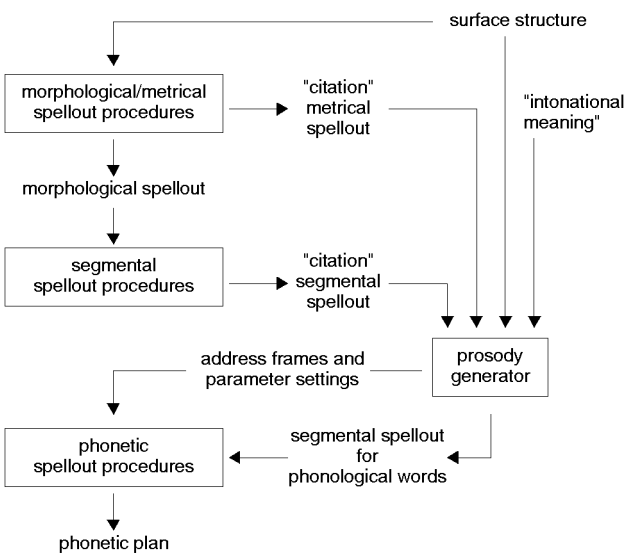

Figure 1. The model of prosody generation [2: 366].

2. METHODS

2.1. Materials

The study recruited healthy native german subjects (five females, four males, mean age 26.2 years, range 21-32 years). All participants were right-handed as determined by strandardized inventory, and none of them had a history of neurological disorders. Informed consent had been obtained from each subject. Subjects were paid for the participation in the experiment. The subjects were asked to produce a sentence like sequence consisting of five syllables [dadadadada] with various pitch-accent types and locations (the FOCUS condition), with various boundary tone types (the MODUS condition), and with various kinds of emotional state marking (the AFFECT condition). As a baseline for the statistical analysis they were asked to produces the logatomes [dadadadada, dididididi, dododododo, dududududu] in a monotonous voice (with a syllable frequency of ca. 5 Hz). The material is summarized in table 1.

We used reiterrant syllables and meaningless words in order to reduce to the minimum the influence of the syntactic, semantic, morphological and segmental factors on prosody generation. The aspects of prosody that were controlled in this experiment were in accordance with the model of prosody given in [2] correlated only with different address frames and parameter settings (cf. Fig. 1).

Stimulus	Paradigm	
	1 (FOCUS)	4 (MONOTONOUS)
○ ○ ○ ● ○	dadada**da**da H*L	dididididi
○ ○ ○ ○ ●	dadada**da**da H*L	dadadadada
● ● ○ ○ ○	**da**dada**da**da H*L H*L	dududududu
○ ○ ○ ● ●	**da**dada**da**da L*H L*H	dododododo
	2 (MODUS)	4 (MONOTONOUS)
•	dadada**da**da H*L L%	dididididi
?	dadada**da**da L*H H%	dadadadada
	3 (AFFECT)	4 (MONOTONOUS)
☺	dadada**da**da H*L [HAPPY]	dududududu
☹	dadada**da**da H*L [SAD]	dododododo

Table 1. Visual presented stimuli and reaction paradigms.

2.2. Procedure

Subjects lie supine in the MR scanner (1.5 T whole body scanner, Siemens Vision), the heads being secured by means of a foam rubber in order to minimize movement artifacts. The stimuli were presented visually every 15 sec. for a period of three seconds. The pauses between the stimuli were 12 sec. long. Subjects were producing the required item immediate after stimulus presentation. Every 60 sec. there was a paradigm change, initiated by an acoustic instruction. Each stimulus has been presented eight times. In four out of these eight presentations the 'prosodic' reaction was required. In the other four cases the subjects were rendering the item in a monotonous manner. Fig. 2 illustrates the data collection procedure used in the experiment. The material and the procedure have been validated in a pilot study [3] performed outside of the MR scanner.

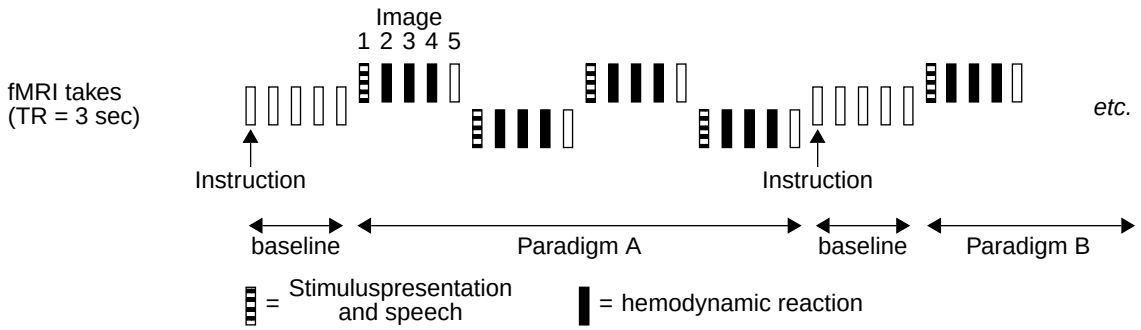

Figure 2. Schematic event related fMRI scanner protocol.

2.3. Analysis

2.3.1. Neuroimaging.

fMRI [functional Magnetic Resonance Imaging] technology uses the blood oxygen level-dependent (BOLD) contrast effect as an indirect marker of brain activation. Local neuronal activity gives rise to a decline in blood oxygenation which, in turn, causes an increase of blood flow. The hemodynamic response outweights the oxygen demand yielding accumulation of oxyhemoglobin within the respective region. Since magnetic properties of oxyhemoglobin are different from that of deoxyhemoglobin, imaging sequences allow to detect the change of MR signal within the activated areas. Twenty-eight parallel axial slices (thickness = 4 mm, gap = 1 mm) were acquired across complete brain volume by means of multi-slice echoplanar imaging sequence T2* EPI (TE=39ms, TR=3s, $\alpha=90^0$, FOV=192mm, 64^2 matrix).

2.3.2. Statistical analysis.

The assumption which is tested against in the cognitive brain research is that the brain is equipotential, with each behavior requiring the interaction of the entire structure. The established method is STATISTICAL PARAMETRIC MAPPING (SPM) [4]. The fMRI data from our experiment was processed by means of SPM96 software package. Each mean image was coregistered and movement correction and space normalization procedures have been performed. The normalized fMRI data were filtered (Gaussian filter, six millimeter full width half maximum [FWHM]). Since prior fMRI studies revealed a delay of the hemodynamic response extending from three to six seconds only the images within this time window (cf. the takes marked in bold in Fig. 2) were considered in the analysis [5]. For optimal localization of significantly activated areas SPM(t)-maps were superimposed on the structural MR images averaged across all nine subjects (Fig. 3).

3. RESULTS

The significant neural activity correlating with our experimental tasks is presented in the SPM images on the following page. For clarity we encircled the most relevant areas. We found a circuit of enhanced neural activity in both left and right anterior portions of the superior temporal gyrus and in the cerebellum (also left and right) in all tasks (rest was used as a baseline for subtraction). This cerebellar-temporo/frontal link may play a critical role in all tasks demanding rapid production of linguistic associations. The monotonous speech is characterized by increased bilateral neural activity in the primary motor cortex, and in the superior lateral hemisheres of the cerebellum (cf. Fig. 3). By subtracting task 4 (monotonous speech) from task 1 (i.e. simulation of FOCUS) we registered enhanced activity in the left temporal superior gyrus (area 38/L, cf. Fig. 3). Subtracting the baseline from task 2 (i.e. simulation of linguistic MODUS) revealed neural activity enhancement in the posterior part of the right superior temporal gyrus (area 22/R, cf. Fig. 3). Subtracting the baseline from task 3 (i.e. simulation of AFFECT) revealed neural activity in the anterior part of the right superior temporal gyrus (area 38/R, cf. Fig. 3).

4. DISCUSSION

The results support the view that both hemispheres subserve the processing of prosodic features of speech. They suggest that this processing is highly localized (superior temporal gyrus). Furthermore, the lateralization is not consistent with the distinction between linguistic vs. emotional functions of prosody. Rather, it is the case that prosodic features which require a short address frame (e.g. focused syllable) are lateralized differently than prosodic features requiring a long address frame (the whole intonational phrase for linguistic modus and paralinguistic affect). Prosodic frame length and not the linguistic/affective function is a basis of lateralization.

ACKNOWLEDGMENTS

This research is being supported by the grant nr. DO/536/2-1 from the German Science Foundation. Correspondence to: dogil, joemayer@ims.phonetik.uni-stuttgart.de.

REFERENCES

[1] Baum, S. and M. Pell, forthcoming. The neural bases of prosody. In M. Lynch (ed.) *The cognitive science of prosody: Interdisciplinary perspectives*.

[2] Levelt, W. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

[3] Mayer, J. 1999. Prosody processing in speech production: Preevaluation of a fMRI study. *Proceedings from ICPhS 99*, San Francisco.

[4] Frackowiak, R.S.J. (ed.) 1997. *Human Brain Function*. San Diego: Academic Press.

[5] Wildgruber, D., Erb, M., Klose, U. and Grodd, W. 1997. Sequential activation of supplementary motor area and primary motor cortex during self-paced finger movements in human evaluated by functional MRI. *Neuroscience Letters*, 127, 161-164.

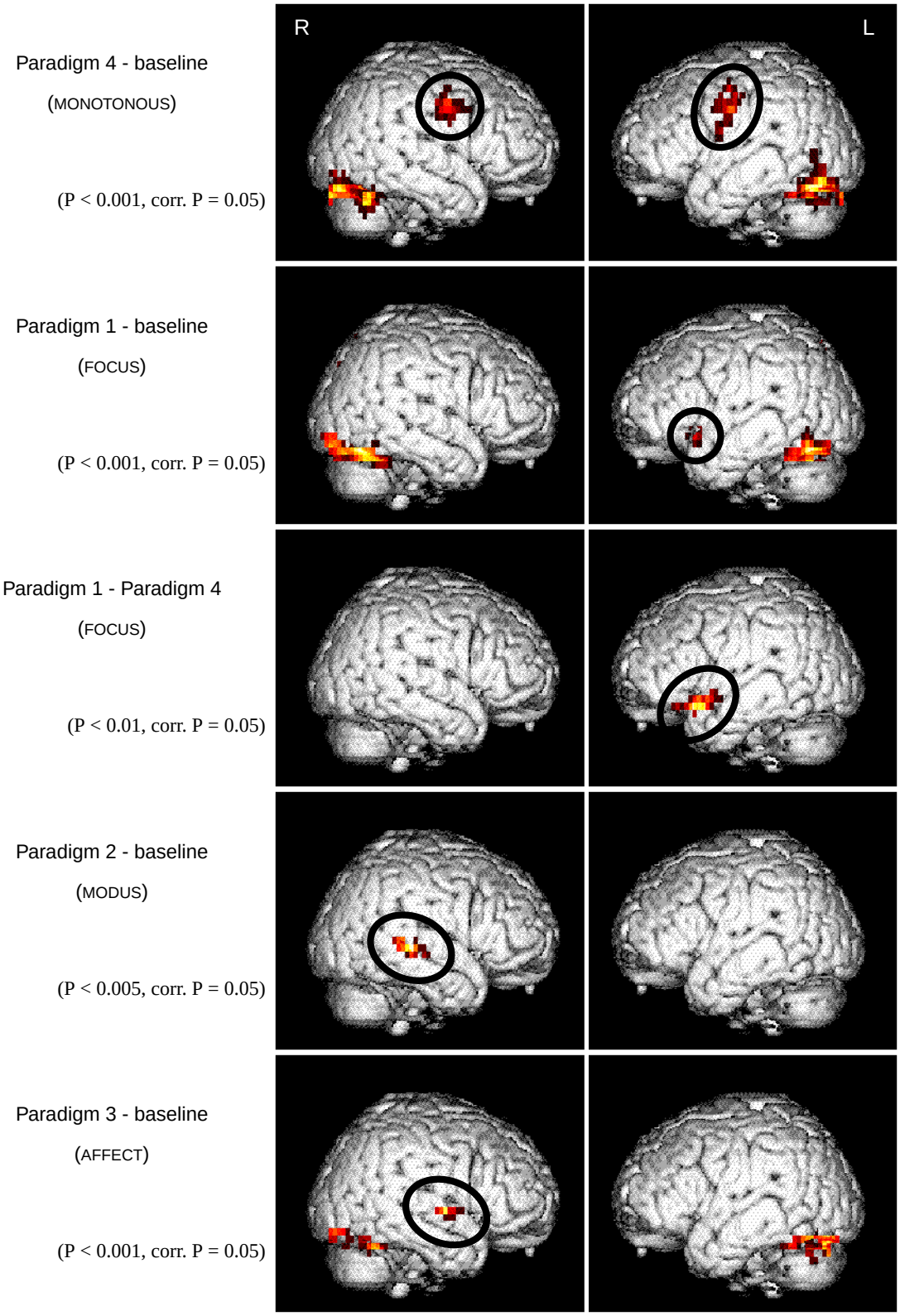

Figure 3. Statistical Parametric Maps (SPM96, Wellcome Department of Cognitive Neurology, London).