ELECTROPHYSIOLOGY OF EMOTIONAL
PROSODY PRODUCTION AND PERCEPTION

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (PhD) at the Center for Systems Neuroscience Hannover awarded by the University of Veterinary Medicine Hannover

BY

Milan Arsic

Medical Doctor
born in Nis, Serbia

Hannover, October 2008
Thesis Supervisor:
Prof. Dr. med. Reinhard Dengler

Referees:

1. Prof. Dr. med. Reinhard Dengler
2. Prof. Dr. med. Eckart Altenmüller
3. Prof. Dr. Elke Zimmermann

External referee:

Prof. Dr. Thomas Münte
Department of Neuropsychology
Otto-von-Guericke University Magdeburg
Postfach 4120
39106 Magdeburg
Germany

Date of the final exam: 31.10.2008.

Funding: The Deutsche Forschungsgemeinschaft (German Research Foundation)
Statement of research

I herewith declare that I autonomously carried out the PhD thesis entitled “Electrophysiology of emotional prosody production and perception.”

The following third party assistance has been enlisted: analysis of behavioral data in the first experiment by Dr. Janine Born, analysis of Event-Related Brain Potentials (ERPs) in the second experiment by Dr. Christine Schröder.

I did not receive any assistance in return for payment by consulting agencies or any other person. No one received any kind of payment for direct or indirect assistance in correlation to the content of the submitted thesis. I conducted the project at Hannover Medical School, Department of Neurology with clinical neurophysiology. The thesis has not been submitted elsewhere for an exam, as thesis or for evaluation in a similar context. I hereby affirm the above statements to be complete and true to the best of my knowledge.

Milan Arsic

Signature
Abstract

Electrophysiology of Emotional Prosody Production and Perception
Milan Arsic

Processing emotional prosody of a speech stream enables us to recognize the emotional state of the speaker. For instance, the same word can have a completely different meaning depending on whether it is intonated in a happy or angry tone of voice. Although emotional prosody processing fulfills such an important role in human social interactions, its underlying neural correlates are still not fully understood. This thesis describes two experiments performed in order to investigate emotional prosody production and perception using two different approaches.

Two different techniques offering the possibility to investigate emotional speech in real time based on objective measures were used, the repetitive Transcranial Magnetic Stimulation (rTMS) and the Event-Related Brain Potentials (ERPs). The rTMS is a non-invasive, painless method that enables temporary modulation of brain functions. Depending on the stimulation parameters, rTMS can activate or inhibit neuronal activity of the cortex. ERPs provide continuous acquisition and online measurements of electrical brain activity with a high time resolution in the range of milliseconds. Additionally, ERPs are considered as reliable tools for studying pre-attentional and attentional cognitive processes as well.

In the first experiment, rTMS was used to examine emotional prosody production. The rTMS was applied over the left and right dorsolateral prefrontal cortex (DLPFC) during
two separate sessions using a 10 Hz frequency at 100% of the resting motor threshold. Three series of stimulation were delivered with 10 minutes long breaks between the first and the second series. Additionally, sham stimulation was performed by positioning a figure-eight-shaped coil at the angle of 45° to the skull. After rTMS, 16 healthy subjects (8 female) pronounced a semantically neutral word “ANNA” in happy, neutral or sad emotional intonation, and answered a mood questionnaire. The goal was to test if stimulation over the left and right DLPFC can provoke transient mood changes, and further on, to determine their correlation with emotional speech. Fundamental frequency $F_0$ (its mean and standard deviation) was analyzed and compared between different stimulation conditions to test for the influence of rTMS on emotional speech production. Analysis of speech showed a statistically significant increase of the standard deviation of fundamental frequency after the right, but also after sham stimulation. When the mean fundamental frequency was used as a speech parameter, there were no significant differences observed after any type of stimulation and any of three intonations. A transient mood decrease occurred after the left stimulation and increased after the right one. The result of this experiment could not show that the rTMS delivered at these parameters was able to modulate emotional prosody production. However, rTMS was capable to transiently influence the mood of healthy subjects in a lateralized manner. This finding is similar to the effect observed in the previous studies done with healthy people and confirms the involvement of the DLPFC in the regulation of affect.

In the second experiment done in this dissertation, ERPs were used as a tool to explore how emotional prosody perception depends upon differences in valence and arousal. According to Russell (1980), each emotion can be described in a two-dimensional space defined in terms of valence (positive vs. negative) and arousal (low vs. high). Twenty healthy subjects (10 women) participated in the experiment. The dataset comprised different semantically neutral words spoken by two professional native speakers. Pronounced words differed with respect to emotional intonation (happy, relaxed, angry, and sad). The stimuli were presented in an oddball paradigm. In the passive condition, stimuli were delivered via headphones while the subjects were reading, where in the active one two experimental
conditions were considered namely, “valence” and “arousal.” In the former one, a series of frequent standard words spoken in negative prosody (e.g., sad) was violated by infrequent deviants of positive prosodic words (e.g., relaxed). In the latter one, arousal was different for standards and deviants (e.g., standard: sad; deviant: angry). The same prosodic combinations were also presented in the opposite manner, such that standard stimuli were used as deviant ones. The task was to evaluate deviant tone according to either its valence (positive, negative) or arousal (calm, aroused) by pressing one of the available keyboard buttons.

Results showed no significant differences for the passive condition. For the active condition, statistically significant difference was observed in the valence task and active condition. Thus, the mean amplitude of the $P3b$ component elicited by deviant stimuli was higher for high arousal stimuli happy and angry when compared to low arousal deviants relaxed and sad. Likewise, behavioral data showed that high arousal stimuli were rated more accurately. Most interestingly, this phenomenon occurred in the task where subjects had to focus on the valence of the target stimuli. These results conform to previous studies and show once again that the level of arousal of a stimulus has a high impact on emotional prosody processing, probably by allocating attentional resources so that it is processed as precisely as possible.
Zusammenfassung

Elektrophysiologie der Verarbeitung und Produktion emotionaler Prosodie
Milan Arsic


Hierbei wurden zwei verschiedene neurophysiologische Methoden verwand, zum einen die repetitive transkranielle Magnetstimulation (rTMS) sowie die der Ableitung ereigniskorr- relie rer Hirnpotentiale (ERPs). RTMS ist eine schmerzfreie, nicht invasive Untersuchungsmethode, mit der eine vorübergehende Veränderung der Hirnfunktion erreicht werden kann. Abhängig von den Stimulationsparametern führt sie zu einer Inhibition oder Aktivierung neuronaler Aktivität des Kortex. Hingegen ermöglicht die Ableitung von ERPs die Abbildung der Hirnaktivität mit einer hohen zeitlichen Auflösung (im Bereich von Millisekunden) und die Differenzierung pre-attentiver sowie attentiver kognitiver Prozesse.

Im ersten Experiment wurde rTMS zur Untersuchung der Produktion emotionaler Prosodie
ZUSAMMENFASSUNG


unterschieden sich in ihrer emotionalen Prosodie (fröhlich, erleichtert, ärgerlich, traurig).


Die Ergebnisse zeigen für die passive Aufgabe keine Unterschiede im Verlauf der ERPs. Hingegen findet sich ein signifikanter Unterschied in der Valenz- Aufgabe mit einer höheren mittleren Amplitude der P3b Komponente für diejenigen abweichenden Reize oder Deviants die sich durch ein hohes Arousal auszeichneten (fröhlich und ärgerlich). Entsprechend findet sich bei den Verhaltensdaten eine höhere Trefferquote für die Stimuli mit hohem Arousal. Interessanterweise zeigt sich dieser Effekt in der Valenz- Aufgabe, wo also abweichende Wort nach seiner Valenz als positiv oder negativ eingeschätzt werden sollte. Die Ergebnisse korrelieren mit vorangegenden Studien und zeigen, dass der Arousal- Grad eines Stimulus sehr wohl einen Einfluss auf die Verarbeitung emotionaler Prosodie nimmt, wahrscheinlich indem die Aufmerksamkeit insbesondere auf einen erregt klingenden Stimulus gelenkt wird und somit eine möglichst effiziente Verarbeitung gewährleistet ist.
Contents

1 Introduction 1
   1.1 What are emotions? ........................................... 1
   1.2 Communication through emotions .......................... 3
   1.3 Prosody and the brain ........................................ 5
   1.4 Prosodic deficits .............................................. 7
   1.5 Perception of emotional prosody ............................ 7
   1.6 Expression of emotional prosody ............................. 8
   1.7 Acoustic measures of emotional speech ...................... 10
   1.8 Electrophysiological and neuroimaging methods used in prosody research ............. 12
   1.9 Repetitive Transcranial Magnetic Stimulation (rTMS) .................. 13
      1.9.1 Principles of rTMS ........................................ 13
      1.9.2 rTMS in prosody research ................................. 15
   1.10 Event-Related Brain Potentials (ERPs) ...................... 17
      1.10.1 Principles of ERPs ........................................ 17
      1.10.2 ERPs in prosody research ................................. 19
   1.11 Thesis overview ............................................. 20
      1.11.1 Objectives ............................................ 21

2 Methods and materials 22
   2.1 The rTMS Experiment .......................................... 22
      2.1.1 Subjects ............................................. 22
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.2</td>
<td>The rTMS procedure</td>
<td>23</td>
</tr>
<tr>
<td>2.1.3</td>
<td>Stimulation protocol</td>
<td>25</td>
</tr>
<tr>
<td>2.1.4</td>
<td>Speech tasks</td>
<td>26</td>
</tr>
<tr>
<td>2.1.5</td>
<td>Speech recordings</td>
<td>27</td>
</tr>
<tr>
<td>2.1.6</td>
<td>Self assessment questionnaires</td>
<td>27</td>
</tr>
<tr>
<td>2.1.7</td>
<td>The sound analysis</td>
<td>28</td>
</tr>
<tr>
<td>2.1.8</td>
<td>Statistical analysis</td>
<td>28</td>
</tr>
<tr>
<td>2.2</td>
<td>The ERP experiment</td>
<td>29</td>
</tr>
<tr>
<td>2.2.1</td>
<td>Subjects</td>
<td>29</td>
</tr>
<tr>
<td>2.2.2</td>
<td>Stimulus material</td>
<td>29</td>
</tr>
<tr>
<td>2.2.3</td>
<td>An Oddball paradigm</td>
<td>30</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Behavioral data recordings</td>
<td>32</td>
</tr>
<tr>
<td>2.2.5</td>
<td>Electrophysiological recordings</td>
<td>32</td>
</tr>
<tr>
<td>2.2.6</td>
<td>Chamber setup</td>
<td>32</td>
</tr>
<tr>
<td>2.2.7</td>
<td>Data analysis</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>Results</td>
<td>35</td>
</tr>
<tr>
<td>3.1</td>
<td>The rTMS experiment</td>
<td>35</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Speech Task</td>
<td>35</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Behavioral data</td>
<td>37</td>
</tr>
<tr>
<td>3.2</td>
<td>The ERP experiment</td>
<td>38</td>
</tr>
<tr>
<td>3.2.1</td>
<td>ERP data</td>
<td>38</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Active condition</td>
<td>39</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Behavioral results</td>
<td>43</td>
</tr>
<tr>
<td>4</td>
<td>Discussion</td>
<td>47</td>
</tr>
<tr>
<td>4.1</td>
<td>The rTMS experiment</td>
<td>47</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Summary</td>
<td>51</td>
</tr>
<tr>
<td>4.2</td>
<td>The ERP experiment</td>
<td>51</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Passive condition: pre-attentive stimulus processing</td>
<td>51</td>
</tr>
</tbody>
</table>
4.2.2 Active condition: attentive stimulus processing .......................... 54

Bibliography 57
List of Figures

1.1 Russell’s “Circumplex Model of Affect.” ............................................. 3
1.2 Bühler’s Organon Model. Figure adopted from Scherer and Bänziger (2004). 5
1.3 (a) Three-stage working model for the processing of emotional prosody. Sensory processing (Stage 1): Acoustic analysis is mediated by bilateral auditory processing areas. Integration (Stage 2): Processing along the auditory “what” pathway integrates emotionally significant acoustic information to derive an emotional “gestalt.” This pathway projects from the superior temporal gyrus (STG) to the anterior superior temporal sulcus (STS) and might be lateralized to the right hemisphere (RH). Cognition (Stage 3): Emotional information derived at the level of the STS is made available for higher-order cognitive processes. For example, explicit evaluative judgments of emotional prosody are mediated by the right inferior gyrus (IFG) and orbitofrontal cortex (OFC), whereas the integration of emotional prosody into language processing recruits inferior frontal gyrus in the left hemisphere (LH). Contextual or individual significance might facilitate or enhance processing at any of the three stages. (b) Schematic presentation of brain areas implicated in vocal emotional processing in a right sagittal view: primary, secondary, and tertiary auditory cortex (light blue) extending to the anterior portion of the superior temporal sulcus (dark blue), from where projections reach inferior frontal gyrus and orbitofrontal gyrus (green). Arrows (yellow) indicate presumed processing directions (colors/numbers correspond to the processing stages outlined in (a). Figure adopted from Schirmer and Kotz (2006). . . . 9
1.4 The spatial and temporal resolution of electrophysiological and neuroimaging techniques. Depending on the chosen technique, one selects the question that he/she can ask. Abbreviations used: CT Computerized Tomography, EEG electroencephalography, ERPs Event Related Brain Potentials, MEG magnetoencephalography, fMRI functional Magnetic Resonance Imaging; PET Positron Emission Tomography. From Walsh and Cowey (1999).

1.5 Examples of cortical activations as a response to high frequency (3.125 HZ) supratreshold stimulation over the motor cortex. The maps represent (A) dorsal, (B and C) lateral, and (D and E) medial views of significant activations projected onto a 3D surface reconstruction of a template brain (Montreal Neurological Institute, MNI). Increased activity (red-yellow) was found in left M1/S1, PMd, bilateral SMA and auditory cortices, postcentral sulcus, and left ventral posterior middle temporal gyrus. In the left hemisphere, additional activity was observed along the inferior precentral sulcus. Decreased BOLD MRI signal intensities (blue) were found in the right M1/S1 region.

1.6 (A) A scheme of pyramidal cell during neurotransmission. An excitatory neurotransmitter is released from the presynaptic terminals, causing positive ions to flow into the postsynaptic neuron. This creates a negative extracellular voltage (represented by the “−” symbols) yielding a small dipole. (B) Folded sheet of cortex containing many pyramidal cells. When a region of this sheet is stimulated, the dipoles from the individual neurons summate. (C) The summed dipoles from the individual neurons can be approximated by a single equivalent current dipole (shown as an arrow). (D) Example of a current dipole with a magnetic field travelling around it. (E) Example of the magnetic field generated by a dipole that lies just inside the surface of the skull. Adopted from Luck (2005).
1.7 Idealized waveform of the computer-averaged auditory event-related potential (ERP) to the brief sound. The ERP is generally too small to be detected in the ongoing EEG (top) and requires computer averaging over many stimuli presentations to achieve adequate signal-to-noise ratios. The logarithmic time display allows visualization of the waves (I-VI), the mid-latency components (No, Po, Na, Pa, Nb), the vertex potential waves (Pl, Nl, P2), and task-related endogenous components (Nd, N2, P3 and slow wave). Figure adopted from Rugg and Coles (1995).

2.1 EEG 10 – 20 system side view.

2.2 EEG 10 – 20 system top view.

2.3 Active stimulation delivered over the left DLPFC.

2.4 Sham stimulation delivered with coil positioned 45° off the skull.

2.5 Schematic presentation of experimental design and the rTMS pattern in one of the three stimulation series.

2.6 Computerized paradigm for speech induction. First, visual cue was presented, suggesting to the participants what type of intonation to produce. After 1.5 sec, it was followed by another visual sign (pictogram of a mouth), telling the participants to start by pronouncing the word “Anna.”

2.7 Schematic representation of the oddball paradigm shown for the valence task, (angry: standard/ happy: deviant) in the upper row, and (sad: standard/ relaxed: deviant) in the lower row. To avoid influence of physical differences on ERP waveforms, all stimuli were presented both as standards and deviants in different runs.

2.8 Schematic representation of the oddball paradigm shown for the arousal task (happy: standard/ relaxed: deviant) in the upper row, and (sad: standard/ angry: deviant) in the lower row. To avoid the influence of physical differences on ERP waveforms, all stimuli were presented both as standards and deviants in different runs.
2.9 Recording setup showing placement of the electrodes on Easycap and those used for eye movement monitoring. ........................................ 33
2.10 Chamber setup for performing the ERP experiment. .................. 34

3.1 The ERPs for sad intonations (presented as standards) and relaxed voice intonation (presented as deviants) in the valence task and passive condition. ...................................................... 38
3.2 The ERPs for angry prosodic intonation (presented as standards) and happy prosodic intonations (presented as deviants) in the valence task and passive condition. ................................................. 39
3.3 The ERPs for angry prosodic intonations (presented as standards) and sad prosodic intonations (presented as deviants) in the arousal task and passive condition. ....................................................... 40
3.4 The ERPs for happy prosodic intonations (presented as standards) and relaxed prosodic intonations of voice (presented as deviants) in the arousal task and passive condition. .............................................. 41
3.5 The grand average waveforms elicited by all four prosodic intonations (happy, angry, relaxed and sad), serving as standards and deviants, in all runs in passive condition. ......................................................... 42
3.6 The ERPs elicited by the deviant stimuli in all prosodic intonations (happy, angry, relaxed, and sad) in the passive condition. ......................... 43
3.7 The ERPs for attended stimuli in the active valence task for all prosodic intonations of the voice (happy, angry, relaxed, and sad). ...................... 44
3.8 The ERPs for attended stimuli in active arousal task for all prosodic intonations (happy, angry, relaxed, and sad). ........................................ 45
3.9 The ERPs for attended stimuli calculated for both tasks, for all prosodic intonations of the voice (happy, angry, relaxed, and sad). .................. 46
## List of Tables

1.1 Acoustic speech parameters. ............................................. 11

2.1 Stimuli parameters. ....................................................... 30

3.1 Table shows mean values of the mean fundamental frequencies $F_0$ (Hz) for the different intonations (neutral, happy, and sad) and rTMS conditions (pre-rTMS, left, right, and sham). The asterisk symbol * ($p < 0.05$) indicates a significant interaction, while the standard error of the mean (SEM) is shown in italics. ............................................. 36

3.2 Table shows mean values for standard deviation of fundamental frequencies $F_0$ (Hz) for the different intonations (neutral, happy, and sad) and rTMS conditions (pre-rTMS, left, right, and sham). The asterisk symbol * ($p < 0.05$) indicates significant interaction, while the standard error of the mean (SEM) is shown in italics. ............................................. 36

3.3 Table shows values for 6 scales and 14 subscales of EWL-k questionnaire and changes after different rTMS conditions (pre-rTMS, left, right, and sham). The asterisk symbol * ($p < 0.05$) indicates significant interaction. .................. 37

3.4 Table shows mean values (presented in percentage) of correct responses to the deviant stimuli for the different prosodic intonations (angry, happy, relaxed, and sad) and different tasks (valence, arousal, as well as valence and arousal calculated together). The standard deviation (SD) is indicated in italics. .. 44
Chapter 1

Introduction

1.1 What are emotions?

Charles Darwin’s work “The expression of the emotions in man and animals” (Darwin, 1872) established the basis for emotional expressiveness research. Darwin was the first to describe the relation between the expression of an emotion and the internal state of its sender. He asserted that human emotions fulfill adaptive functions, such as organizing the body’s response to various challenges in the surrounding environment and in interactions with others.

Besides Darwin, William James tried to answer the question “What is an emotion?” (James, 1884). He proposed that stimuli that provoke emotions first induce changes in the viscera and the autonomic nervous system and that perception of these signals subsequently produces emotional experience. This theory is considered as one of the first about emotions and is known as the James-Lange theory to recognize the contribution of another scholar, Carl Lange (Lange, 1885). Later on, Walter Cannon (Cannon, 1927) together with Philip Bard (Bard, 1928) performed experiments on animals that were able to display emotional expressions, although their internal organs were separated from the central nervous system. They opposed the point of view of the James-Lange theory and concluded that the viscera has insufficient output to the brain to be important in inducing emotional experience. Cannon noted that emotions have primarily adaptive functions and are one of the strongest motivational forces of human behavior. Their role is to ensure survival of the organism,
1.1. What are emotions?

society and family. Ekman, Friesen, and Ellsworth (1972) named anger, disgust, fear, joy, sadness, and surprise as the six basic emotions, because they appear among all cultures and are universally associated with and recognizable by facial expression characteristics. They also appear to serve identifiable biological functions related to the survival needs of the individuals and the whole species. Nevertheless, other researchers (Tomkins, 1962; Plutchik, 1980; Panksepp, 1982) confided more different emotions, such as expectancy and shame to be the basic ones, while other scholars tried to adopt a multidimensional approach to emotions (Wundt, 1911; Russell, 1980).

The German psychologist, Wilhelm Wundt (Wundt, 1911), was the first one who tried to classify emotions along three dimensions: “pleasure” or “valence”, “arousal” and “dominance.” A two-dimensional view of emotions has been proposed by Russell (1980). Russell’s circumplex model of affect has two primary dimensions, “valence” (as a continuum from positive to negative) and “arousal” (ranging from calmness to excitement) (see Figure 1.1). The multidimensional approach offers a conceptual and experimental framework for exploring the neural basis of emotions, and is used in this thesis. In general, it also provides a good theoretical basis for understanding the widespread comorbidity among mood disorders and anxiety (Posner et al., 2005).

The third approach has been proposed by Scherer (1984), who argued that emotion elicitation and differentiation is understood as a process of event evaluation or appraisal. Appraisal models the way in which an individual assesses the personal significance of an event for its well-being based on a number of criteria and dimensions. The result of this appraisal process is an emotion, which is then expressed or externalized in psychological symptoms and particularly in the motor expressive movements of the face, body, and voice (Scherer, 2004). He also stated, that, in different theories the term emotion is used to describe different affective states. Hence, Scherer (2000) tried to sort classes of affective states into distinct categories:

- **Emotions** (e.g., angry, sad, joyful, fearful, proud, elated, desperate)

- **Moods** (e.g., cheerful, gloomy, irritable, listless, depressed)
1.2 Communication through emotions

Following Darwin's pioneering work, most of the research on emotional communication has been performed on facial expressions. Darwin proposed the view that all emotions are innate, and there should be no, or only small, differences in emotional expression between different cultures. The theory of discrete emotions is one of the most prominent in this field of research and is based on studying emotional facial expressions (Tomkins, 1962; Panksepp, 1998). Izard (1998) and Ekman et al. (1972) performed cross-cultural studies and provided

Figure 1.1: Russell's “Circumplex Model of Affect.”

- Interpersonal stances (e.g., distant, cold, warm, supportive, contemptuous)
- Preferences/Attitudes (e.g., liking, loving, hating, valuing, desiring)
- Affect dispositions (e.g., nervous, anxious, reckless, morose, hostile)
evidence for the assumption that seven to nine emotions are innate. When compared to facial emotional expressions, research on emotional vocal expression has received much less attention. However, the idea that vocal cues are connected with the emotional state of the speaker has been noticed very early. Both Cicero and Aristotle proposed that each emotion is associated with a distinctive tone of voice. Like the face, the voice can reliably convey some discrete emotions, such as anger and sadness (Juslin and Laukka, 2003) with a great degree of cross-cultural similarity (Scherer et al., 2001).

Vocal emotional communication serves a highly important function in social interaction; it allows individuals to communicate information to others. In humans, vocal affect expression is based on phylogenetically older parts of the brain similar to those which are morphologically similar to the same structure in the brain of non-human primates. However, what is regarded to be special in humans is the ability of a much greater voluntary control of vocalization (Jürgens, 2002). Such advantage in the control of vocalization is very effectively used in social communication. Hence, Scherer (1989) introduced the “push” and “pull” effects of emotions.

“Pull effects” include physiological responses of the body. For example, elevated arousal of the sympathetic nervous system may influence the breathing pattern, change the shape of the vocal tract and thus, the emotional vocal expression. “Push effects,” on the other hand, refer to a strategic production of emotional expressions for communicative purposes (Krebs and Dawkins, 1984). Russell and colleagues (Russell et al., 2003) similarly agree that emotional expressions are not sent to any and all, but are rather directed to the receiver in a way that is beneficial for the sender. For example, if the sender is on one side angry with the receiver, but on the other side needs him to achieve the aim, he would rather speak in a polite way, instead of shouting at him. Similarly, the receiver does not only receive (encode) the emotional cues but also distinguishes truthful from misleading information. Since affect expressions reflect “push” and “pull” phenomena, spontaneous and strategic effects may be hard to differentiate in real life situations. The function of vocal emotional expression is effectively explained by Bühler’s Organon Model (Bühler, 1934) displayed in Figure 1.2. According to this model, emotional expression has the following characteristics:
1.3. Prosody and the brain

The word *prosody* comes from the Greek word *prosoidia*, meaning “song sung with accompaniment.” Prosodic phenomena are used to structure the speech flow and are perceived as stress and accentuation, or as other modifications of intonation, rhythm and loudness (Werner and Keller, 1994). One of the first to study speech prosody in brain-damaged patients was Monrad-Krohn (1947). He investigated the case of Astrid L., a Norwegian woman wounded in the Second World War. She sustained a shrapnel injury of the fronto-temporo-parietal region of the left hemisphere that subsequently caused Broca’s aphasia. Although the wound recovery process was successful, the alterations in the normal accent caused problems in her verbal communication. During the German occupation of Norway,
listeners always perceived her as a German speaker and, hence, she was socially isolated. Monrad-Krohn (1947) observed: “She never had the natural Norwegian accent when she had to link several words into sentence. What above all characterized her speech was her broken foreign accent, her completely changed melody of language.” However, he could not explain the fact that, besides the acquired foreign accent, this woman preserved her ability to sing. The foreign accent phenomena is usually the result of brain damage, and the patients tend to have symptoms associated with Broca’s aphasia. In all reported cases of the foreign accent phenomena, abnormal prosodic features, including alterations of stress, rhythm, and intonation, are mentioned. In the case of Astrid L., one of the features which presumably contributed to the impression of a foreign accent in her speech was a failure to produce the Norwegian pitch accents. This accent distinction is phonemic and is used to distinguish between pairs of words which consist of the same segmental phonological string (Moen, 1996). Upon his research, Monrad-Krohn divided prosody into four types:

1. **Intrinsic (linguistic) prosody** defines and clarifies the meaning of the sentence using appropriate intonation and pauses, which are equivalent to interpunction signs in the written language.

2. **Intellectual prosody** defines attitudinal information and changes the meaning of the spoken message. The very effective example can be intonation of the sentence “You are smart” and the difference in meaning that can be produced when different parts of the sentence are stressed. If the sentence is stressed on the word “are,” it expresses appreciation of someone’s intellectual abilities; however, if the sentence is stressed on the word “smart,” with raised intonation at the end of the sentence, the expected lexical meaning changes and the sarcasm becomes obvious.

3. **Inarticulate prosody** represents the use of paralinguistic elements of the speech (different from emotional prosody) such as sighs and grunts.

4. **Emotional or affective prosody**. As stated by Ross (2000): “Emotional prosody inserts moods and emotions such as happiness, sadness, fear and anger, into speech. When coupled with gestures, affective prosody greatly influences the content and the
impact of the message. The paralinguistic features of the language, as examples by affective prosody, may thus play an even more important role in the human communication than the exact choice of the words.”

1.4 Prosodic deficits

Besides working on prosody categorization, Monrad-Krohn also described different disorders of prosody expression in various neurological disorders (Monrad-Krohn, 1947):

- **Dysprosody** is defined as a change of voice quality following left hemisphere damage, with difficulties in pronunciation and disturbed patterns of stress and intonation, while affective prosody is preserved.

- **Aprosody** is defined as a disturbance in the modulation of speech intonation, observed, for example, in Parkinson’s disease.

- **Hyperprosody** is defined as the excessive use of prosody as observed in manic patients. When having very few words at their disposal, they used them excessively to overcome their deficit.

Monrad-Krohn also predicted that neurological patients could suffer from deficits in prosody perception, but did not classify this condition as he did for prosody production.

1.5 Perception of emotional prosody

To date, one of the most debated questions in the investigation of prosody perception is the identification of brain areas that control this process. Ross (1981) has examined patients with focal brain damage and proposed the first theory regarding this problem. He claimed that the right hemisphere (RH) is in absolute control of emotional prosody perception. Lancker and Sidtis (1992) proposed a model that distinguishes between processing of linguistic versus emotional prosody. In a study with left and right brain-damaged patients, they showed task dependent shifts from the right hemisphere (RH) (involved more in emotionally-based tasks) to the left hemisphere (LH) (involved more in linguistically-based
1.6 Expression of emotional prosody

Besides the emotional prosody perception, Ross (1981) also investigated the spontaneous prosody production and prosodic repetition. He concluded that the prosodic function of the right hemisphere (RH) mirrors the left hemisphere (LH) function for speech production of right-handed individuals. According to this global view, aprosodias can be classified in a similar manner as aphasias, and are lateralized to the right hemisphere (RH). However,
1.6. Expression of emotional prosody

Figure 1.3: (a) Three-stage working model for the processing of emotional prosody. Sensory processing (Stage 1): Acoustic analysis is mediated by bilateral auditory processing areas. Integration (Stage 2): Processing along the auditory “what” pathway integrates emotionally significant acoustic information to derive an emotional “gestalt.” This pathway projects from the superior temporal gyrus (STG) to the anterior superior temporal sulcus (STS) and might be lateralized to the right hemisphere (RH). Cognition (Stage 3): Emotional information derived at the level of the STS is made available for higher-order cognitive processes. For example, explicit evaluative judgments of emotional prosody are mediated by the right inferior gyrus (IFG) and orbitofrontal cortex (OFC), whereas the integration of emotional prosody into language processing recruits inferior frontal gyrus in the left hemisphere (LH). Contextual or individual significance might facilitate or enhance processing at any of the three stages. (b) Schematic presentation of brain areas implicated in vocal emotional processing in a right sagittal view: primary, secondary, and tertiary auditory cortex (light blue) extending to the anterior portion of the superior temporal sulcus (dark blue), from where projections reach inferior frontal gyrus and orbitofrontal gyrus (green). Arrows (yellow) indicate presumed processing directions (colors/numbers correspond to the processing stages outlined in (a). Figure adopted from Schirmer and Kotz (2006).

Deficits to modulate emotional speech have been observed in patients suffering from damage to the right and left brain hemisphere (Pell and Baum, 1997; Behrens, 1989; Shapiro and Danly, 1985). Schirmer et al. (2001) argue that there is a particular contribution of left hemisphere to prosodic production and superiority for the processing of linguistic prosody. They also propose right hemisphere specialization for choosing right affective cues i.e. the appropriate $F_0$ pattern, lending support for the functional-lateralization hypothesis. The aforementioned studies investigated emotional prosody production in brain-damaged patients, whereas only a few studies tried to examine emotional prosody production with help of neuroimaging methods. Reason is that the study of speech production with an Magnetic Resonance Imaging (MRI) or Positron Emission Tomography (PET) scanner has been
considered technically difficult. Speaking involves the jaw, tongue, lip and larynx movements which inevitably lead to changes of the volumes of the cavities close to the brain. Unfortunately, the brain-imaging techniques are susceptible to motion and volume-change artefacts. The technical developments in fMRI technique-stronger magnetic fields and better interpretation of temporal properties of Blood-Oxygen-Level Dependent (BOLD) signal have led to the establishment of an experimental paradigm which allows for studies of overt speech (Dogil et al., 2002). Evidence from functional imaging studies also points to the fact that both hemispheres are involved in emotional speech production (Mayer et al., 1999, 2002). Mayer et al. (1999) argue that prosodic frame length and not the linguistic/affective functions, is a basis of lateralization. Hence, Mayer et al. (2002) give further support for the functional lateralization hypothesis. However, to achieve speech coordination, intrahemispheric communication has to ensure that the articulatory-verbal and affective prosodic elements are behaviorally unified and temporally coherent (Klouda et al., 1988; Ross et al., 1997). It is possible that intrahemispheric lesions lead to a deficit in affective prosody production by interrupting the integration of affective and propositional aspects of speech.

1.7 Acoustic measures of emotional speech

Quantification of the acoustic speech parameters is considered to be the most suitable and accurate method to describe changes in emotional speech. It allows an objective, unbiased analysis of speech samples obtained during emotional prosody production. According to Banse and Scherer (1996), there are several acoustic variables that constitute emotional prosody:

- the level, the range and the contour of fundamental frequency $F_0$ (perceived as pitch)
- the vocal energy i.e., the amplitude (perceived as vocal intensity)
- the distribution of the energy in the spectrum
- the location of formants $F1$ and $F2$ (related to perception of articulation)
- the variety of temporal phenomena.
Table 1.1 summarizes major acoustic speech parameters used in emotional speech processing.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_0$</td>
<td>Fundamental frequency</td>
</tr>
<tr>
<td>$F_1$</td>
<td>the first formant</td>
</tr>
<tr>
<td>$F_2$</td>
<td>the second</td>
</tr>
<tr>
<td>$F_0$ perturbation</td>
<td>Slight variations in the duration of glottal cycles</td>
</tr>
<tr>
<td>$F_0$ mean</td>
<td>Fundamental frequency (vibration rate of vocal folds as averaged over a speech utterance)</td>
</tr>
<tr>
<td>$F_0$ range</td>
<td>Difference between highest and lowest $F_0$ in an utterance</td>
</tr>
<tr>
<td>$F_0$ variability</td>
<td>Measure of dispersion (e.g., standard deviation of $F_0$)</td>
</tr>
<tr>
<td>$F_0$ contour</td>
<td>Fundamental frequency values plotted over time (intonation)</td>
</tr>
<tr>
<td>$F_1$ mean</td>
<td>Frequency of the first (lowest) formant (significant energy concentration in the spectrum) averaged over an utterance</td>
</tr>
<tr>
<td>$F_2$ mean</td>
<td>Mean frequency of the second formant</td>
</tr>
<tr>
<td>Formant bandwidth</td>
<td>Width of the spectral band containing significant formant energy</td>
</tr>
<tr>
<td>Formant precision</td>
<td>Degree to which formant frequencies attain values prescribed by phonological system of a language</td>
</tr>
<tr>
<td>Intensity mean</td>
<td>Energy values for a speech sound wave averaged over an utterance</td>
</tr>
<tr>
<td>Intensity range</td>
<td>Difference between highest and lowest intensity values in an utterance</td>
</tr>
<tr>
<td>Intensity variability</td>
<td>Measure of dispersion of intensity values in an utterance (e.g., standard deviation)</td>
</tr>
<tr>
<td>Frequency range</td>
<td>Difference between $F_0$ and highest point in the frequency spectrum where there is still speech energy</td>
</tr>
<tr>
<td>High-frequency energy</td>
<td>Relative proportion of energy in the upper frequency region (e.g., $&gt; 1$ kHz)</td>
</tr>
<tr>
<td>Spectral noise</td>
<td>Aperiodic energy components in the spectrum</td>
</tr>
<tr>
<td>Speech rate</td>
<td>Number of speech segments per time unit</td>
</tr>
</tbody>
</table>

Table 1.1: Acoustic speech parameters.

There is a considerable agreement on the acoustic cues that differentiate between certain emotions. Hence, *anger* as a high arousal emotion is characterized by increased mean $F_0$, $F_0$ range and high frequency energy. On the contrary, expression of *sadness* is described by the lowest mean $F_0$, $F_0$ range and mean energy values. *Joy* is an emotion of positive valence and high arousal going along with an increased mean $F_0$, $F_0$ range, $F_0$ variability and mean energy.

Possible approaches to investigate emotional prosody production and perception can be summarized as follows:

1. Recordings and acoustical analysis of natural speech samples during different, highly emotional situations such as journalists reporting from high-risk locations, talk shows or sports events on the TV (Johannes et al., 2000; Cowie and Douglas-Cowie, 1996; Duncan et al., 1983).

2. Induction of emotions in an experimental setting, for example by presenting a funny video, followed by a recording and an acoustical analysis of speech samples.

3. Using simulated (portrayed) vocal expressions where experiment participants or professional actors are asked to produce a certain emotional prosody (as if they were
happy or sad), followed by a recording and analysis of obtained speech samples (Banse and Scherer, 1996; Scherer, 1986).

In this thesis, the third method was used in the investigation of emotional prosody production and perception. In the first experiment, participants were asked to produce a certain emotional intonation upon the visual pictograms displayed on the computer screen. Such a method was previously described and used by Scherer (2003) as well as Hammerschmidt and Jürgens (2007). In the second experiment, two professional actors were asked to simulate prosodic intonations of different arousal and valence. The obtained speech samples were further presented to the group of healthy participants in order to investigate the emotional prosody perception.

1.8 Electrophysiological and neuroimaging methods used in prosody research

Until the last decade of the Twentieth Century, data about brain regions controlling prosody production and perception was collected and conclusions drawn mostly from studies performed with neurologically impaired populations. The studies included the patients with unilateral lesions of the left (LH) or the right hemisphere (RH) or the basal ganglia (BG). Studies with healthy subjects were mainly performed using dichotic listening procedures or the Wada technique (named after neurologist, Jun A. Wada).

Within the past two decades, empirical evidence has also begun to appear from studies using different electrophysiological and neuroimaging methods. The Event-Related Brain Potentials (ERPs) were introduced earlier, at the end of 1960s, but employed later in speech and especially in prosody research. Other methods encompass electrophysiological procedures such as repetitive Transcranial Magnetic Stimulation (rTMS) and neuroimaging techniques such as functional Magnetic Resonance Imaging (fMRI), magnetoencephalography (MEG) and Positron Emission Tomography (PET). Nowadays, these techniques, having different temporal and spatial resolution, allow researchers different approaches and more precise identification of neural networks controlling emotional prosody production and perception (Figure 1.4). The following sections detail electrophysiological techniques used
throughout this thesis and describe their application in prosody research.

Figure 1.4: The spatial and temporal resolution of electrophysiological and neuroimaging techniques. Depending on the chosen technique, one selects the question that he/she can ask. Abbreviations used: CT Computerized Tomography, EEG electroencephalography, ERPs Event Related Brain Potentials, MEG magnetoencephalography, fMRI functional Magnetic Resonance Imaging; PET Positron Emission Tomography. From Walsh and Cowey (1999).

1.9 Repetitive Transcranial Magnetic Stimulation (rTMS)

1.9.1 Principles of rTMS

Anthony Barker, from the University of Sheffield, was the first to build and use the transcranial magnetic stimulation (TMS) device (Barker et al., 1985). According to Barker’s research, it is possible to stimulate the nerves and the brain by applying a strong magnetic field over the skull. The stimulation is based on the principles of Faraday’s Law of Induction (or Law of Electromagnetic Induction) that states: “The electromagnetic field (emf) that is induced in a circuit is directly proportional to the time rate of change of magnetic flux
through the circuit” and is defined as follows:

\[ Emf = -N \times \frac{d\Phi}{dt} \]  \hspace{1cm} (1.1)

where \( Emf \) denotes electromotive force expressed in volts (V), \( N \) is the number of turns in a wire and \( \Phi \) stands for the magnetic flux expressed in weber (Wb).

The rTMS is delivered by a figure-of-eight stimulation coil, also known as a double cone or butterfly coil. This type of coil consists of two planar wings with the core made of two threads of wire, where the maximal current is produced in the middle, at the intersection of the two wings. The short but strong electrical current of approximately 3000 A is generated by the capacitors in the rTMS device and then passed through the coil. Further on, the coil current \( I \) produces a time-varying magnetic field \( B \) with a field strength up to 2T. This rapidly-changing magnetic field \( B \) generates an electric field \( E \) that influences the neural activity and the resting potential of the neuronal cells (Walsh and Cowey, 1999). The field’s strength reaches its maximum on the coil surface, while on the cortex, its strength directly depends on coil-cortex distance (Thielscher and Kammer, 2004).

It is assumed that rTMS, applied over the motor cortex, influences the cortical neurons trans-synaptically (Amassian et al., 1998). So far, it is not known how rTMS influences neurons of the other cortical regions besides the motor cortex. However, it is generally accepted that the high frequency stimulation with a repetition rate above 1 Hz increases neuronal excitability, and, that low frequency stimulation with a repetition rate below and equal to 1 Hz leads to the suppression of neuronal activity (Wassermann, 1998). Moreover, by combining rTMS with neuroimaging techniques such as the fMRI or PET, and using behavioral tests, one can indirectly conclude how rTMS influences certain brain functions. Figure 1.5 shows examples of cortical activations as a response to high frequency rTMS.

Although rTMS is generally considered to be a safe method, it was found to produce epileptic seizures even in healthy subjects. This was noticed to happen especially if the stimulation intensity was high, i.e., close to the resting motor threshold (Pascual-Leone et al., 1993; Wassermann, 1998) and high frequencies (more that 15 Hz) when short inter-train intervals were used (Chen et al., 1997). So far, eight accidental seizures have been
reported. In four cases, the seizures occurred after rTMS application over the primary motor cortex; in three cases, they happened after stimulation over the prefrontal cortex; and in one case, after the stimulation over the left primary auditory cortex. According to the safety guidelines described by Wassermann (1998), possible contraindications of magnetic stimulation are:

1. **Absolute contraindications** - metal in cranium, intracardiac lines and increased intracranial pressure.

2. **Relative contraindications** - pregnancy, childhood, heart disease, cardiac pacemaker, implanted medication pump, tricyclic antidepressants, neuroleptics and family history of epilepsy.

Besides these serious adverse effects, discomfort due to irritations of the scalp and facial nerves provoked by stimulation, as well as mild headaches were also reported among participants.

During the past two decades, rTMS has been used to investigate and map healthy brain functions involved in vision, memory, and muscle control (Pascual-Leone et al., 1996; Paus et al., 1997; Fitzgerald et al., 2006). Moreover, numerous studies were conducted to explore the potential of rTMS as a therapeutic tool for different neurological and psychiatric diseases. In the vast number of studies, magnetic stimulation was examined as a possible tool to treat major depressive disorder (George et al., 1995), schizophrenia (Nahas et al., 1999), obsessive-compulsive disorder (OCD) (Sachdev et al., 2001), post-traumatic stress disorder (PTSD), bipolar disorder (Belmaker, 1999), dystonia (Siebner et al., 1999), tinnitus (Plewnia et al., 2000) and Parkinson’s Disease (PD) (Shimamoto et al., 2001).

### 1.9.2 rTMS in prosody research

The rTMS has previously been used in the research of speech production and related brain functions. Nevertheless, the most recognized modulation of brain function by rTMS is the induction of a speech arrest (Epstein et al., 1999; Pascual-Leone et al., 1991). When rTMS is delivered over Broca’s area, it is able to temporarily interrupt fluent speech, thus causing a “virtual lesion.” Up to now, rTMS was successfully used in this manner to temporarily
modulate emotional prosody perception. Stimulation over fronto-parietal operculum delivered with a frequency of 1 Hz interrupted the emotional prosody perception at the word level (van Rijn et al., 2005). In another study by Barrett et al. (2004a), rTMS was used to modulate affective speech production in combination with PET scanning to localize brain areas that were affected by the stimulation. After 10 Hz stimulation over the left DLPFC, subjects read emotional sentences with a lower fundamental frequency $F_0$ and reported increased self-negative affect. To summarize, rTMS is a reliable research tool able to provoke changes in brain regions involved in emotional prosody processing.
1.10 Event-Related Brain Potentials (ERPs)

1.10.1 Principles of ERPs

Event-Related Brain Potentials (ERPs) are extracted from the electroencephalogram (EEG) and comprise various positive and negative waveforms of different duration that reflect the brain response to changes (events) in the external or internal environment of the organism (Picton and Hillyard, 1988). As stated by Luck (2005), ERPs reflect postsynaptic neuronal activity. Postsynaptic potentials, typically, last tens to hundreds of milliseconds and can be measured under certain circumstances. Hence, two conditions must be met to allow for recording the summated voltages of postsynaptic potentials. Firstly, they have to be spatially aligned, and secondly, they have to occur at the same time across thousands of neurons. Most likely, this process occurs in the cortical pyramidal cells that are aligned perpendicular to the surface of the cortex (see Figure 1.6).

The alignment and firing of pyramidal cells satisfy the constraints for an observable signal. Since even the summated voltages of postsynaptic potentials are small, they have to be extracted from background activity by averaging. The aim of averaging is to enhance the signal and reduce random noise to nearly zero, thus improving the signal-to-noise ratio by a factor proportional to the square root of the number of trials (Kutas and Dale, 1997). Furthermore, averaged EEG epochs yield a single vector representing the neural activity at each time point reflecting the stimulus processing (Rugg and Coles, 1995). Figure 1.7 shows steps of the measurement procedure.

Since the process of speech perception happens quite quickly, in the range of milliseconds, ERPs are a reliable tool for studying the underlying cognitive processes. The hallmark of the speech perception research was a study by Kutas and Hillyard (1980). They identified the N400 component which corresponds to an incongruous semantic end of a given sentence, e.g. “He spread his warm bread with his...socks.”

Another ERP component denoted as the mismatch negativity (MMN) is recognized as a marker of automatic, pre-attentive, feature analysis in the auditory cortex (Näätänen and Alho, 1995; Näätänen, 2001). It is a negative-going wave, largest at central midline scalp sites, and typically peaks between 160 and 220 ms (Luck, 2005). It is caused by the
1.10. Event-Related Brain Potentials (ERPs)

Figure 1.6: (A) A scheme of pyramidal cell during neurotransmission. An excitatory neurotransmitter is released from the presynaptic terminals, causing positive ions to flow into the postsynaptic neuron. This creates a negative extracellular voltage (represented by the “-” symbols) yielding a small dipole. (B) Folded sheet of cortex containing many pyramidal cells. When a region of this sheet is stimulated, the dipoles from the individual neurons summate. (C) The summated dipoles from the individual neurons can be approximated by a single equivalent current dipole (shown as an arrow). (D) Example of a current dipole with a magnetic field travelling around it. (E) Example of the magnetic field generated by a dipole that lies just inside the surface of the skull. Adopted from Luck (2005).

Presentation of a physically deviant stimulus (in pitch, duration, intensity) to the standard stimulus, in an oddball paradigm. Such detection implies that there is a memory trace of the physical standard against which the deviant can be compared. It occurs without active attention of the listener, usually while reading a book or watching a silent move.

Opposite to this pre-attentive processing, another ERP wave known as the $P300$, reflects the stimulus evaluation and categorization process. The $P300$ is not a unitary component and can be divided to different subcomponents. A subcomponent of $P300$, the $P3b$, reflects the evaluation of task-relevant stimuli that are presented in a stream of standard stimuli in a so-called oddball paradigm. The $P3b$ reaches maximal amplitudes over parieto-central
1.10. Event-Related Brain Potentials (ERPs)

Figure 1.7: Idealized waveform of the computer-averaged auditory event-related potential (ERP) to the brief sound. The ERP is generally too small to be detected in the ongoing EEG (top) and requires computer averaging over many stimuli presentations to achieve adequate signal-to-noise ratios. The logarithmic time display allows visualization of the waves (I-VI), the mid-latency components (No, Po, Na, Pa, Nb), the vertex potential waves (P1, N1, P2), and task-related endogenous components (Nd, N2, P3 and slow wave). Figure adopted from Rugg and Coles (1995).

areas and has a latency of at least 300 ms up to 900 ms (Rugg and Coles, 1995). It is known that latency and amplitude of this component depend on the task relevance of the stimulus, as well as on how difficult the categorization task is (Johnson, 1986). Donchin (1981) suggested that $P3b$ reflects the process that he named “context updating” (updating one's representation of the current environment). However, his hypothesis has not been completely accepted as different opinions and criticism have occurred (Verleger, 1988).

1.10.2 ERPs in prosody research

To date, a lot of data regarding prosody perception have been gathered and various components have been identified using different types of paradigms and stimuli. An ERP component known as the Closure Positive Shift (CPS) was described by investigating processing phrase boundaries (IPh) in sentence perception. It is assumed that CPS reflects purely
prosodic processes (Steinhauer et al., 1999). Comparison of sentences spoken in the negative or positive prosodic tones to neutrally-intonated ones revealed a positive component peaking around 200 ms after the stimulus onset named $P_{200}$ (Alter et al., 2003). In another experiment, when instead of words or sentences, meaningless emotional exclamations were used as stimuli (Bostanov and Kotchoubey, 2004), a negative component, the $N_{300}$ was observed. Wambacq et al. (2004) found that the $P_{350}$ component also was modulated by prosodic factors and had a higher amplitude for stimuli carrying the prosodic, rather than for those carrying the semantic information. An ERP study by Schirmer et al. (2005) revealed gender differences in the perception of emotional prosody already at a pre-attentional level. To summarize, measuring ERPs enables us to identify markers of pre-attentive and attentive processes of emotional prosody perception. In the second experiment described in this thesis, ERPs are used to test the influence of the dimensions of valence and arousal on pre-attentive and attentive processing of emotional prosody.

1.11 Thesis overview

The work presented in this thesis considers two different experiments. In the first experiment, rTMS was used to investigate brain areas involved in the production of emotional prosody. The rTMS was applied over the right and the left dorsolateral prefrontal cortex (DLPFC) and as a sham stimulation. After the stimulation, the subject’s voice was recorded when producing happy, sad, and neutral prosody. Fundamental frequency $F_0$ of these speech recordings was evaluated and compared for all stimulation conditions. Additionally, the effect of stimulation on the mood of the participants was assessed.

In the second experiment, ERPs were used to investigate attentive and pre-attentive emotional prosody perceptions. A set of prosodic stimuli was presented in a passive and active oddball paradigm to a group of healthy participants. The ERP components and behavioral data (correct responses) were recorded and compared across conditions.
1.11.1 Objectives

The main objective of this thesis was to investigate neural correlates of emotional prosody production and perception, using two different electrophysiological methods. The tasks performed to confirm the initial hypothesis can be summarized as follows:

• **Emotional prosody production** was examined using high frequency rTMS to transiently module the activity of the DLPFC. The hypothesis was that the left DLPFC stimulation would influence the acoustic parameters of non-emotional speech, while the stimulation over the right DLPFC would lead to changes in emotional prosody production. Moreover, by stimulating the left and right hemispheres, results should enable us to draw conclusions about possible lateralization of emotional speech production. In addition, mood changes after rTMS were documented by a self assessment questionnaire. The mood of healthy participants was assessed in order to observe whether they correlated with emotional speech changes.

• **Emotional prosody perception** was examined using ERPs. In an oddball paradigm, single words were presented to healthy participants differing either in arousal or valence dimension of emotional prosody. The hypothesis was, firstly, that when presented passively, arousal and valence dimensions would influence already pre-attentive processing, and that this influence would be reflected by differences in the MMN; secondly, influence of the arousal and valence dimensions on attentive prosody perception would be reflected by behavioral data (hit rate) and the $P3b$ component.
Chapter 2

Methods and materials

This chapter details the laboratory settings and techniques used in rTMS and ERP experiment. The basic principles of these techniques have been introduced in the previous chapter. Experiments design and paradigms used to test the emotional prosody production and perception are detailed and further explained.

2.1 The rTMS Experiment

2.1.1 Subjects

Sixteen right-handed, healthy subjects (8 women and 8 men, mean age = 24.5 yrs.; SD = 3.22 yrs.) participated in the experiment. They were recruited by advertisement and received a reimbursement of 50 euros for their participation in the experiment. All participants were native German speakers and had no history of neurological, psychiatric and hearing illness or speech problems. None of the subjects had contraindications for the magnetic stimulation. In order to control for depression, Beck’s Depression Inventory (BDI) (Beck et al., 1996) has been applied and none of the subjects scored higher than nine points (mean = 1.31; SD = 1.4). The study was approved by the Ethics Committee of the Hannover Medical School and all the participants gave written informed consent prior to the start of the experimental procedure.
2.1. The rTMS Experiment

2.1.2 The rTMS procedure

Repetitive transcranial magnetic stimulation (rTMS) was performed using the Magstim Rapid device (Magstim Company Ltd, Whitland, UK) and a figure-of-eight shaped coil. The resting motor threshold (RMT) was determined the day before rTMS procedure and speech recordings. The single-pulse TMS was applied over the motor cortex and the motor evoked potentials (MEPs) were recorded from the first dorsal interosseus muscle (FDI) of the contralateral hand. RMT was defined as the minimal intensity expressed as a percentage of the maximal stimulator output capable of inducing MEPs $> 50\mu V$ in at least five out of ten consecutive trials. Mean RMT was 45.25% (SD = 7) of the maximum stimulator output. Stimulation sites were kept stable using a standardized EEG cap according to the 10-20 system (Jasper, 1958). Stimuli were delivered at the position $F_3$ for the left side DLPFC and $F_4$ for the right DLPFC as suggested by Mottaghy et al. (2000). Figure 2.1 and Figure 2.2 show an example of the 10–20 EEG system from the side and the top view, respectively. The coil was positioned in a way that its anterior tip was the closest to the cortical sites with its remaining part tilted away from the skull (Figure 2.3).

![Figure 2.1: EEG 10 – 20 system side view.](image)

Sham stimulation was carried out in the same manner as the active one, except that the
2.1. The rTMS Experiment

Figure 2.2: EEG 10 – 20 system top view.

Figure 2.3: Active stimulation delivered over the left DLPFC.
coil, rather than being placed tangentially to the skull was positioned at an angle of 45° to the skull (Hoffman et al., 2000) as shown in the Figure 2.4. rTMS was applied in the randomized order of sessions. To avoid carry-over and learning effects, at least a five-day break was allowed between the two consecutive experimental sessions.

Figure 2.4: Sham stimulation delivered with coil positioned 45° off the skull.

2.1.3 Stimulation protocol

Three series of stimulation, each of a 2.5 min duration, were delivered using a frequency of 10 Hz and an intensity of 100% of the resting motor threshold (RMT). Each of the 2.5 min series consisted of 15 1-s trains with a 10-s between train-intervals resulting in a total of 450 stimuli in 27.5 min. In accordance with the procedure used by Barrett et al. (2004a), the first and the second series of rTMS were followed by a time break of 10 min duration. During given breaks, subjects were laying in a comfortable chair with their eyes closed, wearing ear plugs. There was no contact between examiners and the subjects during this period, except at the very end, when the next series of stimulations was announced. Immediately before and after the rTMS, subjects also completed a self-assessment mood questionnaire (EWL-k). After the stimulation, the completion of the questionnaire was
2.1. The rTMS Experiment

2.1.4 Speech tasks

Subjects had to pronounce the word “Anna” in happy or sad (emotional) as well as in neutral (non-emotional) intonation as requested by a visual cue displayed on the monitor in front of them (see Figure 2.6). As mentioned above, speech recordings were performed right after each rTMS session. In order to obtain basic speech values for each subject, on a separate day, and before the start of rTMS sessions, speech recording session was also performed. In each recording session, nine blocks of visual cues were displayed. Each block consisted of 6 items, yielding a total of 18 items per intonation. Block variations were randomized, to avoid immediate repetition of the same emotional condition. Prior to each block, subjects underwent a test trial of five examples. As previously described by Hammerschmidt and Jürgens (2007), a single word was used rather than a whole sentence since emotional expressiveness varies throughout the sentence. Hence, acoustical analysis expanded to the whole sentence would “dilute” the acoustic characteristic of each emotion.
2.1.5 Speech recordings

Speech samples were recorded using an AKG-C-420 microphone (AKG acoustics, Vienna, Austria). Furthermore, they were digitized into wave-format (at 44.1 kHz) and stored on a tape using an external sound card, Creative SoundBlaster® Extigy (Creative Technology Ltd, USA). Speech intensity was calibrated to the average level of spontaneous speech (Bakken and Orlikoff, 2000). The procedure was performed individually, for each subject utilizing digital sound level meter 329 Voltcraft® (Conrad Electronics, Wernberg, Germany).

![Figure 2.6: Computerized paradigm for speech induction. First, visual cue was presented, suggesting to the participants what type of intonation to produce. After 1.5 sec, it was followed by another visual sign (pictogram of a mouth), telling the participants to start by pronouncing the word “Anna.”](image)

2.1.6 Self assessment questionnaires

Eigenschaftswörterliste (EWL-k)

Self-reported feelings of activation and mood were assessed by a characteristic trait word list (Eigenschaftswörterliste (kurz)-EWL-k). The EWL-k list created by Janke and Debus (1978) consists of 123 adjectives describing the actual state of mood on 6 scales (working activity, general inactivity, extroversion/introversion, general well-being, arousal, and fear)
2.1. The rTMS Experiment

and 14 subscales (activity, inactivity, fatigue, indifference, extroversion, introversion, self-confidence, exalted mood, vigorousness, sensitivity, irritation, anxiousness, depressiveness, and dreaminess).

Beck’s Depression Inventory (BDI)

Beck’s Depression Inventory (BDI) (Beck et al., 1996) is a 21 multiple-choice item, self-report mood questionnaire widely used to measure the absence or presence of depression in terms of its intensity, severity, and depth. Each of the questions in this questionnaire had four possible answers. The test was evaluated by rating each question on a 0-3 scale and calculating the total score to determine the severity of depression. A score in the range from 0-9 indicated that the person had no depression, 9-18 pointed to moderate depression, 19-29 moderate to severe, and 30-63 indicated severe depression.

2.1.7 The sound analysis

Speech samples were evaluated using the speech analysis software PRAAT (Boersma and Weenink, 1996). Sound clips containing both “a” vowels were cut out from the speech stream of “Anna” separately for each subject and each intonation (happy, neutral, and sad), three rTMS and one pre-rTMS condition. The obtained mean values of the fundamental frequencies $F_0$ and its standard deviation parameter that reflects modulation of fundamental frequency were further used to compare emotional prosody production between different emotions and stimulation conditions.

2.1.8 Statistical analysis

Two-way repeated measures analysis of variance (ANOVA) was conducted for the mean fundamental frequency $F_0$ including the factors emotion (corresponding to neutral, happy and sad intonations) and stimulation (corresponding to pre-rTMS, left, right, and sham rTMS conditions). The same test was performed for the measure standard deviation of the fundamental frequency $F_0$. Post hoc analysis was performed using Student’s t-test for paired samples, with the level of statistical significance set at $p < 0.05$. Mean scores on the characteristic trait word list (EWL-k) were analysed separately, for each session using
2.2 The ERP experiment

2.2.1 Subjects

Twenty healthy subjects (10 male, 10 female) participated in this experiment. All subjects were right-handed by self report with neither hearing problems nor history of any neurological or psychiatric diseases. The participants were recruited by public advertisement and were paid 25 euros for participating in the study. The age range was between 19 and 35 years (mean = 27.20, SD = 4.72). All subjects signed written consent prior to the beginning of the experiment. The study was approved by the Ethic Committee of Hannover Medical School.

2.2.2 Stimulus material

Auditory stimuli consisted of semantically neutral words spoken in different emotional intonations with either negative or positive valence and high or low arousal. The words were recorded by two professional speakers (male and female) in a happy, sad, angry, and relaxed prosody. After pre-evaluation, 12 semantically-neutral words were used: Bürger (citizen), Zeitraum (space), Gegend (area), Kaufhaus (store), Tasche (bag), Kasten (box), Treppe (stairs), Plattform (platform), Denkmal (monument), Kreuzung (crossroad), Beitrag (article) and Bahnhof (railway station). For pre-evaluation, an on-line questionnaire was answered by 20 participants. After an auditory presentation of the recorded words, subjects had to recognize and rate the emotional tone of each presented word on a scale from 0 to 5. The mean values of fundamental frequency $F_0$, duration and intensity of the presented emotional intonated words are summarized in Table 2.1.

The values of fundamental frequency $F_0$ (Hz) differed significantly between angry and sad ($p < 0.0001$) and angry and relaxed intonated words ($p < 0.0001$), as well as between happy and sad ($p = 0.0001$) and happy and relaxed intonated words ($p < 0.0001$). The duration of the two-syllable words differed significantly when comparing angry and relaxed ($p < 0.0001$), happy and sad ($p = 0.03$) and happy and relaxed spoken words ($p < 0.0001$).
2.2. The ERP experiment

### Table 2.1: Stimuli parameters.

<table>
<thead>
<tr>
<th>Prosodic intonation</th>
<th>Speech parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Duration [s]</td>
</tr>
<tr>
<td>Happy</td>
<td>0.55</td>
</tr>
<tr>
<td>Angry</td>
<td>0.52</td>
</tr>
<tr>
<td>Sad</td>
<td>0.57</td>
</tr>
<tr>
<td>Relaxed</td>
<td>0.92</td>
</tr>
</tbody>
</table>

#### 2.2.3 An Oddball paradigm

Stimuli were divided and presented in different pairs to allow assessing the impact of arousal and valence dimension on emotional prosody perception. In a so-called “valence task”, words that differed in valence while having the same arousal dimension were combined: angry (negative valence; high arousal)/happy (positive valence; high arousal) and sad (negative valence; low arousal)/relaxed (positive valence; low arousal) (see Figure 2.7). Respectively, in the “arousal task”, angry (negative valence; high arousal)/sad (negative valence; low arousal) and happy (positive valence; high arousal)/relaxed (positive valence; low arousal) pairs were used (see Figure 2.8). All stimuli served both as deviants and as standards, thus allowing for reducing the influence of physical differences on ERP waveforms. For the same reason, stimuli were randomized in the way that the standard and the deviant were always spoken by the speaker of the same gender. Stimuli were presented in an oddball paradigm, with a passive and active condition. In the passive one, subjects were reading the material of their own choice without paying attention to the stimuli that were presented via earphones. In the active one, subjects had to classify valence or arousal of deviant stimulus and to indicate their choice by pressing the appropriate button on a keyboard. In the passive condition, the interstimulus interval (ISI) was set to 550 ms, while in the active to 1250 ms. 800 stimuli were presented in four blocks in the passive and eight blocks in the active condition, such that 80% of stimuli were introduced as standards and 20% as deviants.
2.2. The ERP experiment

Figure 2.7: Schematic representation of the oddball paradigm shown for the valence task, (angry: standard/ happy: deviant) in the upper row, and (sad: standard/ relaxed: deviant) in the lower row. To avoid influence of physical differences on ERP waveforms, all stimuli were presented both as standards and deviants in different runs.

Figure 2.8: Schematic representation of the oddball paradigm shown for the arousal task (happy: standard/ relaxed: deviant) in the upper row, and (sad: standard/ angry: deviant) in the lower row. To avoid the influence of physical differences on ERP waveforms, all stimuli were presented both as standards and deviants in different runs.
2.2.4 Behavioral data recordings

As aforementioned, behavioral data (hit rate) were obtained from subjects in the active part of the experiment in parallel with the EEG recordings. In the valence task, subjects had to classify deviant stimulus as sounding positive or negative. Accordingly, in the arousal task, deviant stimulus had to be rated as sounding calm or arousing.

2.2.5 Electrophysiological recordings

Brain electrical activity was recorded using the Neuroscan system (Herndon, Virginia, USA). An EEG recording cap (EASYCAP GmbH, Herrsching-Breitbrunn, Germany) with 28 silver chloride electrodes according to the 10–20 electrode EEG system (Jasper, 1958) was fixed to the participant’s scalp. Vertical and lateral eye movements (VEOG) were monitored through the electrodes placed on the canthus and below the orbit of the left eye with the left mastoid electrode as a reference. Figure 2.9 presents a typical measurement setup. Impedance for all electrodes was measured before starting each EEG recording and kept under 5 KOhm. The EEG signal was amplified and band-pass filtered from 0.1 to 50 Hz and further on digitized at 250 Hz.

2.2.6 Chamber setup

Participants were seated in a medical exam chair with an arm rest. The computer screen was placed at eye level at a distance of approximately 1.5 m and was used to provide information about the experimental procedure. During the active condition, a response keyboard with appropriate buttons was placed in front of each participant. Figure 2.10 shows the chamber setup used for performing the experimental procedure throughout the ERP study done in this thesis.

2.2.7 Data analysis

The ERP data analysis was carried out using the Event-Related Potential Software System ERPSS from the Event-Related Potential Laboratory, University of California at San Diego (Hansen, 1993). The EEG was filtered on-line (band pass 0.1 – 50 HZ). Averages
2.2. The ERP experiment

Figure 2.9: Recording setup showing placement of the electrodes on Easycap and those used for eye movement monitoring.

were obtained after artefact rejection based on individualized amplitude criteria on eye channels/frontal electrodes. Data were acquired from 11 subjects in the valence task and 9 subjects in the arousal task. All the trials with voltage values exceeding 50µV in any EEG channel were considered as artefacts and were excluded from further analysis. Averages for each participant and overall grand averages for all participants were calculated for each experimental condition using a 100 ms prestimulus baseline. Repeated measures analysis of variance (ANOVA), including the factors emotion (happy, angry, relaxed, and sad) and electrode site as well as the between-subject factor group were calculated using the Huynh-Feldt epsilon correction. The level of statistical significance was set at $p < 0.05$. 
Figure 2.10: Chamber setup for performing the ERP experiment.
Chapter 3

Results

This chapter details the results of the rTMS and ERP experiment that were conducted to delineate emotional prosody perception and production. Material and methods used in this investigation have been introduced in the previous chapter. Statistical data analysis was done and the statistical significance of the output data was assessed and further detailed.

3.1 The rTMS experiment

3.1.1 Speech Task

Mean values of the mean fundamental frequency $F_0$ and their corresponding standard errors of the mean (SEM) acquired in different experimental conditions are summarized in Table 3.1.

Statistical analysis for mean $F_0$ values was performed using two-way repeated-measures ANOVA. As it can be seen in Table 3.1, significant differences were observed for factor emotion ($F(2, 30) = 38.37, p = 0.0001$) and no significance for the factor stimulation ($F(3, 45) = 0.2, p = 0.087$). There was also no significant interaction between these two factors ($F(6, 90) = 0.95, p = 0.46$). The same statistical procedure, two-way repeated-measures ANOVA, was done for the standard deviation of the fundamental frequency. Regarding the $F_0$ modulation measure (standard deviation of the fundamental frequency $F_0$), a statistically significant effect was observed for the main factor emotion ($F(2, 30) = 79.25, p = 0.0001$), and no main effect for the factor stimulation ($F(3, 45) = 2.62, p = 0.06$). There was statis-
3.1. The rTMS Experiment

<table>
<thead>
<tr>
<th>Emotional condition</th>
<th>rTMS condition</th>
<th>pre-rTMS</th>
<th>left rTMS</th>
<th>right rTMS</th>
<th>sham rTMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral</td>
<td>153.17</td>
<td>146.52</td>
<td>150.36</td>
<td>147.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.75</td>
<td>11.52</td>
<td>11.72</td>
<td>11.42</td>
<td></td>
</tr>
<tr>
<td>happy</td>
<td>191.1</td>
<td>189.62</td>
<td>200.78</td>
<td>190.64</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.46</td>
<td>12.84</td>
<td>15.46</td>
<td>12.43</td>
<td></td>
</tr>
<tr>
<td>sad</td>
<td>150.16</td>
<td>140.49</td>
<td>148.69</td>
<td>149.56</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.67</td>
<td>13.56</td>
<td>13.57</td>
<td>12.47</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: Table shows mean values of the mean fundamental frequencies $F_0$ (Hz) for the different intonations (neutral, happy, and sad) and rTMS conditions (pre-rTMS, left, right, and sham). The asterisk symbol * ($p < 0.05$) indicates a significant interaction, while the standard error of the mean (SEM) is shown in italics.

There was a statistically significant interaction between these two factors ($F(6, 90) = 2.23, p = 0.046$). Post hoc analysis for happily intonated words showed a statistically significant increase after the right-sided rTMS compared to the pre-rTMS condition ($T(15) = 3.0, p = 0.01$). However, this effect was also observed after the sham stimulation compared to pre-rTMS condition ($T(15) = 2.4, p = 0.03$). Standard deviations of $F_0$ values and their corresponding standard errors of the mean (SEM) for the different experimental conditions are summarized in Table 3.2.

<table>
<thead>
<tr>
<th>Emotional condition</th>
<th>rTMS condition</th>
<th>pre-rTMS</th>
<th>left rTMS</th>
<th>right rTMS</th>
<th>sham rTMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>neutral</td>
<td>14.66</td>
<td>13.97</td>
<td>15.80</td>
<td>16.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.16</td>
<td>1.81</td>
<td>2.60</td>
<td>2.91</td>
<td></td>
</tr>
<tr>
<td>happy</td>
<td>48.054</td>
<td>56.59</td>
<td>61.454*</td>
<td>59.91*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.80</td>
<td>5.44</td>
<td>5.82</td>
<td>5.81</td>
<td></td>
</tr>
<tr>
<td>sad</td>
<td>16.72</td>
<td>17.76</td>
<td>18.20</td>
<td>19.19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.76</td>
<td>2.78</td>
<td>2.46</td>
<td>2.28</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2: Table shows mean values for standard deviation of fundamental frequencies $F_0$ (Hz) for the different intonations (neutral, happy, and sad) and rTMS conditions (pre-rTMS, left, right, and sham). The asterisk symbol * ($p < 0.05$) indicates significant interaction, while the standard error of the mean (SEM) is shown in italics.
3.1.2 Behavioral data

Statistical analysis showed significant differences in several subscales after different rTMS conditions. Compared to the values before the stimulation, sham stimulation yielded significant differences for the scales inactivity ($p = 0.0008$), introversion ($p = 0.02$) and irritation ($p = 0.01$). Additionally, significant differences were observed for the subscale dreaminess after left ($p = 0.0001$), right ($p < 0.0001$) and sham stimulation ($p = 0.05$). All three rTMS conditions also led to changes in the subscale anxiousness with different $p$ values for the left ($p = 0.0008$), right ($p = 0.0009$) and sham stimulation ($p < 0.0001$). Left stimulation led to changes in the scale sensitivity ($p = 0.03$) and irritation ($p = 0.04$), and, most interestingly, in the scale depressiveness ($p = 0.01$), indicating a more depressed mood following this stimulation condition. Correspondingly, tests revealed an elevated mood after the right side stimulation ($p < 0.0001$). The values for different subscales in the EWL questionnaire, for all stimulation and one pre-stimulation conditions are summarized in Table 3.3.

<table>
<thead>
<tr>
<th>EWL-k</th>
<th>EWL subscales</th>
<th>pre-rTMS</th>
<th>right rTMS</th>
<th>left rTMS</th>
<th>sham rTMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Working activity</td>
<td>Activity</td>
<td>0.36</td>
<td>0.37</td>
<td>0.33</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>Inactivity</td>
<td>.33</td>
<td>.37</td>
<td>.37</td>
<td>.41*</td>
</tr>
<tr>
<td>General inactivity</td>
<td>Fatigue</td>
<td>.30</td>
<td>.31</td>
<td>.37</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td>Indifference</td>
<td>.31</td>
<td>.26</td>
<td>.21*</td>
<td>.28</td>
</tr>
<tr>
<td>Extraversion/introversion</td>
<td>Extraversion</td>
<td>.22</td>
<td>.21</td>
<td>.28</td>
<td>.23</td>
</tr>
<tr>
<td></td>
<td>Introversion</td>
<td>.35</td>
<td>.37</td>
<td>.35</td>
<td>.42*</td>
</tr>
<tr>
<td>General well-being</td>
<td>Self-confidence</td>
<td>.07</td>
<td>.07</td>
<td>.08</td>
<td>.14*</td>
</tr>
<tr>
<td></td>
<td>Exalted Mood</td>
<td>.06</td>
<td>.29***</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td>Arousal</td>
<td>Vigorousness</td>
<td>.26</td>
<td>.23</td>
<td>.26</td>
<td>.30*</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>.21</td>
<td>.28</td>
<td>.20*</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td>Irritation</td>
<td>.22</td>
<td>.06</td>
<td>.30**</td>
<td>.30*</td>
</tr>
<tr>
<td>Fear</td>
<td>Anxiousness</td>
<td>.02</td>
<td>.24***</td>
<td>.20***</td>
<td>.07*</td>
</tr>
<tr>
<td></td>
<td>Depressiveness</td>
<td>.10</td>
<td>.21</td>
<td>.23*</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>Dreaminess</td>
<td>.20</td>
<td>.3***</td>
<td>.44***</td>
<td>.27*</td>
</tr>
</tbody>
</table>

Table 3.3: Table shows values for 6 scales and 14 subscales of EWL-k questionnaire and changes after different rTMS conditions (pre-rTMS, left, right, and sham). The asterisk symbol * ($p < 0.05$) indicates significant interaction.
3.2 The ERP experiment

3.2.1 ERP data

Passive condition

Figure 3.1 to Figure 3.4 show grand average waveforms of four out of overall eight standard/deviant combinations. For the frontal and central electrode sites, Fz and Cz, and time window 100-300 ms, ANOVA demonstrated a significant effect for factor electrode \( (F(1,20) = 4.10, p = 0.04) \), but no significance for the factor intonation \( (F(1,20) = 0.74, p = 0.63) \) and intonation × electrode interaction \( (F(1,20) = 0.52, p = 0.82) \).

![Waveforms for sad intonations and relaxed voice intonation](image)

Figure 3.1: The ERPs for sad intonations (presented as standards) and relaxed voice intonation (presented as deviants) in the valence task and passive condition.

Additionally, deviants and standards were compared across the blocks, such that the same physical stimulus served one time as a standard and another time as a deviant. For the electrode sites Fz and Cz and time window 100-300 ms, no significant effects were observed for neither main factors electrode \( (F(1,20) = 0.11, p = 0.73) \) and intonation \( (F(1,20) = 0.74, p = 0.63) \)
3.2. The ERP experiment

The corresponding grand average waveforms are shown in Figure 3.5. In addition, comparison of ERPs evoked by deviants of all four prosodic intonations was also performed (electrodes Fz/Cz; time window 100-300 ms) (see Figure 3.6). The level of statistical significance was not reached for the factor electrode \((F(1, 20) = 0.24, p = 0.6)\), factor intonation \((F(1, 20) = 0.25, p = 0.86)\) and intonation \(\times\) electrode interaction \((F(1, 20) = 0.05, p = 0.9)\).

3.2.2 Active condition

Valence task

The grand average waveforms for the active valence task are shown in Figure 3.7. Deviant stimuli elicited a positive component over the parieto-occipital electrodes with the maximum between 400-600 ms, suggesting the \(P3b\) component. A significant effect was observed...
between the ERPs elicited by deviant stimuli of different intonations on the occipital Oz electrode, \textit{intonation} \times \textit{electrode} interaction ($F(1, 11) = 4.19, p = 0.01$). Post hoc comparison showed significantly higher mean $P3b$ amplitude values for happy intonation when compared to the mean amplitudes for both sad ($p = 0.027$) and relaxed intonations ($p = 0.05$). No significant effect was noticed between amplitudes evoked by angry and happy ($p = 0.59$) and angry and relaxed voice intonations ($p = 0.12$). However, a trend to significance was observed between amplitudes evoked by angry and sad intonations ($p = 0.08$).

Furthermore, the ERP amplitudes measured over the right and the left electrodes were compared to control for a lateralization effect. They were compared to those acquired over the midline electrode site. The test was performed for the occipital electrode sites, $O1/Oz/O1$, and 400-600 ms time window. Most interestingly, angry intonated words, elicited higher amplitudes over the left hemisphere ($F(1, 11) = 5.00, p = 0.036$), but not over the right one ($F(1, 11) = 0.60, p = 0.44$). Happy intonated words did not modulate the
3.2. The ERP experiment

Figure 3.4: The ERPs for happy prosodic intonations (presented as standards) and relaxed prosodic intonations of voice (presented as deviants) in the arousal task and passive condition.

amplitude differently and did not produce a significant effect neither over the left \(F(1, 11) = 0.58, p = 0.45\) nor over the right hemisphere \(F(1, 11) = 0.79, p = 0.38\).

**Arousal task**

The analysis of the attended stimuli in the arousal task showed no significant difference (ANOVA, Oz, 400-600 ms electrode × emotion interaction) \(F(1, 9) = 0.11, p = 1\) (see Figure 3.8). In order to control for the lateralization effect, a similar analysis as in the valence task was performed. There were no significant differences between amplitudes evoked by angry intonation neither over the right \(F(1, 9) = 0.82, p = 0.37\) nor over the left hemisphere \(F(1, 9) = 1.26, p = 0.2\). Likewise, there were no significant differences observed for the ERPs elicited by happily spoken words over the right \(F(1, 9) = 0.83, p = 0.37\) and the left hemisphere \(F(1, 9) = 0.17, p = 0.68\).
3.2. The ERP experiment

Figure 3.5: The grand average waveforms elicited by all four prosodic intonations (happy, angry, relaxed and sad), serving as standards and deviants, in all runs in passive condition.

Overall comparison

In order to achieve higher statistical power and significance of the observed data sample, the grand average waveforms for target stimuli from both tasks were calculated together as shown in Figure 3.9. Similarly to previous comparisons, a positive component over the parieto-occipital electrodes and with the maximum between 400-600 ms was observed. However, there was a significant electrode × intonation interaction found (Oz, 400-600 ms) ($F(1,20) = 2.7, p = 0.05$). Furthermore, after post hoc analysis, a statistically significant difference was observed between ERPs for happy and relaxed ($p = 0.02$), happy and sad ($p = 0.02$) and trend to significance between angry and relaxed intonations ($p = 0.08$). There were no significant differences between ERPs for angry and happy ($p = 0.22$) and angry and sad emotional intonations ($p = 0.2$).

In order to control for a lateralization effect, the amplitudes measured over the right and the left occipital electrode sites (O1/O2) were compared to those measured over the
3.2. The ERP experiment

Figure 3.6: The ERPs elicited by the deviant stimuli in all prosodic intonations (happy, angry, relaxed, and sad) in the passive condition.

midline electrode site (Oz). $P3b$ amplitudes elicited by angrily spoken words were significantly higher over the left electrode sites ($F(1, 20) = 5.08, p = 0.03$); such a difference was not observed over the right electrode sites ($F(1, 20) = 1.49, p = 0.22$). There were no statistically significant differences between the ERPs elicited by happy intonation over the left ($F(1, 20) = 0.79, p = 0.19$) and right hemisphere ($F(1, 20) = 0.79, p = 0.38$). Differences were not observed between EPRs elicited by sad intonations over the left ($F(1, 20) = 1.57, p = 0.21$) and the right ($F(1, 20) = 1.18, p = 0.28$) and the ERPs evoked by the relaxed voice intonations over the left ($F(1, 20) = 0.36, p = 0.54$) and the right hemisphere ($F(1, 20) = 0.51, p = 0.47$) as well.

3.2.3 Behavioral results

The mean values of correct responses to deviant stimuli for both valence and arousal tasks are shown in Table 3.4. Similarly to grand average waveforms analysis, in order to achieve
3.2. The ERP experiment

Figure 3.7: The ERPs for attended stimuli in the active valence task for all prosodic intonations of the voice (happy, angry, relaxed, and sad).

higher statistical power, the correct responses for both tasks together were calculated, in addition.

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Deviants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Angry</td>
</tr>
<tr>
<td>valence</td>
<td>88.4</td>
</tr>
<tr>
<td></td>
<td>7.5</td>
</tr>
<tr>
<td>arousal</td>
<td>56.1</td>
</tr>
<tr>
<td></td>
<td>17.6</td>
</tr>
<tr>
<td>valence</td>
<td>73.80</td>
</tr>
<tr>
<td>and arousal</td>
<td>2.15</td>
</tr>
</tbody>
</table>

Table 3.4: Table shows mean values (presented in percentage) of correct responses to the deviant stimuli for the different prosodic intonations (angry, happy, relaxed, and sad) and different tasks (valence, arousal, as well as valence and arousal calculated together). The standard deviation (SD) is indicated in italics.

In the valence task, a statistically significant difference was observed between recognition
3.2. The ERP Experiment

Figure 3.8: The ERPs for attended stimuli in active arousal task for all prosodic intonations (happy, angry, relaxed, and sad).

of angry and happy ($p = 0.001$), angry and sad ($p < 0.0001$) and angry and relaxed emotional tones of voice, as well ($p < 0.0001$). The difference was also significant between the recognition of happy and sad ($p = 0.001$) and happy and relaxed intonations ($p = 0.0002$).

In the arousal task, there were no significant differences between the correct responses and recognition of the prosodic intonations (ANOVA; $F(3, 8) = 0.7, p = 0.55$). Later, exact responses from both tasks (valence and arousal) were gathered and compared. It was observed that deviant angry stimuli were better recognized than those spoken sadly ($p < 0.001$) or relaxed ($p = 0.03$). Likewise, happily-spoken deviants were better recognized than low arousal intonations like sad ($p = 0.003$) and relaxed ($p = 0.01$). There were no differences found between the recognition of happily and angrily intonated words ($p = 0.2$).
Figure 3.9: The ERPs for attended stimuli calculated for both tasks, for all prosodic intonations of the voice (happy, angry, relaxed, and sad).
Chapter 4

Discussion

In this chapter, the findings of the two experiments are discussed with reference to the relevant literature. In the first part, conclusions based on the results of the rTMS experiment are drawn. Some limitations are discussed and further research directions are detailed. The second part focuses on the ERP study and describes main outcomes when using this procedure in the research on emotional speech.

4.1 The rTMS experiment

The goal of this experiment was to investigate neural correlates of emotional prosody production using high-frequency rTMS. The stimulation was delivered in separate sessions, bilaterally, over the right and left dorsolateral prefrontal cortex (DLPFC), and also as a placebo stimulation. The hypothesis was that rTMS would influence emotional and non-emotional speech in a different way so that the left and the right stimulations would produce different effects. Afterward, it was hypothesized that the left DLPFC stimulation would influence the acoustic parameters of non-emotional speech, while the stimulation over the right DLPFC would lead to changes of emotional prosody production. With regard to the prosodic speech alternations found in depressed patients (Alpert et al., 2001), the mood of our healthy participants was assessed in order to observe whether changes in emotional speech and mood are correlated.

For the acoustic speech parameter “mean of fundamental frequency $F_0$,” there were no
changes triggered by the stimulation for all three speech intonations. Statistically significant differences were observed for the parameter “standard deviation of fundamental frequency $F_0$” after right and after sham stimulations as well. However, since a non-specific effect after sham stimulation was also found, the hypothesis that rTMS administered at these parameters could induce changes in emotional prosody production, was not confirmed. Nevertheless, mood alterations following real rTMS were observed. When stimulation was applied over the left DLPFC, a transient depression was observed, whereas stimulation over the right DLPFC yielded a transient mood elevation. These mood changes were not correlated with changes in emotional speech, as it was previously observed by Barrett et al. (2004a).

Initially, the idea to investigate the influence of rTMS on emotional speech was based on the study by Barrett et al. (2004a). They showed that the stimulation of the left DLPFC led to $F_0$ changes in paralinguistic aspects of speech. Using Positron Emission Tomography (PET) in combination with rTMS, these authors also showed an increased connectivity between the left DLPFC and anterior cingulate cortex (ACC) after rTMS. They suggested that stimulation of the left prefrontal cortex indirectly affects the activity in the subcortical and cortical brain circuits, provoking speech variations similar to those in depressed patients. Indeed, this region is strongly involved in the regulation of normal affect and in the pathophysiology of depression (Mayberg, 2003; Drevets, 2001; Drevets et al., 1992).

Although this approach was quite promising and innovative, the study was not sham-controlled. The experiment performed in this thesis, using the same stimulation paradigm and procedure, has shown that sham stimulation induced the same changes in affective speech modulation as the active one. Therefore, it is not very probable that this type of stimulation is capable in inducing changes in emotional speech. However, by using this stimulation, it was possible to produce mood changes in a lateralized fashion. This is in line with the previous findings that rTMS over the left DLPFC results in an increase of self-rated sadness, opposite to the stimulation of the right side that induces an increase in self-rated happiness (Dearing et al., 1997; George, 1996; Pascual-Leone, 1996).
The possibility that mood changes do not have to be followed by those in affective speech has been previously discussed by Barrett, Pike, and Paus (2004b). In this study authors investigated the neural activity associated with the speech production during affective states. They proposed that a patient with a lesion affecting ACC would not display affect-induced changes in speech production, even though she/he may still become sad. Although Paus (2001) previously suggested that the ACC modulates aspects of motor output in response to changes in emotion and motivation, the critical aspect of the ACC contribution in control of these processes is still not determined.

It should be mentioned that some researchers failed to demonstrate mood changes after rTMS (Mosimann et al., 2000; Baeken et al., 2006, 2007). In all these studies, different procedures and experimental designs were used which might explain the diversity in findings and the discrepant results. For example, the presence or absence of changes were reported only after one-sided stimulation and without placebo stimulation. Some researchers considered only subjects of one gender or an unequal number of subjects having different genders (Mosimann et al., 2000; Nedjat et al., 1998; Baeken et al., 2006, 2007; Barrett et al., 2004a). Performing experiments on female participants only, or on more female than male healthy volunteers might influence the investigations of the mood (but also of the speech) in several ways. In such cases, the results of mood induction cannot be generalized, since mood changes observed in female participants could depend on actual hormone levels (Sanders, 1983). Women also tend to express their emotions more intensively than men (Kring and Gordon, 1998). Brain-imaging studies reported that females, when compared to male participants, showed increased brain activations during emotional tasks (Wager et al., 2003). Also, acute rTMS can produce different changes of 11C-αMtrp constant $K^*$ (an index of 5HT brain synthesis) in male and female healthy volunteers (Sibon et al., 2007). Additionally, similarly to Barrett et al. (2004a), many studies were not sham-controlled, thus they did not allow comparison of the results after the real and placebo stimulation.

In the experiment presented here, a single-blinded, crossover design was used. The equal number of female and male participants was examined, and bilateral and sham stimulations were applied. Under such controlled conditions, rTMS leads to different changes on the
subscales of the self-assessment mood questionnaire. However, the main effects on the mood, as seen in depression scales, and elevated mood were not bilateral or influenced by the placebo effect. Furthermore, the effect of an elevated mood after the right-sided stimulation was strongly confirmed, with a high level of statistical confidence. Besides the change in the depression scale, there was the trend of more negative emotions after the left-stimulation, specifically in subscales anger and drowsiness. It should be mentioned that, after both active stimulations, higher anxiety and dreaminess were found. In the questionnaire, there were also higher working inactivity and introversion after sham stimulation. These effects, especially bilateral anxiety, can rather be a consequence of the quite robust stimulation procedure that comprises different changes (discomfort due to irritation of the scalp and of facial nerves, noise during the stimulation) that easily affect the subjective feeling of participants and sensitivity to psychological questionnaires. This is observable especially in the case of sham stimulation which produces the same scalp sensations and may also generate higher noise levels due to the specific coil position. In addition, as it was previously suggested by Loo et al. (2000), some sham forms and procedures might be partially active and still provoke some neuronal activity yielding the enhancement of the above-mentioned effects.

In this study, rTMS was not used in conjunction with neuroimaging methods. Therefore, the brain network activated by rTMS could not be shown. In general, there is evidence from previous PET and fMRI studies with healthy subjects showing that rTMS strongly affects the brain activity (Kimbrell et al., 2002; Nahas et al., 2001). A neuroimaging study by Knoch et al. (2006) confirmed that rTMS delivered to the right and left DLPFC increased the blood flow level at the site of stimulation. Unfortunately, these authors did not assess the mood of the participants in their experiment so the possible correlation of rTMS and mood changes remained unknown. In general, it is known that hemispheric asymmetry in the frontal lobes exists (Toga and Thompson, 2003) and that the left and the right frontal regions are differently involved in experiencing negative and positive emotions (Davidson et al., 2004). In spite of the neuropsychological and psychopathological significance of the left and right DLPFC, mood changes after rTMS in healthy people and especially depressive
patients are still not fully understood (Gross et al., 2007).

4.1.1 Summary

In this experiment, rTMS delivered at the specific parameters did not produce changes in emotional prosody production. The stimulation induced mood changes with the influence of the sham stimulation on different mood questionnaire subscales. The exact brain activation pattern in both hemispheres and the explanation regarding the level of activity in the case of sham stimulation could not be precisely identified without the help of neuroimaging methods. Although the use of magnetic stimulation did not lead to results expected according to the initial hypothesis, one should not exclude this method of the investigation of emotional prosody. Using additional stimulation parameters and larger number of stimulation sessions delivered on consequent days (as done in studies with a depressive patients) could be helpful in the analysis of emotional speech and mood changes. Another possible direction to address this problem could be to perform experiments that combine rTMS with neuroimaging methods.

4.2 The ERP experiment

4.2.1 Passive condition: pre-attentive stimulus processing

In this part of the experiment, the high temporal resolution of ERPs was used to investigate pre-attentive processing of emotional prosody and its dimensions of arousal and valence. A series of semantically neutral words (e.g., “Bahnhof” and “Treppe”) spoken in a distinct emotional intonation (for example: happy; 80%) was presented and interrupted by words spoken in a different emotional tone (for example: angry; 20%).

The hypothesis was that, if one is able to establish a standard emotional intonation with the presented words, then any other word that differs with respect to emotional tone, namely every deviant, would elicit the so-called mismatch negativity (MMN). However, the obtained results show that the MMN, which represents the marker of preattentive change detection, was not found. Additionally, when the same category of stimuli, for example, all happy intonated words, serving as a standard or a deviant, respectively, were compared, no
4.2. The ERP experiment

significant differences were found. MMN is an ERP component evoked approximately 100-
200 ms upon the stimulus onset, using a stimulus sequence where acoustically deviant stimuli
are presented in a series of homogeneous stimuli. It is evoked by any discriminable change
that occurs in some repetitive aspect during the auditory stimulation (Näätänen and Alho,
1995; Tervaniemi et al., 1999; Näätänen, 2001). Even though in the present experiment,
standard and deviant stimuli differed in more than one physical attribute (happily spoken
words are louder and have a higher frequency when compared to sadly spoken words),
the MMN component was not found. The question is how to explain the absence of the
MMN. Previous studies have shown that emotionally intonated words in a series of neutrally
spoken words can provoke an MMN (Schirmer et al., 2002; Schröder et al., 2006). The same
holds true for different emotional timbres in music (Goydke et al., 2004). In each of the
above-referred studies, a distinct emotional tone was represented by one specific stimulus.
A typical example is the word “Anna” spoken in certain emotional intonations (Schröder
et al., 2006) or the syllables “dada” (Schirmer et al., 2002). Thus, the observed results
can be explained by either dissimilarities in emotional processing or by simple differences
among physical attributes of the used stimuli. To exclude the possibility that stimuli only
differ in terms of associated physical attributes irrespectively of the emotional tone, in this
experiment a set of words spoken by different speakers was employed. In this way, each
presented word differed from the other. In order to detect the deviant, those physical
attributes being crucial for emotional prosody had to be processed. Nevertheless, following
this approach, it is more difficult to establish a series of standards that can be violated by
a deviant, since standards also differ from each other. Such an approach points to more
complex acoustical structures of the deviant stimuli.

The results of this experiment, can be interpreted as a consequence of a different position
of the neural generators for the MMN, when simple tones and more complex stimuli are
used. There exists evidence that neural generators for the MMN, when evoked by the
complex acoustic signals, are located more medially in comparison to the generators for the
simple tones. When the MMN is elicited by pure tones, the corresponding neural generators
are located more laterally, at approximately 5.6 cm from the midline axis (Alho et al., 1996;
Tiitinen et al., 1993). The generators for the MMN evoked by contrast vowels /ba/ vs. /da/ and /da/ vs. /ga/ are located more medially, at approximately 4.5 cm from the midline of an axis drawn between preauricular points (Diesch and Luce, 1997). In the case of the complex stimuli, neural generators for the MMN might be placed even more medially than for the vowels, with a greater distance to the scalp electrodes, thus, rendering the signal more difficult to be observed. Another explanation might be due to different types of auditory cortex activations and lateral inhibition, when simple tones, in contrast to words, are used. May et al. (1999), investigated the stimulation of auditory cortex with pure tones in an oddball paradigm and pure tones only. According to their study, stimulation of auditory cortex with pure tones in an oddball paradigm resulted in specific adaptation and lateral inhibition of the auditory cortex. In addition, they argue that the MMN to frequency change in the oddball paradigm can be seen as the result of the activity of auditory neurons responding to a deviant tone, under inhibitory influence exerted by standards. As suggested by O’Leary et al. (1996), auditory perception of spectrally complex stimuli yields a broad activation of the auditory cortex. In electrophysiological animal studies on non-human primates, it was demonstrated that complex sounds evoke an extended activation in the auditory cortex (Rauschecker et al., 1995). Also, when comparing the perception of the acoustically-controlled complex speech (sine wave analogue) to the complex non-speech stimuli, larger portions of brain areas were activated (Vouloumanos et al., 2001).

According to the studies reviewed above, it can be seen that stimulation with spectral complex stimuli, such as words, may result in a more wide-spread activation of the auditory cortex. Presentations of a standard stimuli create a broad pattern of adaptation and lateral inhibition, and affect mostly the areas that are subsequently activated by a deviant sound. In these conditions, the overlap between these neuronal populations would be significant, and although the deviant stimulus provokes the neuronal activity, it might not be sufficiently different from the standard to evoke a signal that can be recorded on the scalp.
4.2.2 Active condition: attentive stimulus processing

The main objective of this part of the experiment was to delineate whether attentive processing of emotional prosody differs along the dimensions of valence and arousal. In a so-called oddball paradigm, frequent standard words spoken in a specific emotional prosody, (for example, angry) were interrupted by infrequent deviants (20%). They differ with respect to valence in the valence condition (standard: angry; deviant: happy) and to arousal in the arousal condition (standard: angry; deviant: sad). The participant had to classify deviant stimuli as being positive or negative in the valence condition, and as being of high or low arousal in the arousal condition. As expected, deviant stimuli elicited a $P_{3b}$ component with a parieto-occipital maximum that is usually correlated with stimulus evaluation (Shahin et al., 2006). Interestingly, the results showed significant differences when comparing the data of the arousal and the data of the valence task.

Only in the valence task, $P_{3b}$ waveforms and also behavioral data showed differences between the four emotional tones of voices. Thus, the highest mean amplitudes were elicited when angry and happily spoken words had to be evaluated. This finding correlates with the behavioral data showing a better categorization of angry and happy state compared to relaxed and sad intonations that were, in overall, very low. Hence, stimuli of the high arousal yielded a higher mean amplitude of $P_{3b}$ together with a higher rate of correct classification irrespectively of their prosodic valence.

It is presumed that the $P_{3b}$ component is produced by a neuronal network associated with attention and memory operations (Polich, 2007). It is also known that trait and arousal influence the operability of attention resources and $P_{300}$ measures, observed at the scalp (Kok, 2001; Polich and Kok, 1995). So far, neuroscientific research in the field of emotion and attention has mostly focused on perception of emotional faces and facial expressions. Although brain networks responsible for controlling perception of auditory and visual stimuli differ, some similarities can still be found. The areas that control the voice and the face perception might share the same similar principles of brain organization (Belin et al., 2004). For example, activities of the brain voice-sensitive areas are modulated by the higher threat value of the voice (Grandjean et al., 2005). Such patterns of brain
activation depending on the interaction of emotion and attention have also been found in the fusiform gyrus during the perception of visual stimuli (Vuilleumier et al., 2001). The ERP results showing influence of valence and arousal on attention processing in the visual domain, are not consistent. Experimental data indicate that the valence of the perceived stimulus is taken at the different time points in the information-processing stream (Delplanque et al., 2004; Smith et al., 2003) and that the valence effect can be reflected by the amplitude modulation and the slow positive wave (Smith et al., 2003). Exclusive arousal effects (Carretié and Iglesias, 1995; Cuthbert et al., 2000) and those of both dimensions (Delplanque et al., 2005, 2006) were also demonstrated. In general, stronger emotional effects have been commonly reported, when more arousing, fearful, and happy pictures or facial expressions were compared to neutral ones (Vuilleumier et al., 2002; Schupp et al., 2004; Eimer et al., 2003). Previous studies also reported that the P3 component elicited by a rare target stimuli is sensitive to the arousal dimension, with higher amplitudes obtained for high arousing target stimuli than for less arousing ones (Diedrich et al., 1997).

Research in the auditory domain has shown that emotional stimuli, when compared to neutral ones (Ethofer et al., 2006; Beaucousin et al., 2007; Grandjean et al., 2005) increase the activity in the voice sensitive brain regions i.e., the superior temporal gyrus (STG) (Belin et al., 2002). The responses were found in different conditions, when subjects were instructed to attend either emotional prosody (Ethofer et al., 2006; Beaucousin et al., 2007) or some other speech features (like emotional semantics) (Ethofer et al., 2006) or the speaker’s gender. Recently, using a regression analysis between increased hemodynamical response and acoustic parameters of the speech (Wiethoff et al., 2008), stronger hemodynamical responses for emotional compared to neutral prosody were found, suggesting a strong modulating effect of arousal.

**Lateralization effects**

The reported P3b for angry intonation was significantly higher over the left than over the right hemisphere. This lateralization result is opposite to the theory of emotional valence that predicts a right hemisphere dominance for the negative emotional content and
lateralization toward the left hemisphere for the positive one (Davidson, 1992; Canli et al., 1998). However, such hemispheric asymmetries are not always reliably observed in healthy subjects. Studies that analyzed higher frequency EEG components found higher activity for positive valence at right hemisphere electrodes (Aftanas et al., 1998; Müller et al., 1999). The role of the left hemisphere in the control of attention toward negative stimuli has also been reported (Anderson and Phelps, 2001; Phelps et al., 2001). High arousal of the emotional stimuli also induces higher theta activity over the left hemisphere (Aftanas et al., 2002). The variability of the fundamental frequency correlates with the activity in the left primary auditory cortex and the left angular gyrus (Wiethoff et al., 2008). The left auditory cortex is also involved in the detection of acoustically deviant word stimuli (Tervaniemi et al., 2006). The lateralization toward the left side for angry stimuli might be explained by the influence of physical properties of fundamental frequency, which conveys the most emotional prosody information and correlates with the elevated arousal of the speaker’s voice.

**Summary**

The ERP study presented here consisted of two different parts, in which separately pre-attentive and attentive emotional prosody processing having different degrees of arousal and valence were tested. The first part of the experiment demonstrated that the prosodic intonations of different arousal and valence were not able to elicit the marker of the pre-attentive auditory processing (MMN). The absence of this processing is likely to be the consequence of the complexity of speech sounds used. Moreover, the activation of the auditory cortex during the processing of such complex stimuli may have masked the MMN. In the second part of the experiment, the oddball paradigm and explicit evaluation of the target stimuli were employed. Here, a modulation of the $P3b$ component by the emotional content of the stimuli was shown. The observed modulation (and resources allocation) was influenced by the arousal dimension, but not by the valence of voice intonations. The asymmetry and lateralization effect toward the left hemisphere might also represent the influence of the high arousal, and additionally, acoustic properties of an angry intonation.
Bibliography


Bibliography


M. George. Changes in mood and hormone levels after rapid-rate transcranial magnetic stimulation (rTMS) of the prefrontal cortex, 1996.


M. Gross, L. Nakamura, A. Pascual-Leone, and F. Fregni. Has repetitive transcranial magnetic stimulation (rTMS) treatment for depression improved? A systematic review


C. Lange. *The Emotions*. Williams and Wilkins, Baltimore, Maryland, 1885.


S. van Rijn, A. Aleman, E. van Diessen, C. Berckmoes, G. Vingerhoets, and R. Kahn. Short Communication What is said or how it is said makes a difference: role of the right


Curriculum Vitae

PERSONAL DATA

MILAN ARSIC

Date of Birth: October 19th, 1978
Place of Birth: Nis, Serbia

EDUCATION

• Hanover Medical School and Center of Systemic Neurosciences,
  Hanover, Germany  October 2005 - Present
  PhD. student at the Department of Neurology and Clinical Neurophysiology,
  Research topic: Electrophysiology of emotional prosody production and perception.

• University of Nis, Serbia  1997-2004
  Medical Doctor Degree (MD), Faculty of Medicine.

• Gymnasium “Bora Stankovic”, Nis, Serbia  1993-1997
  High school degree.

WORK EXPERIENCE

• Hanover Medical School, Germany  October 2005 - Present
  Research assistant at the Department of Neurology and Clinical Neurophysiology.

• University Hospital Nis, Serbia  November 2004 - October 2005
  Internship at the Department of Radiology.
  Duties included team work with clinical staff towards learning various radiological
  techniques, obtaining practical skills and gaining experience.
**Scientific Training**

- Course work on methods used in neuropharmacology, behavioral pharmacology, neurophysiology, neuroanatomy and neurochemistry
  Department of Pharmacology, Toxicology and Pharmacy, University of Veterinary medicine, Hanover, Germany, 2008.

- Introduction to application of transcranial magnetic and direct current brain stimulation (TMS/tDCS)
  Department of Neurology, University Clinic Goettingen, Germany, 2007.

- Breast Cancer Screening - Quality Issues in Detection and Diagnosis
  European School of Oncology, Faculty of Medicine, University of Nis, June 2005.

- Specialized training in Emergency Medicine
  Department of Emergency Medicine, University Hospital Nis, Serbia, October 2004.

- Clinical attachments in general surgery, cardiology, gynecology and obstetrics
  University Hospital Nis, Serbia, 2003-2004.

**Advisor for Student Diploma Project**

Magdalena Lefik cand. med., Department of Neurology and Clinical Neurophysiology, Hanover Medical School, February - August 2007.

**Publications**

**Poster and oral presentations at International Conferences**


### Professional membership and activities

• Federation of European Neuroscience Societies (FENS)

• German Neuroscience Society (Neurowissenschaftliche Gesellschaft)


### Languages

• English: Fluent

• German: Good

• French: Basic

• Serbian: Native
Acknowledgements

First of all, I would like to thank my supervisor Prof. Reinhard Dengler, for giving me the opportunity to participate in this project and conduct my work at the Department of Neurology. I would like to thank him for his constant support, encouragement and wise leading through the obstacles of my project and my PhD studies.

I extend my appreciations to my co-supervisors, Prof. Eckart Altenmüller and Prof. Elke Zimmermann for their continuous discussion, crucial comments and scientific supervision.

I thank my colleagues from the working group Dr. Janine Born, Mathias Wittfoth, Anja Fellbrich and my doctoral student, Magdalena Lefik for their cooperation and help. I would particularly like to thank Dr. Christine Schröder for her scientific advices and contributions, and help and assistance during the writing of this manuscript.

My thanks also go to Prof. Wolfgang Baumgartner for his fair and open approach and also to Dr. Stephanie Schwab for her help in my first days in Germany. Also, a word of appreciation goes to Dr. Dagmar Esser and Mrs. Nadja Borsum for their cooperation and help during the course of PhD studies.

I would also like to thank to my colleagues from the Department of Neurology: Zornitza Nikolova, Alexandra Kotsiaris, Danela Ragancokova, Paraskevi Kotsoudaki, Jelena Skuljec, Victoria Gudi, Danai Dima and Friedrich-Emanuel Schlesinger for their collegial support and friendship.

An expression of gratitude goes to my friend Dejan and his lovely family, who generously offered me their friendship and made me feel like at home.

I give my special thanks to my parents for their unconditional love and support. I also give special thanks to my sister Ivana, for her older sister care, love and support at any
time it was needed.

Finally, I thank to my wife Marija, for her love both in the time of difficulties and happiness.

Last, but not the least, I thank to all those that I could not mention here, but who were, at least in my thoughts and all the time with me!