Neural mechanisms for fast recognition of auditory emotion

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium

(Dr. rer. nat.)

genehmigt durch die Fakultät für Naturwissenschaften

der Otto-von-Guericke-Universität Magdeburg

von Diplom-Psychologin Katja N. Spreckelmeyer

geb. am 13.04.1976 in Braunschweig

Gutachter: Prof. Dr. Thomas F. Münte

PD Dr. Sonja Kotz

eingereicht am: 25.04.2006

verteidigt am: 19.07.2006

Acknowledgments

To collect the data for my thesis I worked in three different laboratories at different places of the world. Accordingly, many people contributed to its realization, and I want to thank all of them. Above all, I would like to express my deepest gratitude to my supervisor, Professor Thomas F. Münte of the Department of Psychology, Otto-von-Guericke University Magdeburg. His sharp eye for details and superb analytical skills have been instrumental in the success of the project. I also want to thank the members of his lab who I could always turn to for advice.

Some of the data was collected during my stay at the Department of Cognitive Science, University of San Diego, California. I am deeply grateful to Professor Marta Kutas and Thomas Urbach who gave me a warm welcome, and who taught me an incredible lot about ERP research. Thanks also to the other members of the Kutas lab.

During the course of my thesis work my 'home base' was the Institute for Music-Physiology and Musicians' Medicine, Hanover School of Music and Drama. I want to thank the director, Professor Eckart Altenmüller, for his encouragement and guidance during my research. I am also grateful to past and present members of the lab, especially Dr. Michael Grossbach and Hauke Egermann. I also want to express my gratitude to Professor Hans Colonius of the Carl von Ossietzky University Oldenburg for introducing Fechnerian Scaling to me.

The work could not have been done without the generous support of the *Gottlieb-Daimler- and Karl Benz-Foundation*, the *Lienert-Foundation*, the *Deutsche Akademische Austausch Dienst*, and the *Studienstiftung des deutschen Volkes*.

I also want to thank the participants of my experiments and the musicians who played and sang the stimulus tones for me. Special thanks to Marcel Babazadeh of *Sennheiser* for professional recording of the tones.

I am very grateful to my parents for the support they provided me through my entire life. Finally, I must give immense thanks to my husband, Hanno. His love and support during the course of my thesis work was of immeasurable value to me.

Contents

1. Foreword **1**

2. Theoretical background **2**

2.1. General introduction 2

2.2. Emotion 4

2.2.1. Emotion theories 4

2.2.2. Fast stimulus evaluation 10

2.3. Acoustical expression of emotion 13

2.3.1. Acoustical correlates of vocal affect expression 13

2.3.2. Origins of vocal affect expression 16

2.3.3. Similar code usage in music and vocal expression of emotion . . . 19

2.3.4. A theoretical framework for the communication of emotions . . . 25

2.3.5. Neural correlates of auditory emotion processing 28

2.4. Summary and implications for the present study 33

3. The method of event-related potential recording **34**

I. Pre-attentive Processing of Emotional Expression in Violin Tones **39**

4. MMN-Exp. I: Are subtle changes in the emotional expression of single tones registered by the brain? **40**

4.1. Introduction . . . 40
4.1.1. Active discrimination as reflected by the P3b . . . 40
4.1.2. Pre-attentive processing as reflected by the mismatch negativity . 42
4.2. Materials and methods . . . 44
4.2.1. Participants . . . 44
4.2.2. Stimulus material . . . 45
4.2.3. Design . . . 45
4.2.4. Procedure . . . 46
4.2.5. Apparatus and recording . . . 46
4.3. Results . . . 49
4.3.1. Active condition . . . 49
4.3.2. Passive condition . . . 52
4.4. Discussion . . . 57
4.4.1. Active condition . . . 57
4.4.2. Passive condition . . . 58

5. MMN-Exp II: Are single tones categorized by the brain based on their emotional expression? 60

5.1. Introduction . . . 60
5.1.1. Aim of the study . . . 60
5.2. Scaling experiment . . . 62
5.2.1. About scaling . . . 62
5.2.2. Materials and methods . . . 63
5.2.3. Results . . . 67
5.2.4. Selection for follow-up experiment . . . 71
5.3. The MMN-study . . . 72
5.3.1. Materials and methods . . . 72
5.3.2. Results . . . 76
5.4. Discussion . . . 79

II. Processing of Vocal Emotion Expression **88**

6. Experiment II-01: Timbre as a code for emotion and identity **89**

6.1. Introduction . . . 89

6.2. Materials and methods . . . 92

6.2.1. Stimulus material . . . 92

6.2.2. Participants . . . 92

6.2.3. Design . . . 93

6.2.4. Experimental procedure . . . 94

6.2.5. Apparatus and recording . . . 95

6.3. Results . . . 98

6.3.1. ERP-experiment . . . 98

6.4. Discussion . . . 107

7. Integration of visual and auditory emotional stimuli **109**

7.1. Introduction . . . 109

7.2. Materials and Methods . . . 113

7.2.1. Stimuli . . . 113

7.2.2. Participants . . . 115

7.2.3. Task procedure . . . 115

7.2.4. ERP recording . . . 116

7.3. Results . . . 118

7.3.1. Behavioral results . . . 118

7.3.2. ERP data . . . 119

7.4. Discussion . . . 127

8. Conclusions **133**

8.1. Summary of key findings . . . 133

8.2. General discussion . . . 134

8.3. Implications for future research . . . 136

8.4. Concluding remark . . . 137

References **138**

Appendix **159**

Erklärung **164**

Lebenslauf **165**

1. Foreword

That recognition of emotion from the voice happens fast becomes obvious when your heart starts pounding the moment you hear a person next to you scream in panic. Likewise, we might know from the first word of a phone caller that there is bad news. This thesis addresses the fast recognition of emotion expressed in the auditory channel. Event-related brain potentials were recorded to examine the underlying mechanisms in the brain. The first chapter reviews the current standard of knowledge and outlines recent models serving as a theoretical framework for the presented experiments. Chapter 3 gives a short introduction into the methodology of recording event-related brain potentials (ERP). Because ERPs permit non-invasive real-time monitoring of physiological processes with a high temporal resolution, they are an ideal tool to study rapid processes in the brain. The presentation of the experiments will be divided into two parts because different levels of emotional processing were addressed. The experiments in part I (MMN-Exp. I and II) were concerned with pre-attentive classification processes of emotionally significant auditory stimuli. The experiments described in part II (II-01 and II-02) examined the time-course of cognitive processes in tasks requiring that the emotional expression was attended. Though the results of each experiment will be discussed in the accordant chapters, a brief summary and overall discussion of all four experiments will be given at the end of part II. Please note that experiment MMN-I and experiment II-02 have already been published elsewhere.

2. Theoretical background

2.1. General introduction

The communication of emotion via the auditory channel has only relatively recently come into the focus of scientific attention. Though already Darwin (1872), in his book "The expression of the emotions in man and animals", stated that *"the cause of widely different sounds being uttered under different emotions and sensations is a very obscure subject"* (Darwin, 1998/1872, p. 90), not much research was done to enlighten the subject until the nineties of the last century (see Juslin & Laukka, 2003, for a review). This long period of neglect is even more surprising given that studying the expression of emotion in the visual domain has a long tradition. One reason might be, that recording and analyzing speech sounds requires considerable technical effort. This problem has been eased, though, by the fast development of digital media technique in recent years. As a consequence, in parallel with an increased interest on emotional processing in the brain, studies have accumulated which addressed the neural basis of emotion recognition from the voice (Morris, Scott, & Dolan, 1999; Kotz et al., 2003; Bostanov & Kotchoubey, 2004; Wildgruber et al., 2005) and, though less extensively, music (Altenmüller, Schürmann, Lim, & Parlitz, 2002; Khalfa, Schön, Anton, & Liegeois-Chauvel, 2005). The main question in emotion communication research is how an emotion, intentionally or unintentionally expressed by a sender, can be decoded by a receiver. In speech, information can be perceived from two parallel channels. On the one hand, semantic information can be understood from the linguistic content of a sentence. At the same time, information about the speaker's age, gender, or emotional state can be derived

from ‘paralinguistic features’, e.g. tone of voice. The research on emotional information conveyed by a speaker’s way of speaking, termed affective prosody, thus faces the problem that it can hardly be studied independently of the semantical channel. Attempts to eliminate semantic meaning from speech samples have been made, for example, by using pseudo-words (Banse & Scherer, 1996) or by applying frequency filters such that the words become incomprehensible (Friend & Farrar, 1994). However, both methods have the disadvantage of largely reducing stimulus authenticity. In contrast, music does not require manipulation because musical meaning relies on only one channel. Also, because music can easily be broken down into structural subcomponents, it allows for systematic manipulation of different features which might play a role in the encoding of emotion. Since music is also a strong carrier of emotion it is an ideal tool to study emotion perception.

Previous studies have revealed the role of ‘dynamic’ changes of musical structure such as rhythm and tempo on the perception of a piece as e.g. happy or sad (Gabrielsson & Juslin, 1996; Peretz, Gagnon, & Bouchard, 1998). It is, however, likely that fast categorization of acoustic events relies on more ‘static’ aspects such as sound quality which can quickly be grasped by the listener. In music these aspects rely mainly on the performer, that is, the way he plays rather than what he plays. Interestingly, such aspects of ‘expressive performance’ and the paralinguistic features known to play a role in emotional expression of speech, bear a clear resemblance. For example, changing the sound of an instrumental tone from bright to dull to express sadness, sounds much like a voice changing from joyful to depressed. Indeed, several lines of arguments suggest that emotional expression in vocal speech and certain aspects of emotional expression of music have the same roots and are likely to be mediated via similar brain mechanisms. Thus, studying emotion-related sound changes in musical tones provides a way to study the mechanisms also underlying the processing of non-linguistic features of emotional speech without having to consider interactions with linguistic information.

The experiments presented in part I addressed the basic neural processing underlying the quick recognition of emotional information conveyed by auditory input. It was assumed

that the early-stage analysis of emotional stimuli can best be studied via simple auditory material. Single tones only varying with regard to their emotional expression were chosen as stimulus material for 3 reasons. First, to study fast evaluation processes of auditory input, the sound material needed to be short. Listeners have been found able to categorize single tones based on their emotional expression (Konishi, Niimi, & Imaizumi, 2000). Single tones, thus, allowed for the presentation of brief stimuli which could still be reliably classified by the listeners. Second, the emotional expression of musical tones can be manipulated without massive acoustical changes (e.g. pitch or duration). Keeping the main sound parameters stable significantly reduced the structural complexity of the auditory material. Third, the subtle changes of acoustical structure, performed to give a tone a certain emotional character, resemble the paralinguistic features of emotional speech. The results thus also have implications for the understanding of affective prosody.

The following section will give a brief introduction into general emotion theories before turning to the auditory communication of emotion in more detail.

2.2. Emotion

2.2.1. Emotion theories

Though no single universally accepted definition of emotion exists, it is generally agreed, that an emotion is a transient inner state which has been triggered by an external stimulus or an internal event (e.g. a memory or a thought). To underline its episodic character the term 'affect' is often used equivalently with 'emotion'. In the following, the two terms will be used synonymously. In contrast, the term 'mood' will be avoided because it is understood as describing a lingering state of weak intensity which is not necessarily related to a concrete stimulus event. Motor behavior, physiological and subjective-psychological reactions are generally regarded to be main aspects of emotion, frequently summarized as 'reaction triad' (e.g. Scherer, 2000). However, the question of how physiological, behavioral, and psychological aspects relate, has been a matter of centuries-long con-

troversy (Solomon, 2004). Early emotion theorists (James, 1884; Lange, 1887; Cannon, 1927) debated whether bodily reactions to a stimulus (in the form of visceral or motor responses) are a necessary prerequisite of consciously perceived emotions. William James (1884) conjectured that emotions are the result of realizing a bodily reaction (e.g. increasing heart beat, sweating) in response to a stimulus. Thus, fear would only be felt after fear-specific reactions would have been acknowledged in one's own body. In a development of the theory, Schachter and Singer (1962) suggested that a state of bodily arousal only leads to an emotion in combination with a corresponding cognitive interpretation of the situation. However, current evidence from quadriplegic patients has shown that brain activity alone can cause an emotion independent of feedback from the body (LeDoux, 1989). It is now widely accepted that emotions are the result of brain activity triggered by external or internal events. In a finer-grained dissection than the emotion-triad, five components have been identified to represent the different physiological and psychological aspects of emotion (Sokolowski, 2002): a subjective, a physiological, a behavioral, an expressive, and a cognitive component.

- *Subjective component*

  The subjective component describes the consciously perceived part of emotion which allows people to talk about how they feel. It is best assessed with rating scales.

- *Physiological component*

  The physiological component includes reactions of the peripheral body system and changes in the central nervous system. Though bodily activation has been assumed to be a necessary concomitant of emotions because it prepares the body for fast adaptive behavior, such as fight or flight (Damasio, 1999), most studies failed to link different patterns of peripheral nervous activation (resulting in changes of heart rate or electrodermal response) to specific emotions (see Cacioppo et al., 2000, for a review). In contrast, thanks to better methods and knowledge gained from patients with brain lesions, changes in the neural system that accompany emotional processing have been studied more successfully in recent years (LeDoux, 2000;

Adolphs, Tranel, & Damasio, 2003; Demaree, Everhart, Youngstrom, & Harrison, 2005; Kawasaki et al., 2005). Some of the results will be presented in more detail in section 2.3.5.

- *Behavioral component and expressive component*

  Motor responses that can be seen in conjunction with emotions have been categorized into different components ('behavioral' and 'expressive') because they are thought of as serving different functions (Sokolowski, 2002). The behavioral component describes action behavior which has the purpose to deal with a certain situation (e.g. running away from a bear or hugging a loved one). Motor activity subsumed in the expressive component includes vocal and facial expression, and serves to communicate emotions to others (e.g. to warn them, Scherer, 1988). Vocal and facial affect expression has been the object of numberless studies ever since Darwin's influential publication "The expression of the emotions in man and animal" (Darwin, 1998/1872). Findings on the vocal expression of emotion will be reviewed below.

- *Cognitive component*

  Neuroscientists (Damasio, 1999; LeDoux, 2002; Davidson, 2003) believe that emotions arise as a consequence of 'affective stimulus processing' in the brain and that emotional-processing circuits involve different brain structures than cognitive-processing circuits. However, it has been acknowledged on the basis of neurophysiological as well as behavioral evidence, that affective and cognitive processing can interact. For example, it has been shown in numerous studies that the emotional valence of a stimulus can bias cognitive processing (Pratto & John, 1991; Windmann, Daum, & Güntürkün, 2002; Davis et al., 2005). Windmann et al. (2002) found that discrimination of words and nonsense-words presented near perceptual threshold (pre-lexically) was better for sad than for neutral words. In addition, cognitive aspects of memory and expertise have been found to influence affective processing (Halberstadt, 2005, see also section 2.2.2 on appraisal).

There is little evidence for specific correlation patterns of all five components. Moreover, it seems that any of the five components can emerge alone as a consequence of emotional processing in the brain. Thus, emotional processing does not necessarily need to become conscious (LeDoux, 2002).

Another point emotion theorists disagree upon is the way emotions should be conceptualized. Whereas some authors favor the use of dimensions (e.g. Schlosberg, 1954, Russell, 1980, Watson & Tellegen, 1985), others prefer to talk of distinct categories (e.g. Ekman, 1992, Izard, 1992, Panksepp, 1998).
In the dimensional approach it is assumed that an emotional state can best be described via several independent continuous dimensions, such as valence, activation, dominance or potency. The two dimensions most researchers agree upon are valence (negative vs. positive) and arousal (excited vs. relaxed). Valence is supposed to reflect the degree to which a stimulus event makes you feel good or bad (Feldman Barrett & Russell, 1999). Davidson (1992) suggested that a basic biologically determined concept of approach and avoidance underlies the positive-negative valence dimension and that the left and the right hemisphere are unequivocally involved in the processing of approach- and avoidance-related emotions (but see Demaree et al., 2005, for a critical discussion of the latter assumption). Arousal is generally understood as the degree of activation that accompanies an emotion. Despite efforts to link self-reported arousal to physiological signs of arousal such as heart rate and skin conductance (Bradley, Greenwald, & Hamm, 1993; Lang, Bradley, & Cuthbert, 1997), the underlying neurophysiological mechanisms are still poorly understood.

In the categorical approach the existence of a certain number of mutually exclusive emotion categories has been postulated. However, the number of categories varies from 4 (Ekman, Levenson, & Friesen, 1983) to 22 (Ortony, Clore, & Collins, 1999), depending on the fineness and the methodology of categorization. In the study of emotional expression the assumption of basic emotions has found wider support than the dimensional approach (Buck, 1984; Ekman, 1992; Izard, 1992; Juslin, 2001) because evidence

accumulates that emotional expressions are perceived categorically in both the face and voice (e.g. Etcoff & Magee, 1992; Laukka, 2003). Laukka (2003) created speech examples by morphing two different prototypical expressions such that the proportion of the two emotions in a sample varied between 100% vs. 0% and 10% vs. 90%. Between all tested emotions (anger, fear, happiness, sadness) the authors found clear categorical boundaries, i.e. as soon as one emotion dominated the speech sample (e.g. 60% vs. 40%), an abrupt shift of judgment towards that emotion was seen in the majority of listeners.

The emotions that have been found to be most reliably communicated via facial and vocal non-verbal expression, even cross-culturally, are happiness, sadness, fear, anger, and, though less frequently studied, disgust and surprise (see Ekman, 1992, Juslin & Laukka, 2003, Elfenbein & Ambady, 2002, for reviews).

The dimensional and the categorical approach do not necessarily exclude each other.

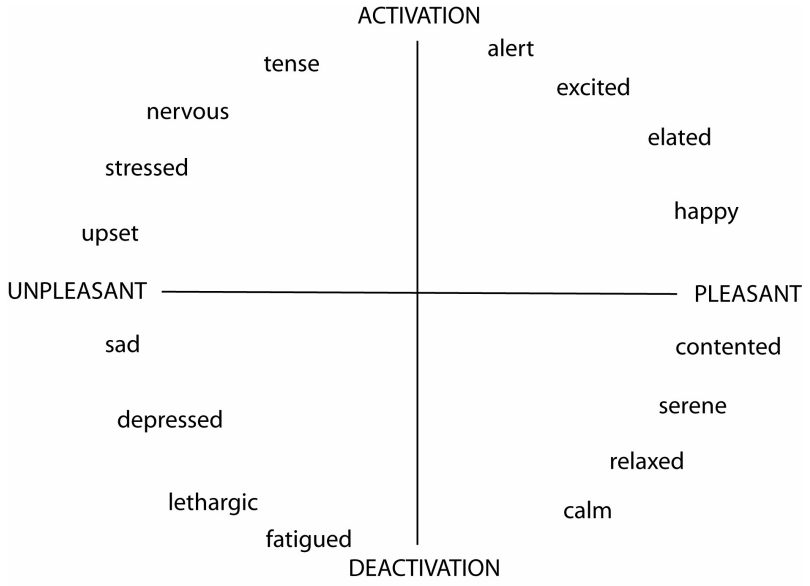

Figure 2.1.: In the circumplex model by Russell and Barrett (1999) 'core affects' can be described by two independent dimensions, degree of pleasantness and degree of activation.

Several researchers made an attempt to integrate discrete emotions into a multi-dimensional model (Roseman, Spindel, & Jose, 1990; Russell & Barrett, 1999; Juslin, 2001). For

example, Russell and Barrett (1999) suggest in their 'circumplex-model' (see Fig. 2.1) that *"core affects"* (including 'sad' and 'happy') are blends of a certain activation level and a certain feeling of pleasure. Distance from the center is interpreted as the *"intensity of a specific named affective state"* (p. 809). This integrating approach has proven particularly useful in the continuous evaluation of musical emotion because it allows gradual variations of arousal or valence level to be registered over the course of a piece within one or across different emotion categories (e.g. in the form of 'EMuJoy' by Nagel, Kopiez, Grewe, and Altenmüller, in press).

This thesis has been based on the categorical approach, in line with the majority of previous studies on emotion expression in voice and music. However, as in the hybrid model by Feldman Barrett and Russell (1999) valence and arousal were considered to be underlying dimensions in the sense that the degree of arousal may be different in different expressions of happiness.

2.2.2. Fast stimulus evaluation

A large number of emotion theorists (Ekman, 1999a; Scherer, 2001; Öhman, 1986; Zajonc, 1985; Lazarus, 1991) included in their theory an automatic 'appraisal mechanism' as part of the emotion elicitation process. It is assumed that the living organism constantly evaluates its environment in search of new stimuli which might require a fast adaptation of behavior. The appraisal process is supposed to consist of *"determining the overall significance of the stimulus event for the organism...The result of this appraisal process - the appraisal outcome - produces emotion episodes when there is sufficient evidence that the perceived significance of the appraised event requires adaptive action or internal adjustment"* (Scherer, 2001a, p. 369). There is agreement that the appraisal needs to happen fast and in many cases pre-attentively. However, how much the specific stimulus-reaction-patterns that trigger emotional reaction are hard-wired in the brain is a matter of debate. On the extreme end, Lazarus (1991) states that if a stimulus event fits certain innate criteria, an emotional (psychobiological) reaction is a mandatory consequence of the appraisal process. In a less behavioristic approach, Ekman (1999a) considers social learning as an important mediator of stimulus expectancies and resulting emotional reactions.

On the neurobiological level, the amygdala has been identified as playing a crucial role in fast and unconscious evaluation processes (LeDoux, 2002). It has been acknowledged as the core structure of an 'emotional-processing circuit' which performs the evaluation of the incoming sensory stimulus and triggers the subsequent emotional response. It is the ideal candidate because it is connected with both sensory input and motor output systems (see Fig. 2.2). Direct connections between the sensory thalamus and the amygdala (bold arrows in Fig. 2.2) allow the "*quick and dirty*" processing (LeDoux, 2002, p. 123) of significant stimulus events without previous (conscious) processing in the sensory cortex. An additional, slower processing route via the sensory cortex (dashed arrows in Fig. 2.2) provides the amygdala with a more accurate stimulus presentation[1]. The two paral-

[1]LeDoux (2002) has developed his model examplarily on 'fear'. He stresses that though the structures involved might not be absolutely identical for other emotions, the basic mechanisms are expected to be the same.

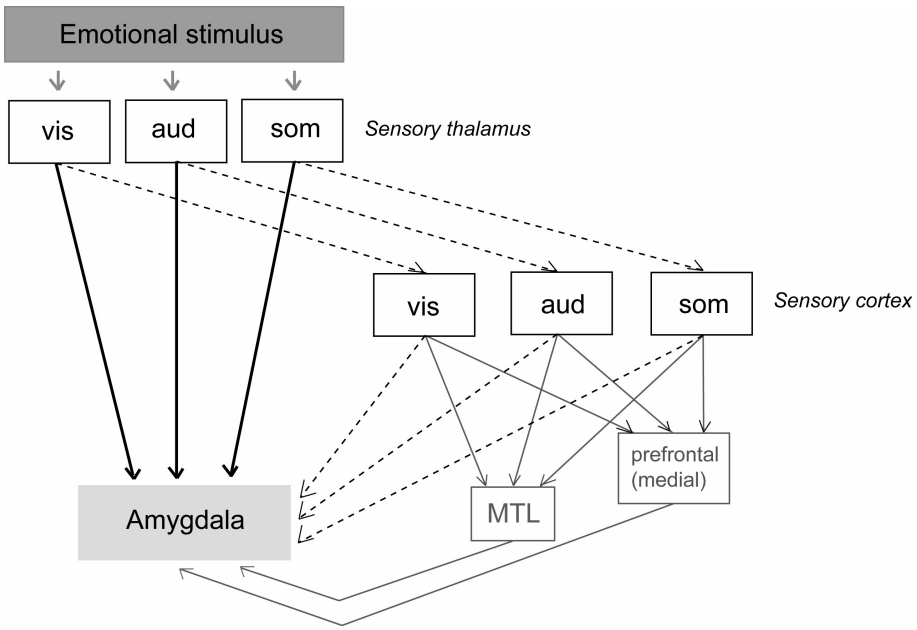

Figure 2.2.: The amygdala as core structure of an emotional processing circuit in the brain (adapted from LeDoux, 2002). Direct connections from the sensory thalamus (bold arrows) constitute the 'quick-and-dirty'-road to the amygdala; a higher processing level involves the sensory cortex (dashed arrows). Abbreviations: vis, visual; aud, auditory; som, somatosensory; MTL, medial temporal lobe system.

lel routes can be linked to different levels of appraisal as proposed by Scherer (2001). In his model of fast 'stimulus-evaluation-checks' (SECs) he distinguishes a 'sensory-motor level' from a 'schematic level'. Similar to the 'quick-and-dirty'-processing suggested by LeDoux (2002), stimulus evaluation on the motor-sensory-level is expected to be based on innate stimulus patterns and reflexive reactions. In contrast, *"on the schematic level, the schemata forming the criteria for the SECs are expected to be largely based on social learning processes, and much of the processing can be expected to occur in a fairly automatic fashion, outside of consciousness. It is likely that response integrations are stored along with the schema-eliciting criteria"* (Scherer, 2001, p. 103). Scherer (2001) and LeDoux (1991) agree that, because emotional processing needs to be fast, the higher level (and slower processing route) will only be chosen if the lower one does not suffice. Both authors also point out that processing via the higher level does not necessarily need to be conscious.

On a neural basis, aspects of social learning are likely to be mediated via the medial prefrontal cortex as well as the memory system in the medial temporal lobe ('medial prefrontal' and MTL in Fig. 2.2). In humans, the prefrontal cortex is known to play an important role in the 'top-down' regulation of perception and behavior (Miller & Cohen, 2001). Patients with prefrontal lesions have severe difficulties interpreting complex non-verbal behavior in the context of social interactions (Mah, Arnold, & Grafman, 2004; Shaw et al., 2005). Connections between the amygdala and the medio-temporal memory system allow for the integration of personal experience into the emotional-processing circuit. The hippocampus, core-structure of the medio-temporal memory system, was found to be relevant for context-learning in emotional situations. Rats with hippocampal lesions did not show freezing (i.e. typical defense behavior) in contexts that had previously been coupled with a threatening stimulus (Anagnostaras, Maren, & Fanselow, 1999).

2.3. Acoustical expression of emotion

2.3.1. Acoustical correlates of vocal affect expression

One of the most stable findings in the study of vocal affect expression is that prototypical acoustical patterns exist for a number of vocal emotions (Kotlyar & Morozov, 1976; Scherer, 1986; Scherer, Banse, Wallbott, & Goldbeck, 1991; Banse & Scherer, 1996). To test for prototypical acoustical similarities, many studies used an experimental setup known as 'standard content paradigm' (Davitz, 1964). In that paradigm, the stimulus material typically consists of the same spoken phrases expressed in different tones of emotion by the encoder (typically a professional actor or narrator). The material is recorded and evaluated in listening experiments to test how well the particular emotion is communicated to the 'decoder'. Afterwards, an acoustical analysis is performed on the stimuli that have been correctly categorized. The use of 'standard content'-stimuli allows deviations in the acoustic profiles to be linked to the emotional tone alone. Despite a heterogeneous use of stimulus material across studies (words, sentences, numbers, vowels, nonsense-words, syllables), correlating patterns were found for distinct emotions (see Juslin & Laukka, 2003, for a review). Table 2.1 summarizes the main findings.

The most frequently studied parameter in voice analysis seems to be fundamental frequency (F0). Physiologically, the fundamental frequency is modulated by the tension of the vocal folds which vibrate under sub-glottal air pressure (Sundberg, 1999). Typical findings for sadness are decreases in mean F0 level and F0 range as well as downward-directed F0 contours (Pittam & Scherer, 1993). In contrast parameters to express joy or happiness in the voice are typically an increase in mean F0 and mean intensity as well as greater F0 variability and F0 range (Pittam & Scherer, 1993). As an explanation for the inconsistent patterns for fear expression (see table 2.1), Juslin and Laukka (2003) suggested that different forms of fear might come with different vocalizations, i.e. panic-like fear is more likely to be expressed via a high pitched loud voice than e.g. a persistent creeping fear.

Voice quality

Table 2.1.: Overview of speech and voice parameters found to be relevant in the expression of certain emotions (adapted from Juslin and Laukka, 2003).

	speech rate	voice intensity	intensity variability	high-frequency energy
happiness	fast	high	high	high
sadness	slow	low	low	low
anger	fast	high	high	high
fear	fast	inconsistent	medium-high	inconsistent
	F0 (Mean)	F0 contours	F0 variability	microstructural irregularity
happiness	high	up	high	regular
sadness	low	down	low	irregular
anger	high	up	high	irregular
fear	high	up	inconsistent	irregular

Besides variations of F0-related parameters many researchers reported emotion-specific alterations which can best be described as changes in 'voice quality'. Voice quality is related to the musical term 'timbre' and shares with it the difficulty to be properly defined. Colloquially, sound specifications of both qualities are typically described with words like dark, light, dull, bright, sharp, metallic, or warm. All are attempts to describe the tonal color of a sound. But timbre is more than just an aesthetical aspect of music. It allows us to distinguish a dog's growl from a lion's growl and one friend's voice from that of another. Timbre thus plays an important role in auditory object recognition (Moore, 2004). The American Standards Association (ASA 1960, p.45) defines timbre as *"[...] that attribute of sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar"*. The reason why timbre is not described more precisely is its multi-dimensional nature (McAdams, Winsberg, Donnadieu, Soete, & Krimphoff, 1995). Apparently, timbre (and voice quality, likewise), depends on many different acoustical features. Besides temporal parameters, such as attack, the most important parameter was found to be the spectral composition of the sound (Grey, 1977; Iverson & Krumhansl, 1993; McAdams et al., 1995). Spectral composition refers to the collectivity of all frequencies present in a sound. Unless it is a pure sine tone, acous-

tical stimuli consist of a number of distinct frequencies (parameterized via cycles per second in Hertz). The lowest frequency is called the fundamental frequency F0. Higher frequencies are referred to as 'partials'. Most instruments and the human vocal tract produce harmonic sounds, which are characterized by harmonic partials with frequencies that are whole number multiples of the fundamental frequency (see Fig. 2.3). The 'spectrum' encompasses all frequencies present in a sound. Among other features of spectral composition timbre depends on the number and the relative intensity of the individual partials. Increasing the intensity of the higher harmonic partials more than that of the lower ones results in an increasing 'brightness'-perception of the sound, gradually turning into 'sharpness' (Meyer, 2004). Banse and Scherer (1996) found a high proportion of energy in the low frequencies ($<$ 1000 Hz) for vocally expressed sadness as compared to all other tested emotions. In their meta-analysis Juslin & Laukka (2003) found high-frequency energy to be increased in happiness and anger expression in both music and speech. In systematically manipulating spectral parameters of sentences spoken with different emotional expression, Lakshminarayanan et al. (2003) demonstrated that emotion recognition was degraded most if the original spectral envelopes per word were replaced by a fixed spectral pattern without changing the original pitch contour. The manipulation eliminated any subtle spectral information that had previously differentiated the speech samples. The results underline the importance of voice quality parameters for emotion recognition.

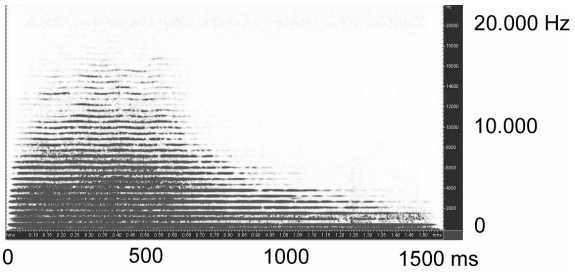

Figure 2.3.: Spectrogram of a violin tone. The tone consists of a fundamental frequency and a number of harmonic partials with higher frequencies (horizontal lines).

2.3.2. Origins of vocal affect expression

Paul Ekman (though originally coming from a dimensional approach, see Ekman, 1957), was very influential in establishing the now widely accepted notion that emotion categories can be brought down to a small number of 'basic' emotions which are universal and based on innate brain patterns (Ekman, 1999a, 1999b). From such an evolutionarist's perspective, emotions have universally evolved from adaptive reactions to prototypical basic life-situations that are common for most organisms, such as danger (fear), competition (anger), loss (sadness), and cooperation (happiness; Juslin, 2001). Strong support for the concept of basic emotions has been drawn from research on the origins of emotion vocalization. Studies in animals indicate that animals, too, express emotions via distinct sounds (Jürgens, 1979; Hauser, 1997). It has been shown that the emotional state of a calling animal can be recognized by the specific acoustic structure of certain calls. It has even been suggested that the same acoustic features are identically used by different species to communicate affective states (Owings & Morton, 1998) and that a "*phylogenetical continuity of emotional [vocal] expression*" (Juslin, 2001, p. 772) exists which finds its continuation in human communication of emotions. From the evolutionary perspective outlined above, vocal expression has developed to facilitate social interaction with the aim to increase the chance of survival (e.g. through cooperative defense, Plutchik, 1980). Several different lines of argumentation are drawn on to support the notion that acoustic patterns in human affect communication evolved from phylogenetically old (primal) communication patterns. In the following, I will briefly introduce them.

- *Evidence for innate patterns of affect vocalization*

  The fact that infant humans (Eibl-Eibesfeldt, 1973; Scheiner, Hammerschmidt, Jürgens, & Zwirner, 2005) and infant squirrel monkeys (Hammerschmidt, Freudenstein, & Jürgens, 2001) spontaneously produce affective vocalizations even when born deaf provides strong evidence that the vocal expression of emotional states is to a large extent innate. Further support comes from reports that electrical stimulation of certain structures in the brain of primates (e.g. the anterior cingu-

late cortex, periaqueductal grey, amygdala) produces vocalizations that cannot be distinguished from voluntarily uttered ones (Jürgens, Maurus, Ploog, & Winter, 1967; Jürgens & Ploog, 1970). In human patients the anterior cingulate gyrus[2] (Jürgens & Cramon, 1982a; J. F. Barrett, Pike, & Paus, 2004), too, has been found to play a major role in affect vocalization.

- *Ontological evidence*

  The assumption of 'pre-wired patterns' of emotion vocalizations gains support from developmental studies. Papoušek, Bornstein, Nuzzo, Papoušek, and Symmes (1990) demonstrated that few months old infants were well able to differentiate different emotional expressions in infant-directed speech. Infants were found to attend longer to photographs of strangers when the tone of voice was approving than when it was disapproving. This line of argument is maybe the weakest because intrauterine learning may also have shaped emotion recognition prior to birth (Moon & Fifer, 2000).

- *Parallels in the acoustic structure of human affect expression and animal calls*

  Acoustic analysis of affect vocalizations in animals and humans have revealed structural similarities (Jürgens, 2003). For example, fundamental frequency was found to increase with increasing stress (e.g. in fear) in men (Banse & Scherer, 1996) and primates (Gouzoules & Gouzoules, 1989; Schrader & Todt, 1993). Aggressiveness was found to be correlated with total pitch range and irregularity of pitch contours in both human (Scherer, 1986) and primate vocal calls (Jürgens, 1979). Several researchers (Darwin, 1998/1872; Fonagy, 1962; Sundberg, 1987; Scherer, 1995) have presumed that some acoustical aspects of emotional vocalizations are a result of the physiological changes (mainly of the autonomic nervous system) that accompany emotional arousal (see Cacioppo et al., 2000, for a review). Though this presumption seems to be derived mainly from intuition, like in Sundberg (1999, p. 210): *"we expect no rapid body movements from a truly sad*

[2]The cingulate gyrus is part of the phylogenetically old limbic system.

person, and, of course, we would not expect any rapid movements in those laryngeal structures that regulate voice fundamental frequency in that person.", it has indeed been found, for example, that increasing tension in the laryngeal muscles, results in a higher fundamental frequency of the voice (Fonagy, 1962; Johnstone, Reekum, & Scherer, 2001).

- *Cross-cultural similarities*

  As a consequence of the supposed universality of emotional expression it may be expected that voice patterns of basic emotional expressions are largely consistent across different languages and cultures. In a meta-analysis by Elfenbein and Ambady (2002), including 11 different cross-cultural studies of non-verbal vocal expression, the mean accuracy level for cross-cultural emotion recognition reached 44% after correction for chance level[3]. However, the accuracy varied considerably across national and/or cultural groups (range 11.7% to 80.4%) and the emotions were consistently better recognized if speakers and listeners were members of the same group (*"in-group-advantage"*, Elfenbein and Ambady, 2002, p. 205). For example, Scherer, Banse, and Wallbott (2001), presenting identical stimulus material (German non-words expressed with angry, sad, fearful, joyful, and neutral voice) to listeners from nine different countries, reported a lower recognition rate[4] in Indonesian listeners (39.5%) than in participants from European countries (52.0% to 61.5%) or the U.S. (59.3%). The authors (Scherer et al., 2001) relate the result to the linguistic aspects of the stimulus material (typical for Indo-European languages but not for Indonesian), which might have weakened the universal character of the affect expressions. Indeed, it has been suggested earlier that spontaneously expressed, reflexive *"Naturlaute und Empfindungslaute"*[5] (Kleinpaul, 1888/1972, p. 164) are more likely to be similar among speakers of different cultures than word-like utterances. In the same vein Wundt (1900) distinguished 'primary inter-

[3] To be able to compare accuracy percentages across studies despite varying numbers of response categories and accordingly varying chance levels, the authors subtracted the portion of the accuracy that was due to chance and rescaled the resulting score relative to 100%

[4] reported are accuracy percentages as corrected by Elfenbein and Ambady (2002)

[5] 'elemental cries and expressions of sentiment' (translation by the editor)

jections' from 'secondary interjections'.

Together these findings indicate that, though there is considerable evidence for a universal (possibly hard-wired) code in vocal affect communication, culture, too, plays a role in shaping emotional communication. Although most likely based on innate mechanisms, vocalization is modulated by social experience. Already the earliest parent-infant-interaction is likely to cause culture-specific modulation (Juslin, 2001). An entertaining example of how language and etiquette might *"domesticate"* primal affect vocalization was given by Scherer (1988, p. 82):

> *"Two people are making their first attempt to eat oysters at home, and upon opening the shell, observe a black worm slither from the oyster. One of them immediately screeches 'Eee!' [Now,] let us take the case of a person eating in a restaurant who happens to observe another diner apparently relishing a dish of oysters containing the black worms. In this situation, some people might exclaim 'Yuck'."*

Scherer (1995) introduced the concept of 'push- and pull-factors' to describe the interacting roles of biological determination and social learning in affect expression. He states that physiological conditions (e.g. muscle tone) mandatorily 'push' vocalization into a certain form, whereas external factors, such as social conventions 'pull' into a different direction. In the given example, language and culture were the pull-factors which modulated the spontaneous outcry ('Eee!') into a socially more acceptable expression ('Yuck!').

2.3.3. Similar code usage in music and vocal expression of emotion

In the music domain, a seminal series of experiments by Hevner (1935, 1936, 1937) investigated which structural features contributed to the emotional expression conveyed by a piece of music. By systematically manipulating individual factors within the same musical pieces, she came to the conclusion that tempo and mode had the largest effects on listeners' judgments, followed by pitch level, harmony and rhythm (Hevner,

1937). Slow tempos, or few beats per minute, were generally rated as sad, whereas fast tempos, or many beats per minute, tended to be rated as happy. Correspondingly, the minor mode was associated with sadness, whereas pieces written in the major mode were predominantly assessed as happy. Concerning pitch level the studies revealed that high pitch was generally associated with happiness and low pitch with sadness. Simple, consonant harmony was more likely perceived as happy than complex and dissonant harmony. Finally, flowing and varied rhythm as opposed to solemn rhythm was rated as conveying happiness and gaiety. Many studies have by now confirmed Hevner's early findings (Scherer, 1995; Gabrielsson & Juslin, 1996; Juslin, 1997a; Peretz et al., 1998). The recognition accuracy of emotion in music was found to be almost as good as in vocal expression. In a meta-analysis on 13 studies of music performance, Juslin & Laukka (2003) calculated a mean accuracy level of π =.88, compared to π =.90 for vocal expression studies, with π =.50 representing chance level[6]. Juslin (1997b), testing musically trained listeners and untrained listeners did not find differences in decoding accuracy between the two groups.

All of the musical parameters discussed so far describe changes in the structure of a musical sequence developing over a period of time such as melody, rhythm or harmony. Many parallels have been found between such *"suprasegmental features"* (Scherer & Zentner, 2001, p. 362) of musical structure and the vocal communication of emotion. For example, both, speech rate and beats per minute are higher in expressions of happiness than of sadness (Juslin & Laukka, 2003). Many researchers indeed hypothesized a close relationship in the development of music and human vocalization (Helmholtz, 1885/1954; Kivy, 1989; Scherer, 1995; Molino, 2000). Although currently there is no agreement on whether music evolved before, after, or in parallel with speech (Brown, Merker, & Wallin, 2000; Brown, 2000), findings of musical instruments dating back to the Middle Paleolithic[7] (Kunej & Turk, 2000) have nourished theories that music

[6] π-values were calculated by the authors based on a procedure by R. Rosenthal & D.B. Rubin (1989, *Psychological Bulletin*, 106, p. 332-337), to be able to compare accuracy levels despite different numbers of forced-choice categories.

[7] approx. 300,000 BC to 30,000 BC

evolved early in the development of mankind. The question of why it has evolved, though, remains a matter of debate. There is some evidence that one of music's earliest functions was to coordinate human social activity, e.g. in the form of work or war songs (Geissmann, 2000). Some authors also suggest that music evolved to enhance parent-off spring communication, which might have increased the offspring's chance of survival (Dissanayake, 2000). There is, however, little doubt that musical structure is much more affected by sociocultural conventionalization (Sloboda, 1990) than affective speech. Composition rules as well as development of different instruments and playing techniques have led to very different forms of musical expression (Gabrielsson & Juslin, 2003). Juslin (1997a, 2001) as well as Scherer & Zentner (2001) therefore suggest that only certain aspects of expressive music bear the same *"iconic signaling characteristics"* (Scherer & Zentner, 2001, p. 364) in the communication of emotion as paralinguistic aspects of speech. Scherer & Zentner (2001) suppose that the acoustical structure of *"segmental features"*, defined as *"individual sounds or tones as produced by the singing voice or specific musical instruments"* (p. 362) bear more resemblance to spontaneous natural affect vocalizations than suprasegmental features (i.e. rhythm, melody, or harmony) because they are both influenced by physiological changes accompanying affective states. The authors assume that because of their gradual transformation by centuries-long socialization, suprasegmental features might by now convey emotions primarily via symbolic coding. To correctly understand each other, the cultural-specific code must be known between performer and listener. In contrast, the appraisal of segmental features is supposed to be based on innate symbolic representations which have emerged from the same evolutionarily evolved mechanisms as required for the evaluation of vocal expression (Scherer & Zentner, 2001). The emotion-specific modulation of segmental features is thus expected to be largely culture-independent. This hypothesis, however, still awaits empirical support.

Juslin (1997a, 2001), in line with Scherer & Zentner (2001), notes that *"the hypothesis that there is an iconic similarity between vocal expression of emotion and musical expression of emotion applies mainly to those aspects of the music that the performer*

can control during his performance" (p. 321). However, the performer's freedom to alter the emotional meaning of a piece of music varies with musical style and cultural background. In pieces with strict structure as typically found in Western classical music, the performer's options to give the performance a personal note are limited to manipulating articulation and sound-aspects of individual tones. Articulation is defined as the proportion of sound to silence in successive notes (Juslin & Laukka, 2003). The relative portion of silence was found reduced ('legato'-way of playing) in sad pieces and increased ('staccato') when happiness or fear were expressed (Gabrielsson & Juslin, 1996). Factors that have been identified to contribute to the emotional expression of single tones (Rapoport, 1996; Juslin, 1997a; Juslin & Laukka, 2003) are

(1) timbre,

(2) attack,

(3) mean pitch,

(4) pitch contour,

(5) vibrato, and

(6) sound level.

Table 2.2 summarizes the main findings how happiness, sadness, fear, and anger are commonly expressed on single tones, based on a review by Juslin & Laukka (2003).

Table 2.2.: Overview of the most important performance related aspects for segmental features (i.e. individual tones) in studies on musical expression of emotion (based on review by Juslin and Laukka, 2003). Abbreviations: incon.=inconsistent, frequ.=frequency, en.=energy.

	high-frequ. en.	attack	mean pitch	pitch contour	vibrato	sound level
happiness	medium	fast	sharp	up	incon.	medium-high
sadness	low	slow	flat	down	small	low
anger	high	fast	sharp	down	large	high
fear	–	incon.	sharp	down	incon.	low

Timbre

As has been outlined in section 2.3.1, timbre can be described via the proportion of high and low frequency energy. Different instruments are characterized by different timbres as a consequence of different materials and shapes. In addition the performer may alter the timbre of a tone by varying embouchure or bow pressure. Musicians have been found to encode sadness with dull timbre which results if the relative portion of high frequency energy is low. Bright timbre was found in happy tones (medium amount of energy in high frequencies), and sharp timbre (high amount of energy in high frequencies) in anger. Parallels can thus be drawn to spectral composition patterns in speech-related vocal affect expression.

Attack

A musical sound event can be described via its tone envelope (or waveform), i.e. the development of amplitude over time. The envelope can be split up into three parts: onset, a sustained middle part, and offset (McAdams, 1993). The onset portion is called attack and can be described via the time from absolute tone onset to the point of maximum amplitude. Tones in sad sequences are characterized by a slow tone attack. In contrast tone sequences categorized as happy or angry mostly consist of tones with a fast attack. Attack was less frequently studied in affective speech and yielded heterogeneous results (see Juslin & Laukka, 2003). In speech material attack is of course also dependent on linguistic aspects of the stimulus material, e.g. the structure of the initial letter.

Mean pitch

Mean pitch, which is typically defined by the F0 of a tone, was found to encode different emotions depending on how well it matched the frequency of the intended pitch level (intonation). In sadly expressed tones the intonation has a tendency to be lower than it should be (flat), whereas in happy tones it is precise or even above supposed level (sharp). In affective prosody, too, sadly spoken phrases had a lower F0 than happy phrases (see table 2.1).

Pitch contour

Pitch contour also refers to intonation and can be linked to F0 contour in affective speech. It describes how well the pitch of a note is maintained for the duration of the tone. Pitch contour is typically found to go down in sad tones and to go up in all other emotions.

Sound level

Consistent with findings in spoken vocal expression, sound level was consistently reported to be low in sad and fearful musical expressions, medium to high in happy tones, and high when anger was expressed.

Vibrato

Vibrato is a parameter specific to music. It is defined as periodic frequency modulation which is often accompanied by amplitude modulation (Rapoport, 1996). It can be parameterized via magnitude and rate. Most emotion-related findings on vibrato are heterogeneous. The reason may be that vibrato is much more a stylistic device than a 'natural' aspect of tone production. How much vibrato is used is very much a matter of personal taste and underlies historical trends in listener preferences (Gärtner, 1974). Nonetheless, relatively consistent findings were reported for sad music which is characterized by slow vibrato with a small amplitude.

Thus, indeed, parallels can be found between segmental features of emotionally expressive music performance and paralinguistic features of vocal emotion expression. It seems that a common code exists to communicate basic emotions. However, one point that has been pondered in both realms is the fact that musicians and speakers use different cue combinations to encode the same emotion. Moreover, single cues usually do not have a mutually exclusive meaning, e.g. increased intensity is used to encode happiness and anger likewise. A model which accounts for these aspects of communication of emotions was developed by Juslin (1997b, 2001) and will be introduced in the next section.

2.3.4. A theoretical framework for the communication of emotions

The model by Juslin (1997b, 2001) is an adaptation of the lens model by Egon Brunswik. Brunswik (1955) introduced the theory of *'probabilistic functionalism'*. In the functionalist perspective, all aspects of human behavior are constantly shaped to increase their survival value. Behavior is seen as the result of interacting genetic and environmental forces (Juslin, 1997b). Brunswik's notion was that the environment that an organism comes into contact with is uncertain and probabilistic. He inferred that to be able to cope with this uncertainty an organism needs a probability-based strategy for survival. Adapted to communication this would mean that the code to express a certain emotion consists of several parallel cues which are partially redundant but not perfectly reliable. As is known from information theory (Mansuripur, 1987), coding the same information in a redundant fashion reduces uncertainty and *"yield[s] a robust communication system that is forgiving of deviations from optimal code usage"* (Juslin, 2001, p. 802). However, there are two sides to the same story. Though through redundant channeling information becomes relatively invulnerable to disturbances, it is also relatively imprecise (Mansuripur, 1987). But, as many authors favoring the basic-emotions approach have pointed out (Panksepp, 1998; Ekman, 1999a; Juslin & Laukka, 2003), in terms of survival, it is more important to make quick inferences based on broad categories than to be able to make subtle discriminations. Brunswik's model (Fig. 2.4, top) is particularly suitable to explain the communication of emotion in vocal and musical expression because it has been found in numerous studies that encoders use a large number of different cues to express the same emotions which are nonetheless understood by the decoder (Bezooijen, 1984; Scherer, 1982a). In the model, 'ecological validity' describes the relationship between an object (the "distal variable") and the cues characterizing it ("proximal cues" X1, X2, X3 etc.) which may be deciphered by an organism. The 'functional validity' is a measure of how much a cue is used by the decoder, i.e. the perceiving organism. Figure 2.4, bottom, depicts Juslin's adaptation of Brunswik's lens-model to explain probabilistic coding of emotion in musical performance. The success

of a communication process between encoding and decoding entity is expressed in the 'functional achievement' value. It is suggested that the performing artist can make use of different combinations of expressive cues. This aspect of the model provides an explanation for the fact that musicians can succesfully communicate emotions to listeners despite different musical styles and instruments. Expressive cues for which functional validity could be shown are tempo, sound level, articulation, attack and timbre (see Juslin & Laukka, 2003, for a review).

In this thesis a close-up view of the modified lens-model was applied by assuming that different features of individual tones are treated as separate expressive cues. Microstructural variations were expected to have a functional validity in that they allow for the attribution of a certain emotion to a performance.

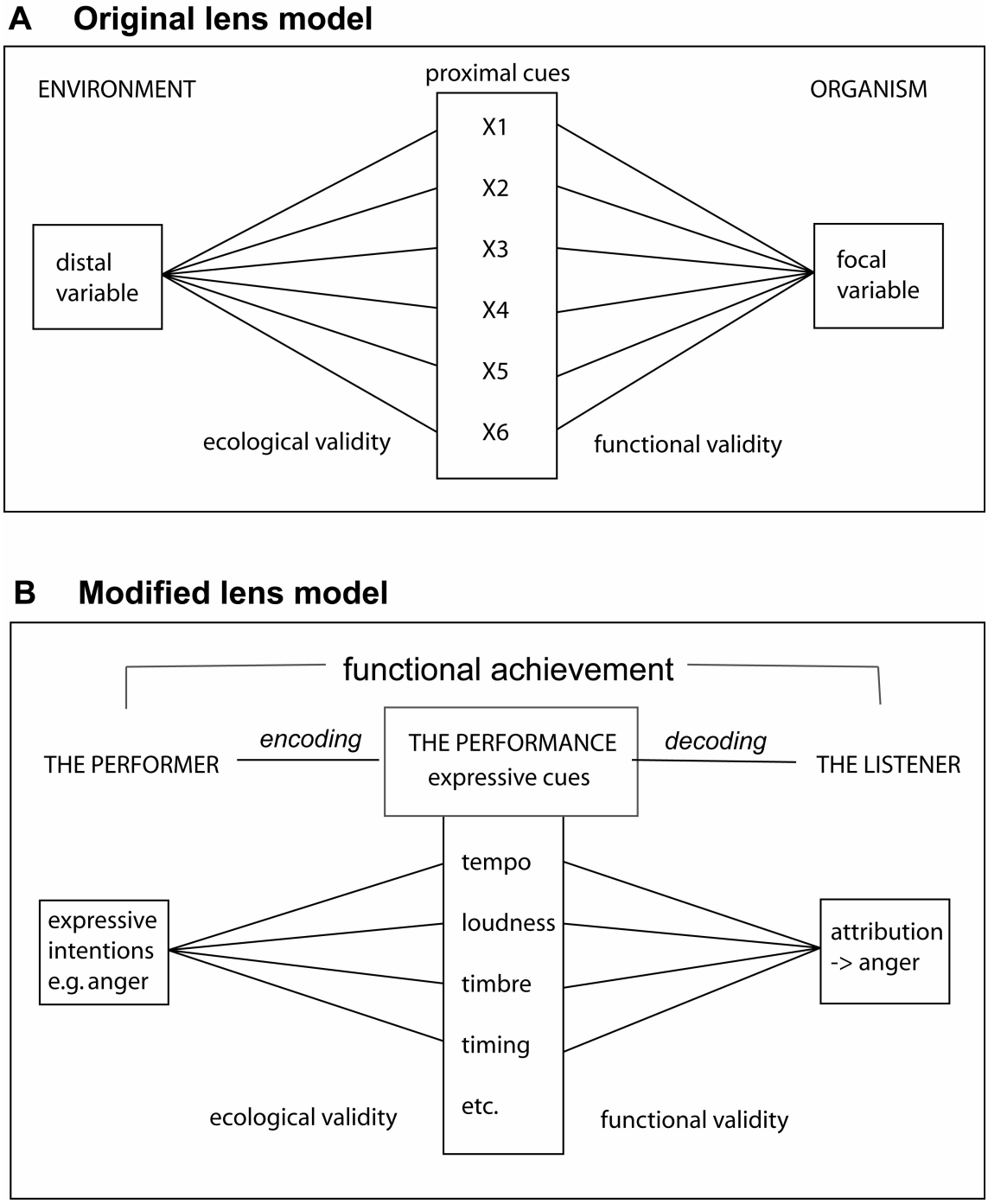

Figure 2.4.: (A) The lens model for behavior as originally introduced by Brunswik (1955). (B) The modified lens model by Juslin (1997b) for musical expression of emotion. In both models, ecological validity stands for the potential usefulness of a cue, whereas functional validity reflects how much the cue is really used by the decoder. Both models were adapted from Juslin (1997b).

2.3.5. Neural correlates of auditory emotion processing

The comprehension of the semantic content of an utterance and the understanding of the emotional expression conveyed by paralinguistic features can be selectively impaired in the sense of a double dissociation (Barrett, Crucian, Raymer, & Heilman, 1999). It has thus been inferred that both abilities are mediated by different brain structures. However, clinical trials as well as brain imaging studies have yielded heterogeneous findings as to where exactly in the brain emotion recognition takes place. The traditional view of a right-hemispheric specialization for the processing of emotional information in a speaker's voice (Ross, 1981; Ross, Edmondson, Seibert, & Homan, 1988; Blonder, Bowers, & Heilman, 1991) has increasingly been challenged. Though some of more recent brain imaging-studies have also found stronger activation in the right than in the left hemisphere in emotion recognition tasks (George et al., 1996; Buchanan et al., 2000; Pihan, Altenmüller, Hertrich, & Ackermann, 2000; Wildgruber et al., 2005), many others (Morris et al., 1999; Kotz et al., 2003; Schirmer & Kotz, 2003; Wildgruber et al., 2004; Ethofer et al., 2006) reported equal involvement of both hemispheres. Structures that have frequently been linked to vocal expression processing are the *sulcus temporalis superior* and adjacent regions (Schirmer & Kotz, 2003; Wildgruber et al., 2005; Ethofer et al., 2006), the *inferior frontal gyrus* (George et al., 1996; Buchanan et al., 2000; Wildgruber et al., 2004, 2005), and the *orbito-frontal cortex* (Wildgruber, Pihan, Ackermann, Erb, & Grodd, 2002; Wildgruber et al., 2004, 2005). Moreover, clinical studies reporting disturbed perception of emotional prosody in Parkinson patients, suggest an additional role of subcortical structures such as the *basal-ganglia* (Breitenstein, Lancker, Daum, & Waters, 2001; Pell & Leonard, 2003). The role of the *amygdala* in the auditory recognition of emotions in vocal sounds is unclear. It was found relevant in processing of non-verbal affect expressions such as screams, laughing, and crying (Scott et al., 1997; Sander, Brechmann, & Scheich, 2003; Sander & Scheich, 2005) and speech prosody (Scott et al., 1997; Morris et al., 1999). But bilateral damage to the *amygdala* yielded an impairment in the recognition of fear in faces, while it spared the recognition of fearful voices (Anderson & Phelps, 1998; Adolphs, Tranel, & Damasio, 2001). Gosselin et al.

(2005) found that patients with amygdala damage showed disturbed perception of scary music but not happy or sad music.

To account for the differing results of hemispheric involvement in auditory perception, a number of alternative models have been introduced (Zatorre, Belin, & Penhune, 2002; Poeppel, 2003; Friederici & Alter, 2004). All models take into account cumulative evidence of a functional hemispheric specialization of the auditory cortical areas on a low level of auditory processing, showing that spectral features of sound are predominantly processed in the right hemisphere and temporal features in the left (Zatorre & Belin, 2001; Boemio, Fromm, Braun, & Poeppel, 2005; Schönwiesner, Rübsamen, & Cramon, 2005). Another important point concerns the temporal development of the perception process. Several models suggest the perception of speech and music to be a multi-staged process (Altenmüller, 2003; Poeppel, 2003; Schirmer & Kotz, 2006). Poeppel (2003) has proposed, and provided evidence (Boemio et al., 2005), that auditory processing at an early representational level happens bilaterally, but that follow-up processing on a higher cognitive level is differently mediated by the right and the left hemisphere. Based on fMRI-findings[8] Boemio et al. (2005) distinguished a short (25-50 ms) and a long time window (200-300 ms), differently used by the auditory cortex to integrate auditory information. Slowly and rapidly varying narrow-band noise segments differently activated the right and the left *sulcus temporalis superior*. Their finding supports earlier statements that perceptual processing of fast changes, e.g. in speech, are likely to be processed on a different time scale than slow pitch changes typical for prosody (e.g. Rosen, 1992). In his model of 'asymmetric sampling in time' Poeppel (2003) assumes that information integrated in the short window triggers processing in the left hemisphere, whereas information from the long integration window is further processed in the right hemisphere. On the basis of Poeppel's model Schirmer & Kotz (2006), developed a three-stage model for the processing of emotional prosody (see Fig. 2.5). It is assumed that a level of early sensory processing, performed by the auditory processing areas in the temporal lobe, is followed by a phase of integration of emotionally significant cues. The integration

[8] FMRI (functional magnet resonance imaging) makes use of oxygen-consumption in the brain to track brain activation.

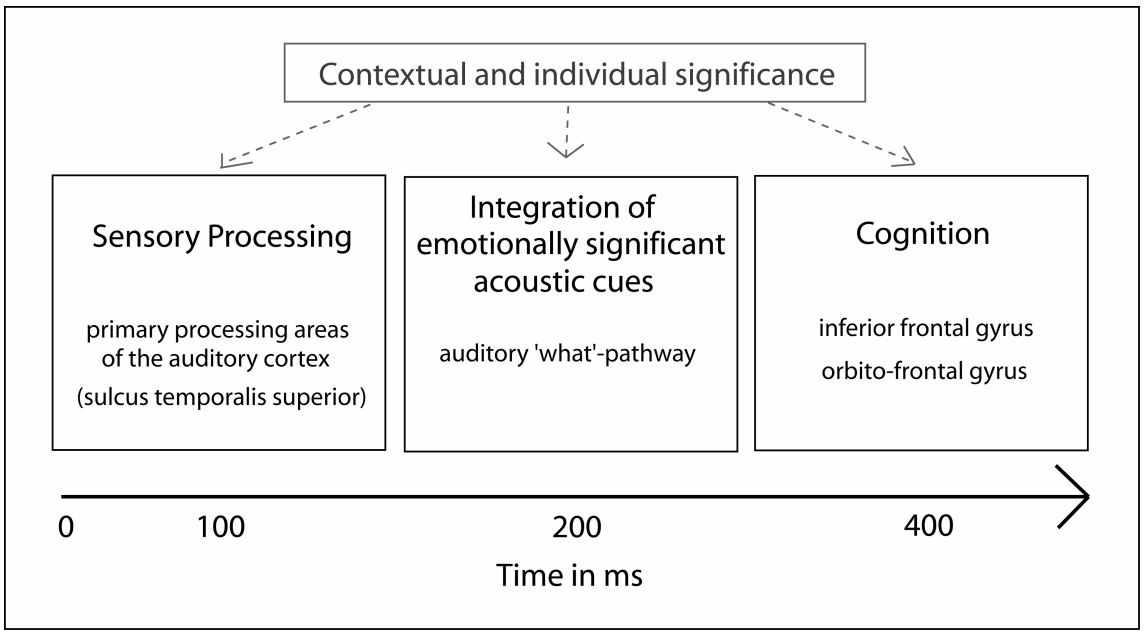

Figure 2.5.: Simplified adaptation of the model for the processing of emotional prosody by Schirmer & Kotz, 2006, p. 25. Three consecutive processing steps are expected to differ with regard to timing, function, and brain structures involved. Contextual and individual significance may modulate any of the three levels.

process requires the recognition of specific acoustic-feature-combinations. It is expected to be a function of the 'what'-pathway in audition. In analogy with visual perception (e.g. Vaina, 1994) the 'what'-pathway has been suggested to perform auditory object-recognition in contrast to the 'where'-pathway, relevant for object-localization. Areas constituting the auditory 'what'-pathway include parts of the *sulcus temporalis superior* and the *gyrus temporalis superior* (Tian, Reser, Durham, Kustov, & Rauschecker, 2001; Kraus & Nicol, 2005). The upper bank of the *sulcus temporalis superior* has been found to be especially sensitive to acoustic patterns characteristic of the human voice (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Belin & Zatorre, 2003). The third level is expected to perform conscious emotional judgment. Schirmer & Kotz (2006) point out that cortical structures found active in emotional processing other than the temporal gyrus, i.e. the *inferior frontal gyrus* and *orbito-frontal gyrus*, are likely to reflect task-dependent cognitive aspects of the evaluation process, e.g. attaching a verbal label to the perceived expression. It is expected that processing on all three levels can be influenced by context

and personal experience.

The model makes an important point by stating that sensory processing happens prior to evaluative judgment. It thus abolishes the idea that one circumscribed structure exists in the brain which is automatically activated by emotional speech and music.

It has to be noted that the studies on emotion expression in audition so far have lacked a differentiated view on the sensory aspects of encoding emotion. Moreover, most studies on emotional prosody merely considered F0 contour as an important cue, failing to control for additional aspects such as voice quality or timing (Morris et al., 1999; Wildgruber et al., 2005; Ethofer et al., 2006). Since there is little controversy that different aspects of sound, e.g. pitch, timbre, and rhythm, are processed in different brain areas (Peretz, 1990; Liegeois-Chauvel, Peretz, Babai, Laguitton, & Chauvel, 1998; Warren, Uppenkamp, Patterson, & Griffiths, 2003), it seems even more important to study emotional expression in audition more systematically. Van Lancker & Sidtis (1992) tested patients with left-hemisphere damage (LHD) and patients with right-hemispheric damage (RHD) on their ability to understand emotional prosody. Given that pitch-processing is regarded as a function of the right hemisphere (Liegeois-Chauvel et al., 1998; Zatorre, Bouffard, & Belin, 2004), the authors expected that RHD-patients would perform worse than LHD-patients because of disturbed processing of F0-contour. They found, however, that though relatively undisturbed in perceiving F0-contour, LHD-patients were as impaired in their ability to recognize emotional prosody as RHD-patients. The authors concluded that cues other than F0-contour also play a role in emotional prosody and that their processing relies on other brain structures. Thus, a more fine-grained approach to study emotional expression in audition is required. It would probably also help to gain a more consistent pattern of activation in imaging studies.

As a consequence of the assumed relation between emotional expression in speech and music, it is likely that the model by Schirmer & Kotz (2006) (Fig. 2.5) also applies to expressive performance in music. The evidence, however, is sparse. Studying the neural basis of emotion in music has mainly focused on emotion experience induced by music (Blood, Zatorre, Bermudez, & Evans, 1999; Blood & Zatorre, 2001; Koelsch, Fritz,

Cramon, Müller, & Friederici, 2006) which might be a completely different matter than emotion recognition in speech or music performance (see Juslin & Laukka, 2004, for discussion). In an emotional valence judgment task of musical pieces from different genres, Altenmüller, Schürmann, Lim, and Parlitz (2002) found bilateral frontal activation patterns, as measured by direct current EEG-potentials. They reported a right hemisphere dominance for positively and a left hemisphere dominance for negatively evaluated pieces. A more recent imaging study did not replicate hemispheric differences but also found (orbito-) frontal activation in response to tempo- and mode-manipulated pieces (Khalfa et al., 2005). With regard to segmental features, no study so far exists which has addressed the neural basis of timbre-related aspects in musical expression of emotion. From lesion studies it is known that timbre processing relies mainly on the integrity of the right hemisphere (Samson, Zatorre, & Ramsay, 2002; Kohlmetz, Müller, Nager, Münte, & Altenmüller, 2003). The fact that RHD-patients are significantly impaired in the evaluation of emotional meaning of music (Kohlmetz et al., 2003), underlines the important role timbre plays in coding emotion. Brain imaging studies confirmed the dominance of the right hemisphere in processing spectral aspects of timbre (Koelsch et al., 2002; Zatorre et al., 2004).

To conclude, the neural basis of processing emotion in speech and music is not yet unveiled. Recent findings from imaging data and lesion studies implicate that a number of brain structures is involved, which to different extents contribute to the analysis, integration, and interpretation of different acoustical features bearing emotional significance. To disentangle the different stages of emotional processing, studies are needed which specifically address individual aspects of the evaluation process.

To study the early aspects of emotional processing this thesis used very simple stimuli by presenting single tones merely differing with regard to emotional expression. Consequently, the number of acoustic cues under study was limited. The use of musical sounds allowed for systematically controlling many acoustic aspects (e.g. pitch, duration), and prevented interference of semantic and phonological aspects typical for speech. Because of the parallels of emotional code usage in musical timbre and voice, the results also have implications for understanding vocal expression in speech.

2.4. Summary and implications for the present study

Several assumptions can be derived from the theoretical framework outlined above.

1. Vocal expression of emotion has evolved from primal affect vocalization in animals.
2. A small number of basic emotions including happiness, sadness, anger, and fear are universal and relatively independent of social learning.
3. Encoding and decoding of vocal expressions of basic emotions is universal and based on innate or highly overlearned brain patterns.
4. Evaluation of emotionally significant cues happens fast and automatically.
5. Encoding of emotion in segmental features of musical performance (individual tones) bears a resemblance to the code used in vocal expression of emotion.

It can thus be hypothesized that evaluation of emotional expressive tones, too, happens fast and automatically based on innate brain patterns. The aim of the experiments described in part I of the thesis (MMN-exp I & II) was to test for a neural basis of categorical processing of emotionally expressive musical tones based on prototypical acoustic features.

In part II, emotional expressivity of vocal music was used to examine the role of timbre in decoding emotion and speaker identity (exp II-01) and to study the multi-sensory integration of emotion in audition and vision (exp II-02). The hypotheses for the individual experiments in part I and II will be presented at appropriate places.

In all experiments, happiness and sadness were chosen as the emotions under study because they are regarded as basic emotions and are thus most likely to be communicated based on an innate code (Juslin, 1997b). In addition, happy and sad are found to be hardly ever confused in recognition of vocal and musical affect (Juslin & Laukka, 2003).

3. The method of event-related potential recording

Event-related potentials (ERP) allow for the study of rapid changes of cortical activation in response to an external stimulus event. The physiological basis is formed by synchronized activity of large neuron populations in response to meaningful events (Silva, 1991). It is assumed that electrical activity can only be measured outside the head if the electrical fields of many individual neurons oriented in the same direction sum up to a dipolar field with opposite charges (positive-negative), producing a strong voltage. Based on this assumption, the signal is expected to stem from pyramidal cell activity, since pyramidal cells make up 75% of the cortex and have a densly packed parallel orientation, vertical to the surface of the skull (Münte, Urbach, Düzel, & Kutas, 2000). Indeed, simultaneous recording of membrane potentials at the pyramidal cells of a rat's cortex and field potentials at an electrode placed on the head, reveal similar potential fluctuations (see Fig. 3.1). However, event-related voltage changes which can be recorded at the scalp of human subjects are very small (in the order of microvolts). During recording they are superposed by the large-scale permanent voltage variation reflected in the electro-encephalogram (EEG). The event-related potential to a distinct stimulus event can be extracted from the background EEG through averaging – a common technique of signal extraction (Glaser & Ruchkin, 1996). On the basis that the ERP-waveforms to stimuli of a certain type have been found to have a relatively fixed form and latency, and that, in contrast, unrelated brain activity can be regarded as random noise, signal quality (i.e. the signal-to-noise-ratio, SNR) increases with an increasing number of stimulus presentation, because the random EEG-activity approximately averages to zero (Coles & Rugg, 1996).

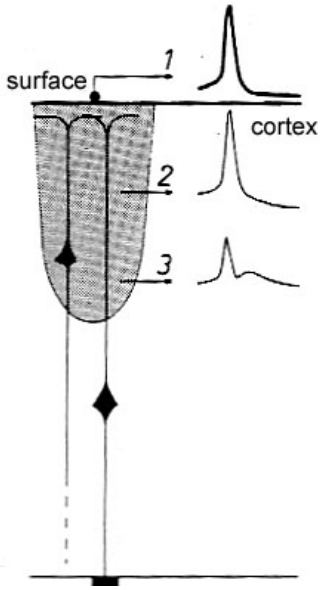

Figure 3.1.: Generation of EEG signal: simultaneous recording of membrane potentials in the cortex (2 and 3) and field potentials at an electrode placed on the head (1) reveal similar potential fluctuations. (Adapted from Speckmann & Elger, 2005.)

ERPs are typically obtained by measuring the difference in voltage between electrodes positioned on the head and a 'reference' electrode placed at a site which is relatively uninfluenced by neural activity ('referential montage', Picton, Bentin, et al., 2000). Common reference sites are the mastoid bones behind the ears (either alone or linked), as well as the nose, though other sites have been tried out (see Dien, 1998, for review and discussion). The recording system consists of an amplifier and an analog-to-digital (A/D) converter. In modern systems, only the signal difference between the scalp electrodes and the reference electrode is amplified, while the rest of the signal is canceled. A/D-conversion is necessary to allow further digital processing. The higher the number of sampling points per second (sampling rate in Hz) the higher the resolution of the digital signal. In ERP-recordings the sampling rate should be at least twice the highest frequency in the measured signal (Picton, Bentin, et al., 2000).

To further increase SNR of the ERP signal, group averages are typically calculated over all participants' data, resulting in a 'grand average'. In the experiments presented in this thesis, ERP data of all participants were normalized prior to averaging as suggested

by McCarthy and Wood (1985). McCarthy and Wood (1985) have shown that a fixed dipole source in a spherical volume conductor (such as the head) can mistakingly yield the impression of topographical differences between two conditions though it only varies in strength but not in position. The normalization procedure applied to the data in this thesis is a common technique to reduce the error. Averaging the EEG over time

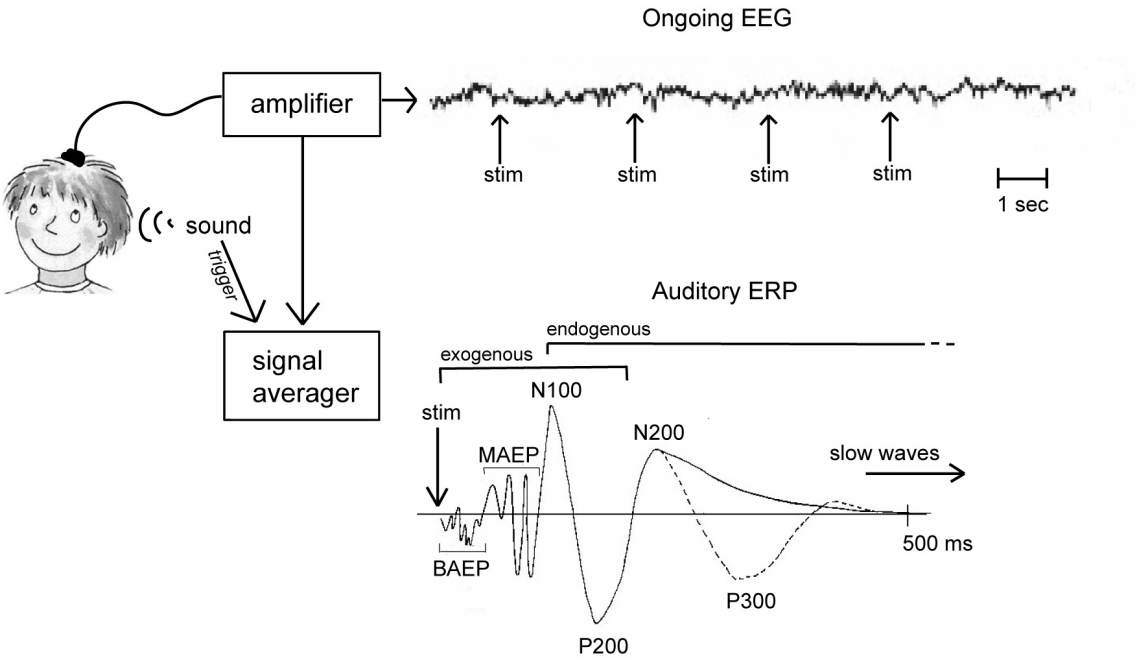

Figure 3.2.: An auditory event-related potential (idealized waveform) is retrieved by averaging over several stimulus presentations. A trigger from the stimulus sound is integrated in the averaging process to indicate stimulus onset. Abbreviations: BAEP=brain stem auditory evoked potentials, MAEP=midlatency auditory evoked potentials, stim=stimulus.

windows locked to a certain stimulus (or different stimuli of the same type) results in an ERP-waveform which consists of characteristic peaks and troughs (see Fig. 3.2). These deflections can be described via their polarity (positive or negative), their time latency following stimulus onset, and their topographical distribution. Based on numerous ERP studies certain peaks (and troughs) have been identified as 'components'. Components can be linked to distinct stimulus classes and/or cognitive processes and are characterized by relatively stable timing and topography. Component names are typically composed of the letter P or N to indicate polarity (positive, negative) and the characteristic peak la-

tency in ms (sometimes abbreviated, e.g. 'N1' for 'N100'). However, component latency may vary significantly and does not always match the number given in the component name. To avoid misinterpretation, some authors prefer to label the components according to the order of appearance. Fig. 3.2 (bottom) depicts ERP-components following auditory stimulation. Auditory evoked potentials with latencies below 10 ms (BAEP) reflect transfer of auditory information in the relay stations of the brain stem, and have great value in clinical diagnostics (Celesia & Brigell, 2005). Like the middle latency potentials (MAEP) they depend largely on physical features of the stimulus (e.g. loudness), and are often termed 'exogenous' components (Altenmüller, Münte, & Gerloff, 2005). In contrast, later components are termed 'endogenous', to underline that they reflect mental processes of the perceiver rather than stimulus characteristics. It has to be noted though, that not all authors draw the same line to distinguish between exogenous and endogenous components. While some consider the N1-P2-complex exogenous (e.g. Rugg & Coles, 1995), stressing its sensitivity to physical stimulus characteristics (e.g. Elfner, Gustafson, and Williams, 1976), others draw the line at 100 ms, regarding N1 and P2 as endogenous components (e.g. Altenmüller et al., 2005). Indeed, findings that N1 and P2 amplitude are strongly modulated by attention (Woldorff & Hillyard, 1991) have increasingly softened the distinction between endogenous and exogenous.

The general approach to examine cognitive processes triggered by external events is to directly compare ERPs recorded under different levels of manipulation. Manipulations can pertain to the stimulus material (e.g. high vs. low tones), presentation (e.g. stimulation of right ear or left ear), state of the perceiver (e.g. alert vs. relaxed), or task (e.g. passive listening vs. counting tones). Parameters commonly used for the statistical analysis of ERP differences under different manipulations are peak latency and amplitude (either peak-to-peak or baseline-to-peak). Finding the exact point in time of maximum peak amplitude can cause a problem because the time point when the maximum is reached can be different across electrodes and conditions. To increase reliability, a 'mean amplitude' is commonly calculated by averaging successive data points within a pre-defined time window (Münte et al., 2000). The time windows used in the present

thesis to calculate mean amplitudes will be reported at the appropriate places.

Artefacts in ERP-recording Electrophysiological data are very susceptible to artefacts, i.e. the EEG contains signals which do not stem from brain activity. Artefacts can either be technical or physiological in nature. Technical artefacts can result from poor recording equipment such as damaged electrodes, or disturbance through electromagnetic devices or metal near the subject's head. Physiological artefacts can derive from the subject's heart rate, muscle activity, sweating, or, most commonly, eye movements. Because the eye is a dipole, eye movements strongly interfere with the surface recording of the electrical activity of the brain (Gratton, 1998). In fact, the electrical signal recorded from the eyes (electro-oculogram, EOG) can be several times larger than the brain-generated scalp potentials (i.e. several hundred microvolts in amplitude).
Many artefacts can be kept out of the ERP-data by applying filters at different frequencies during amplification of the recorded EEG-signal prior to converting it from analog to digital (Münte et al., 2000). Filtering allows for the attenuation of frequency bands that are not relevant for the event-related potential but might e.g. reflect muscle activity. Trials still contaminated by artefacts despite online-filtering need to be excluded from averaging in an artefact-rejection procedure, to guarantee that the averaged ERP reflects only brain activity. Thus, it is essential to record a sufficient number of trials to maintain a good signal-to-noise-ratio.

Part I.

Pre-attentive Processing of Emotional Expression in Violin Tones

4. MMN-Exp. I: Are subtle changes in the emotional expression of single tones registered by the brain?

4.1. Introduction

Based on empirical findings and theoretical considerations it is assumed that tones expressing prototypical emotions (happiness and sadness) are registered by the brain for their potential significance. The aim of the first study was to test if the brain's tools for deviance detection are sensitive to subtle changes characterizing emotional expression in single tones. To this end event-related potentials were recorded in an active and a passive deviant detection task.

4.1.1. Active discrimination as reflected by the P3b

The P300 is a component of the event-related potential that is very sensitive to any kind of change in a stream of events. It is particularly pronounced when the deviant events are attended and task-relevant (Donchin, 1981; Donchin et al., 1984; Münte et al., 2000; Pritchard, 1981, for reviews). Moreover, it has been shown that the P300 amplitude increases with decreasing occurrence probability of the deviant (Duncan-Johnson & Donchin, 1977, 1982). It is assumed that the P300 is not a unitary component but can be broken down to several subcomponents (Johnson, 1986). Thus, the component just described as P300 is often termed P3b and distinguished from a component P3a

which is sensitive to the novelty[1] of an event and seems to reflect a switch in attention triggered by a task-irrelevant stimulus change (Schroeger, 1997). It has a more frontal distribution than the parietally focused P3b. The P3b is best demonstrated in response to task-relevant deviant stimuli within a stream of standard stimuli, a sequence known as oddball paradigm. Its onset latency varies between 300 and 600 ms. Latency and amplitude depend on the difficulty of the categorization task as well as on the task-relevance of the stimulus (Kutas, McCarthy, & Donchin, 1977; Johnson, 1986). Thus, the P3b appears to reflect stimulus evaluation and stimulus categorization processes. It has further been suggested that the underlying processes serve the updating of working memory (Donchin & Coles, 1988a, 1988b), though not everyone agrees on this interpretation (Verleger, 1988).

With respect to musical stimuli, the P3(b) amplitude[2] was found to correlate with the magnitude of pitch deviance in both, musicians and non-musicians (Tervaniemi, Just, Koelsch, Widmann, & Schroeger, 2005). P3(b) latency was found to be shorter in musicians (especially those with absolute pitch[3]) than in non-musicians in pitch discrimination (Wayman, Frisina, Walton, Hantz, & Crummer, 1992) and in instrumental timbre discrimination tasks (Crummer, Walton, Wayman, Hantz, & Frisina, 1994). The results indicate that expertise may influence context updating processes. Crummer et al. (1994) reported that P3(b) latencies (in both, musicians and non-musicians) became longer when differences between different instrumental timbres became increasingly subtle. In response to equal-pitch tones of brass instruments only differing with respect to their size (B-flat vs. F tuba), discrimination accuracy decreased and P3(b)-latencies increased compared to different string instruments (cello vs. viola) or flutes made of different material (wood vs. silver).

The P3(b)'s sensitivity to emotional valence has been demonstrated in picture process-

[1] Novelty is understood in the sense that no similar event has previously occurred in the stream of preceding events.

[2] The authors themselves used the general term P3. To omit confusion it will be called P3b here. However, to point out that the component name differs from that originally used by the authors, b will be given in brackets.

[3] Persons with absolute pitch are able to name or reproduce a tone without the need of a reference tone.

ing (Johnston, Miller, & Burleson, 1986; Keil et al., 2002) and evaluation of emotional prosody in spoken words (Twist, Squires, Spielholz, & Silverglide, 1991). Twist et al. (1991) found prolonged P3(b) latencies in response to semantic compared to prosodic deviants but did not comment on this finding. Targets in their semantic oddball-paradigm were names of colors compared to standard body parts. Target words in the prosodic oddball-paradigm (all words were names of pieces of furniture) were spoken with a rising, i.e. surprised sounding voice compared to monotonely spoken standards. Since different semantic material was used in both tasks, it cannot be ruled out that the reported differences in latency stemmed from stimulus-inherent differences in semantic and/or acoustic processing durations. To elude this problem, stimulus material in the current study was chosen such that physical differences were minimized. Musical tones were used to eliminate semantical meaning and to reduce acoustical variability to timbre features alone.
The aim of the study was to test how fast subtle changes of emotional expression can be recognized and categorized correctly and to study the timing of the underlying evaluation process via latency of the P3b-component. The result will be set into relation with the latency of pitch and instrumental timbre evaluation processes.

4.1.2. Pre-attentive processing as reflected by the mismatch negativity

To test the hypothesis that the evaluation process of emotional expression is mandatory and happens automatically, even in the absence of attention, stimulus material was also presented in a passive oddball-experiment where participants' attention was engaged in a visual attention task during auditory stimulation. To explore if a mismatch between standards and deviants was detected despite the lack of attention, ERPs were analyzed for occurrence of a 'Mismatch Negativity'.
The Mismatch Negativity (MMN) is a frontal negative wave in the event-related-potential that was first described by Näätänen (Näätänen & Michie, 1979). It is typically evoked by an auditory stimulus that differs from a train of preceding stimuli ('standards'). The MMN requires that the deviant tone has a lower probability of occurrence than the stan-

dards. The negative wave is thought to result from the mismatch between an incoming stimulus and the memory trace of the previous standard stimuli in the sensory memory (Picton, Alain, Otten, Ritter, & Achim, 2000). Because the MMN is typically elicited while listeners do not attend the auditory stimulation [e.g. during reading (Näätänen, 1992) or even while asleep (Loewy, Campbell, & Bastien, 1996)] it is assumed to reflect a pre-attentive or automatic *"deviance detection system"* (Schroeger, 1997, p. 256) of the brain. Changes in physical structure such as frequency (Sams, Paavilainen, Alho, & Näätänen, 1985), intensity (Näätänen, Paavilainen, Alho, Reinikainen, & Sams, 1987), timbre (Tervaniemi, Winkler, & Näätänen, 1997) or duration (Näätänen, Paavilainen, & Reinikainen, 1989) evoke as well a MMN-wave as variation of location (Paavilainen, Karlsson, Reinikainen, & Näätänen, 1989) or timing (Boettcher-Gandor & Ullsperger, 1992). It is thus a perfect tool to address the early, automatic stages of sound evaluation.

The deviant stimulus typically results in two negative waves (N1 and the MMN). They can best be depicted in form of a difference wave which results from subtracting the standard response from the deviant response. Amplitude and latency of both components vary according to the nature of the stimulus deviance. The onset latency of the MMN lies at approximately 150 ms for simple, physically deviant stimuli. Deouell and Bentin (1998) found that the peak latency increased with decreasing magnitude of frequency deviance. This finding indicates that the MMN-latency is an indicator of discrimination difficulty. Picton, Alain, et al. (2000) suggested that MMN-latency reflects a combination of discrimination difficulty and duration of the discrimination process itself.

Schirmer, Striano, and Friederici (2005) studied the preattentive processing of emotional expression in spoken syllables in an oddball experiment. They presented speech samples spoken either in a happy or a neutral voice to one group of participants ('happy group') and angry and neutral samples to another group ('angry group'). In the happy group the MMN had a shorter latency if the deviant was happy than if it was neutral. No latency difference was found in the angry group. The authors suggest that the results reflect

a stronger sensitivity of MMN-latency to stimulus valence than to stimulus arousal. However, since the stimulus material was not explicitly tested for either dimension, this consideration remains hypothetical.

With regard to timbre perception, Tervaniemi et al. (1997) have found a MMN in response to pure tones presented in a train of harmonically complex tones. The deviance thus consisted in the total lack of harmonic partials. That the MMN is also sensitive to changes in the particular structure of harmonic partials has been shown for different vowels (Jaramillo et al., 2001; Savela et al., 2003; Jacobsen, Schroeger, & Sussman, 2004), speakers (Titova & Näätänen, 2001), and musical instruments (Toiviainen et al., 1998; Koelsch, Wittfoth, Wolf, Müller, & Hahne, 2004). Toiviainen et al. (1998) demonstrated that the amplitude of the MMN decreased with increasing similarity between standard and deviant synthesized tones with regard to their timbre. In a parallel similarity rating it was proved that perceived similarity was a function of the relative amplitudes of the higher harmonic partials.

So far, no study has addressed the pre-attentive processing of timbre as a mediator of emotional expression in tones that were otherwise stable in pitch and instrumental timbre. The hypothesis for the passive experiment was that if, as assumed, the brain accomplishes a fast and possibly automatic check on every incoming stimulus with regard to the properties encoding its emotional significance, even subtle differences in the acoustic shape of the tone as in tones of different emotional expression would result in a mismatch negativity.

4.2. Materials and methods

4.2.1. Participants

Twelve non-musicians participated in the experiment (11 women, 20 to 36 years of age, mean=26). All participants were right-handed, neurologically healthy and had normal hearing.

4.2.2. Stimulus material

Two sets of four different tones were used. Each set consisted of one standard tone and three different deviant tones. All tones were played by a violinist and a flutist, digitally recorded, and edited to equal length (600 ms) and sound level (65 dB) using cool edit. These edited tones were rated by 10 naive listeners using a 7-point scale (-3 = very sad, 0 = neutral, $+3$ = very happy). Tones used for the experiment had a mean score of >1.7 for the happy and smaller than -1.7 for the sad conditions. In set 1, the standard tone consisted in a violin /c/ played in a happy way. This frequent 'happy standard' was combined with a rare violin /c/ played in a sad way ('sad deviant'), a rare flute /c/ played in a happy way ('instrument deviant') and a happy violin /a/ ('pitch deviant'). For set 2, the sad violin /c/ was used as a standard ('sad standard') and combined with the following deviants: happy violin /c/ ('happy deviant'), sad flute /c/ ('instrument deviant') and sad violin /a/ ('pitch deviant'). In the passive condition, two video films ("Les vacances de monsieur Hulot" and "Playtime", both by Jacques Tati) were presented to the participants with the sound turned off. In order to minimize eye movements, a small video screen (18") at a viewing distance of 130 cm was used.

4.2.3. Design

Each subject participated in two different sessions. The sessions were conducted on two different days separated by at least 1 week. Each session consisted of two consecutive blocks which differed with regard to the stimulus set used. The order of the two stimulus sets was kept stable for each participant between session 1 and 2 but was counterbalanced between subjects. In one session (active condition), participants held a joy stick in one hand and pressed a button with their index finger in response to any deviant tone. The use of the right or the left hand was counterbalanced between all participants. In the other session (passive condition), participants watched a video while the stimulus tones were played in the background. No response to the tones was required. The order of conditions (active or passive) was counterbalanced.

4.2.4. Procedure

Participants were tested individually while seated in a soundproof chamber in front of a computer screen which was replaced by a television set in the passive condition. In each condition, 2600 tones were played to the participants via loud speaker. A series of standard tones was presented, interrupted randomly by emotionally deviant, by instrument deviant, or pitch deviant stimuli. The probability of occurrence was 76.9% for the standard tone and 7.7% for each of the deviant tones. The interstimulus interval was randomized between 400 and 900 ms. No test trials were given but the first trials of each block were excluded from the analysis. Every 10 minutes, there was a short break and a longer 15-minute-break was taken between the two blocks. Each experimental block lasted about 55 minutes. One entire session lasted about two and a half hours. In the active condition participants were instructed to press a button as fast as possible in response to a deviant tone. In the passive condition, participants were instructed to watch the video carefully because they would be asked about it later. Following each block, three questions relating to the content of the film were asked by the experimenter that had to be answered by the participant. During the experiment, the participants looked at a fixation point in the center of the computer screen. In both sessions, participants were asked not to speak and to blink or move their eyes as little as possible.

4.2.5. Apparatus and recording

4.2.5.1. ERP-recording

The EEG was recorded from 30 scalp sites using tin electrodes mounted in an electrode cap based on the 10-20-system for electrode positioning of the International Federation of Clinical Neurophysiology (Klem, Luders, Jasper, & Elger, 1999) with reference electrodes placed at the left mastoid and the tip of the nose (see figure 4.1). Signals were collected using the left mastoid electrode as a reference and were re-referenced off-line to the nose electrode. Blinks and vertical eye movements were monitored by a bipolar montage using an electrode placed on the left lower orbital ridge and Fp1. Lateral eye movements were monitored by a bipolar montage using two electrodes placed on the right and left

external canthus. The eye movements were recorded in order to allow for later off-line rejection. Electrode impedance was kept below 5 kΩ for the EEG and eye movement recording. The EEG was sampled with a Brainlab system (Schwarzer, Munich). Signals were amplified and digitized with 4 ms resolution. Averages were obtained for 1024 ms epochs including a 100 ms prestimulus baseline period. Trials contaminated by eye movements or amplifier blocking within the critical time window were rejected from averaging by a computer program using individualized rejection criteria. On average, 11 % of the trials were excluded from further analysis. ERPs were quantified by mean amplitude and peak latency measures using the mean voltage of the 100 ms period preceding the onset of the stimulus as a reference. Time windows and electrode sites are specified at the appropriate places of the result section. Topographical distributions of the ERP effects were compared by ANOVA designs, with condition (emotion, timbre, pitch) and electrode site as factors. Before computing the statistics, the amplitudes were vector normalised according to the method described by McCarthy and Wood (McCarthy & Wood, 1985). The Huynh-Feldt epsilon correction (Huynh & Feldt, 1980) was used to correct for violations of the sphericity assumption.[4] Reported are the original degrees of freedom and the corrected p-values.

4.2.5.2. Reaction time recording

In the active condition, push-button response latencies were measured from sound onset, with the timeout point (the moment in time after which responses were registered as missing) set at 400 ms post stimulus offset. Timeouts and errors, i.e., wrong responses, were excluded from further analysis.

[4] If the sphericity assumption is not met the averaged F-tests overestimate the strength of the relationships. The Huynh-Feldt Epsilon is a commonly used correction formula.

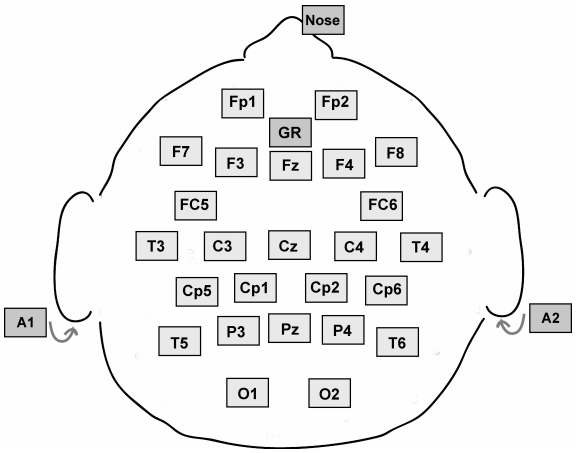

Figure 4.1.: Electrode setup used in the experiment was based on the international 10-20-system (Klem et al., 1999); F stands for frontal, C and z for central, T for temporal, P for parietal, and O for occipital, GR=ground, A1,A2=mastoid electrodes.

4.3. Results

4.3.1. Active condition

4.3.1.1. Behavior

The level of performance was nearly perfect for all deviant target stimuli (misses <1%) as well as for the standards (false alarms <1%). The reaction time was longer for emotion than for pitch or instrument deviants[5] [489 vs. 439 and 428 ms, F(2,16)=29.42, p<0.001], though the effect interacted with emotional valence [F(2,16)=123.8, p<0.001]: Differences in mean reaction times (see Table 4.1) between different types of deviants were only apparent when the standard tone was a happy tone [$F(2,16)=22.45$, $p<0.001$]. Post hoc comparison (Scheffé) revealed that in this condition, the mean reaction to the emotional deviant (sad violin tone) was slower than to the pitch deviant ($p<0.001$) and to the instrument deviant ($p<0.001$).

4.3.1.2. Electrophysiology

The peak latency was quantified in the time window between 300 and 550 ms for the Pz electrode site and subjected to ANOVA with factors 'deviant type' (emotion vs. instrument vs. pitch) and emotion (sad vs. happy). A main effect of deviant type was found [$F(2,22)=7.04$, $p<0.005$] reflecting the fact that the P3b latency was longest for the emotion deviant type (460 ms, S.D.=85 ms), followed by the instrument (402 ms, S.D.=68 ms) and the pitch deviant (383 ms, S.D.=62 ms, see Fig. 4.3). A main effect of emotion was also found [$F(1,11)=8.7$, $p<0.015$] reflecting the overall longer mean latency of sad compared to happy deviants (369 ms, S.D.=81 ms, vs. 441 ms, S.D.=81 ms). However, as can be seen in Fig. 4.4, the P3b peaked much earlier for the happy deviant than for the sad deviant only in the emotion condition (significant deviant type x emotion interaction [$F(2,22)=8.02$, $p<0.005$]).

[5] Due to technical failure reaction time recordings of 3 participants were lost.

Table 4.1.: Reaction times (ms) to deviant stimuli in the active experiment

	Block I standard happy			Block II standard sad		
	Emotion (sad)	Instrument (happy)	Pitch (happy)	Emotion (happy)	Instrument (sad)	Pitch (sad)
Mean (*N*=*9*)	527	383	406	449	472	470
S.D.	107	93	115	104	107	118

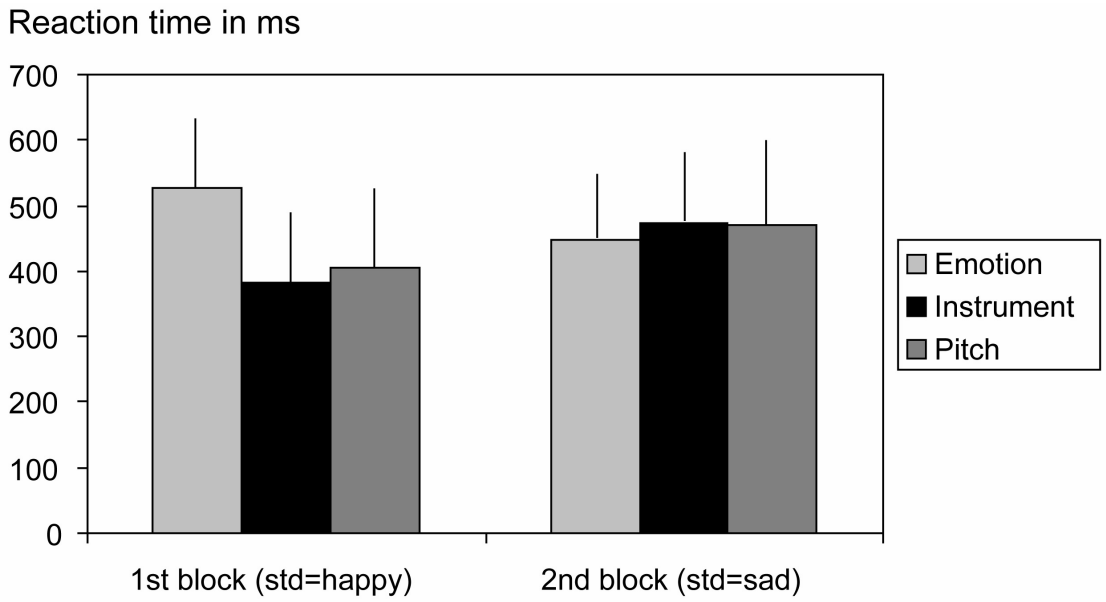

Figure 4.2.: Reaction times in the active condition. The emotional deviant in block 1 was sad, while the instrument and the pitch deviant were happy. In block 2 the emotional deviant was happy while the instrument and the pitch deviant were sad. Error bars reflect standard deviations.

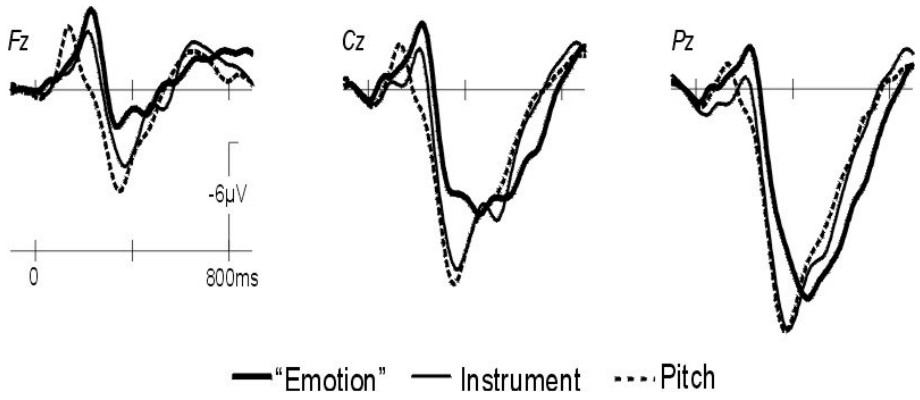

Figure 4.3.: ERPs from the active experiment at three electrodes (Fz, Cz, Pz) for all three target types, collapsed over happy and sad.

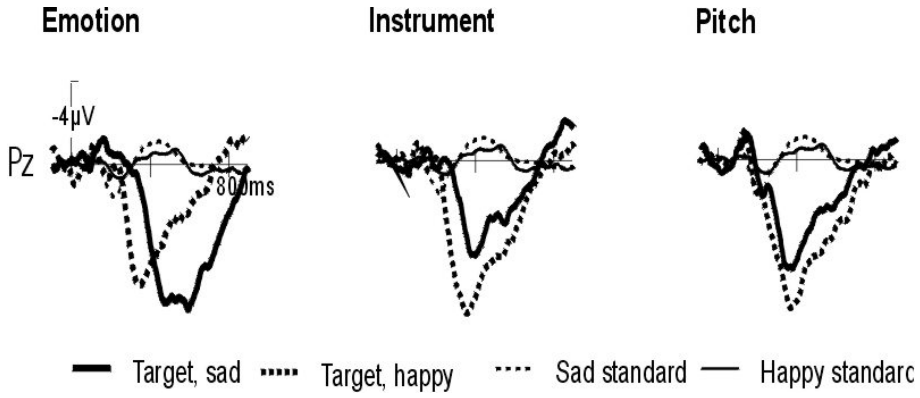

Figure 4.4.: ERPs from the active experiment for the emotion, instrument, and pitch condition (Pz electrode site), separated for happy and sad.

4.3.2. Passive condition

4.3.2.1. Electrophysiology

Fig. 4.5 shows the grand average waveforms for all three deviant types at three scalp positions (Fz, Cz, Pz). Note that the results from the two blocks, using the happy and the sad violin tone as standard stimuli respectively, are given in separate columns. The grand average waveforms to the standard tones show an initial small negative deflection (N1) at around 100 ms. This is followed by a long-duration negative component with a frontal maximum and a peak around 400 to 500 ms.[6] Inspection of the ERPs to the happy and sad standard stimuli (bold curves in Fig. 4.5) suggests that these are different, especially with regard to this long-standing negativity. Statistical analysis (successive 100 ms time-windows, Fz/Cz/Pz-electrodes) indicated a significant difference between sad and happy tones primarily for the tonic negativity (100-200 ms, $F(1,11)=1.78$, n.s.; 200-300 ms, $F=3.42$, n.s.; 300-400 ms, $F=5.1$, $p<0.05$; 400-500 ms, $F=6.77$, $p=0.024$; 500-600 ms, $F=6.32$, $p=0.029$; 600-700 ms, $F=8.87$, $p=0.013$; 700-800 ms, $F=9.3$, $p=0.011$).

Emotion deviants

The current design allows two different ways to compare emotional deviants. Firstly, deviants and standards collected in the same experimental blocks can be compared (i.e. happy standard vs. sad deviant or sad standard vs. happy deviant). These stimulus classes are emotionally as well as physically different. Secondly, deviants and standards can be compared across blocks so the compared stimuli are physically and emotionally same but differ in their functional significance as standard and deviant (i.e. sad standard vs. sad deviant and happy standard vs. happy deviant). Regardless of the comparison (Fig. 4.5, top), emotional deviants elicited a more negative waveform in the 150-250 ms

[6] This negativity is not seen in most MMN studies but was also found, for example, by Bostanov and Kotchoubey (2004) who presented expressive voice stimuli which, like in the present study, had a considerable length. Indeed, such long stimuli are known to give rise to a longstanding, tonic negativity (Keidel, 1971).

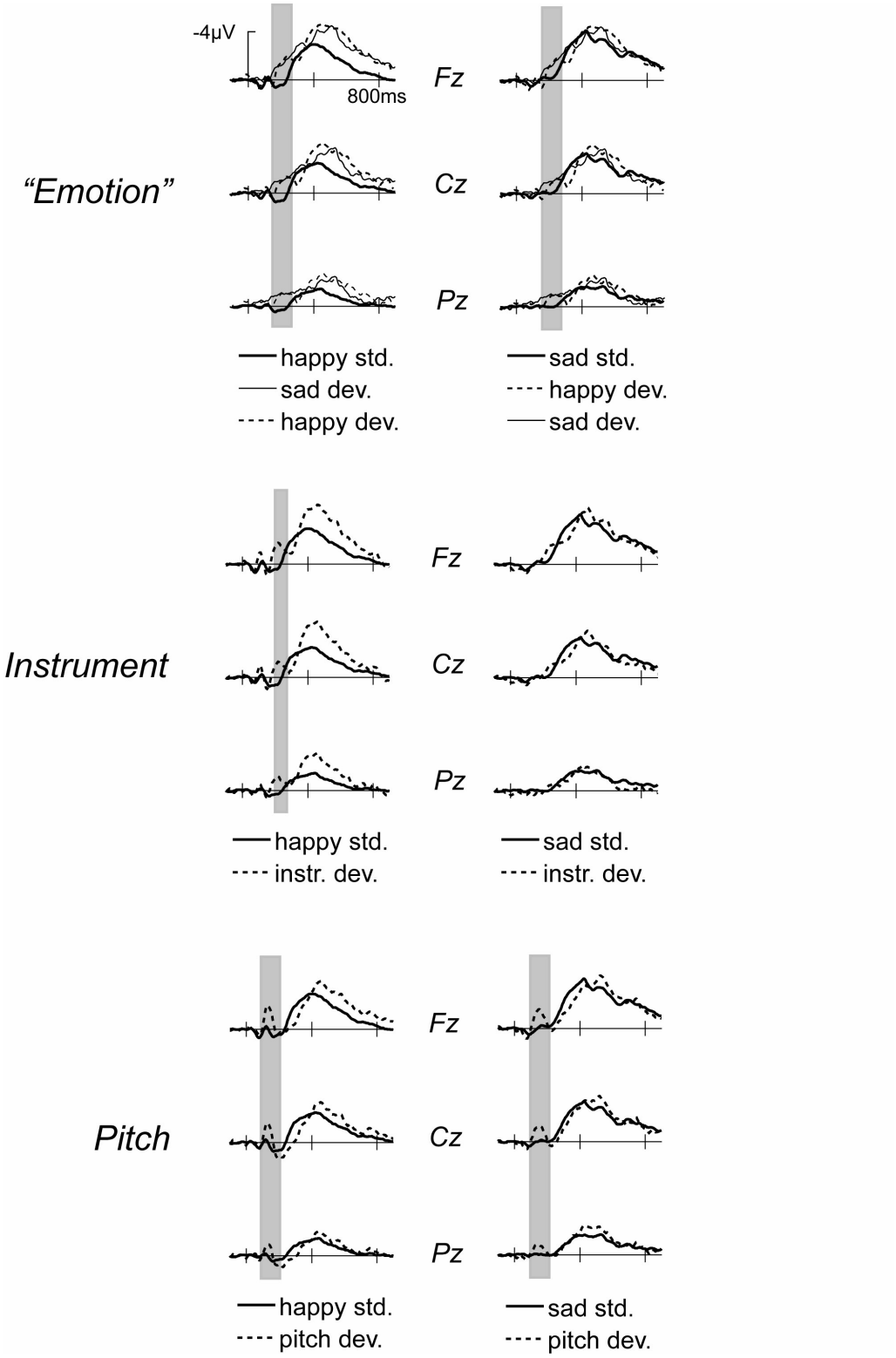

Figure 4.5.: Grand average ERPs from the passive experiment for three midline electrodes (Fz, Cz, Pz). Two columns (separated for happy and sad standard) are presented for each condition (emotion, instrument, pitch) showing the standard and the respective deviant. Time windows with significant mismatch effects are highlighted. See table 4.2 for F-values.

latency range (see Table 4.2 for results of the statistical analysis). Thus, the mismatch response cannot be explained by the fact that physically different tones per se elicited different ERP waveforms.

Instrument deviants

The MMN evoked by instrument deviants is shown in Fig. 4.5, middle. Though visible in both conditions, the comparison between instrumental deviant and standard tone only reached significance between 200 and 250 ms in the happy condition (see Table 4.2 for F-values).

Pitch deviants

Finally, stimuli deviating in pitch evoked an early MMN which was of similar size and morphology for 'happy' and 'sad' stimuli (Fig. 4.5, bottom). Statistical analysis (Table 4.2) revealed significant effects for both pitch deviants in the 100-200 ms time window.

Table 4.2.: Passive experiment; Comparison of standard vs. deviant stimuli; given are the F-values (df=1,11)

Comparison	Standard	Deviant	100-150 ms	150-200 ms	200-250 ms	250-300 ms
Emotion	Happy	Happy	0.10	2.72	22.75**	0.24
Emotion	Happy	Sad	1.33	9.64^{+}	11.28^{+}	3.38
Emotion	Sad	Sad	1.63	6.55^{+}	7.47^{+}	2.72
Emotion	Sad	Happy	0.19	0.06	12.02*	0.24
Instrumental	Happy	Happy	0.22	3.64	25.25**	0.25
Instrumental	Sad	Sad	0.47	0.01	3.84	0.5
Pitch	Happy	Happy	10.10*	2.72	22.75**	17.43**
Pitch	Sad	Sad	4.97^{+}	7.62^{+}	0.13	1.1

** $p < 0.001$
* $p < 0.01$
$^{+}$ $p < 0.015$

Difference waves

To isolate mismatch-related brain activity, deviant minus standard difference waves were computed (Fig. 4.6). These difference waves showed an initial negative peak, identified as the MMN, which was followed by a phasic positivity and finally, the tonic negativity mentioned above. The MMN for the different conditions appeared to differ markedly in latency. This was confirmed statistically by determining the peak latency of the most negative peak in the 100 to 300 time window [Cz site, $F(2,22)=20.3$, $p<0.001$]. Post hoc tests revealed a significant difference between the peak latencies in the pitch and emotion conditions ($p<0.001$) and between pitch and instrument conditions ($p<0.001$). There was no difference between the emotion and instrument conditions, however ($p>0.2$). While the latency of the negativity was very different for the different classes of deviant stimuli, the distribution of all three effects was virtually identical and typical for the MMN, as illustrated by spline-interpolated isovoltage maps[7] (see Fig. 4.6, right panel). This was corroborated by an analysis on the vector-normalized (McCarthy & Wood, 1985) mean amplitudes (taken in 40 ms time windows centered upon the peak latency of the negativity in each condition) which revealed no condition by electrode site interaction [$F(27,297)=0.16$, n.s.].

[7] Interpolation is a method to generate topographical maps of the relative voltage distribution over the head based on relatively few data points. Interpolation algorithms are applied to calculate additional virtual data points.

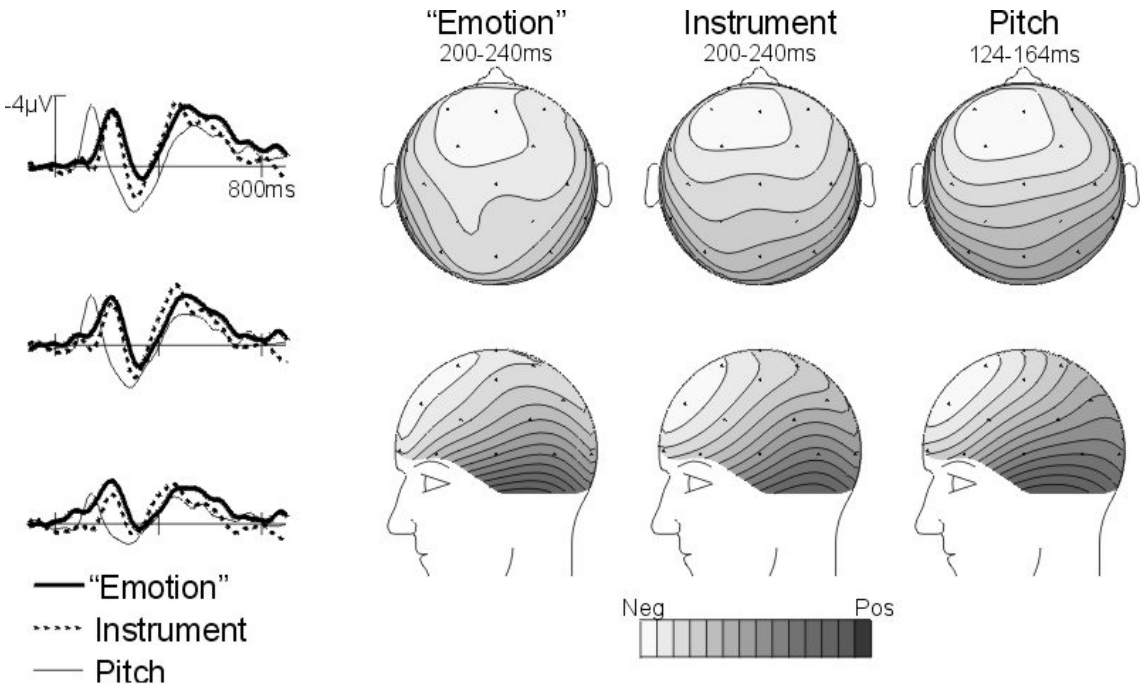

Figure 4.6.: Deviant minus standard difference waves. For these waveforms, data from both versions of the passive task (violin happy /c/ standard and violin sad /c/ standard) were averaged together. All three conditions show an initial negativity differing in latency. The scalp distribution of this negativity is shown on the right side using spline-interpolated isovoltage maps. These maps are based on the mean voltage in the 40 ms time window centered upon the peak latency of the negativity. The distribution of the negativities from the three conditions is virtually identical.

4.4. Discussion

4.4.1. Active condition

In the active condition, deviants of all types could well be detected by the participants. In addition, a P3b occurred in response to all three deviant types. On first sight it appeared that the P3b had a longer latency to emotion deviants than to the pitch and the instrument deviant. This effect, however, was triggered solely by the sad emotion deviants as was reflected by a significant deviant type x emotion-interaction. In correspondence with the delayed reaction time for sad emotion deviants, the prolonged P3b latency indicates that the sad tones required a longer evaluation process than the happy tones. Though from a evolutionists's perspective the privileged processing of happy tones could be interpreted as resulting from their greater significance, a concurrent - and more plausible - explanation lies in the different acoustical structures of happy and sad tones. As has been outlined in section 2.2 musical tones expressing sadness tend to have a slower tone attack than happy tones. Tone envelopes depicted in Fig. 4.7 show that this was also true for the current stimuli. It is likely that the sharp attack of the happy

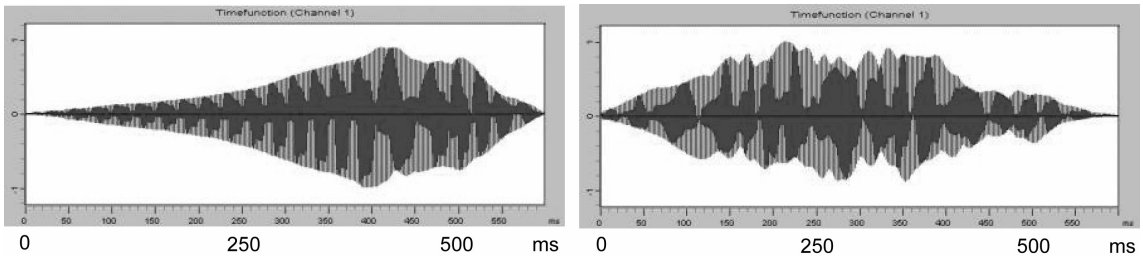

Figure 4.7.: Waveforms depicted for the sad (left) and the happy violin tone (right) represent the amplitude development over time.

tones facilitated (and accelerated) their detection. In addition, particularly at the onset, the happy tone was louder than the sad tone, a feature which is also known to decrease P3-latency (Sugg & Polich, 1995). To summarize, detection of emotional expression in single tones was fast and resulted in a P3b component with a latency similar to that evoked by pitch and instrument deviants. The latency of the emotion-P3b was found to vary as a function of acoustical features encoding emotional valence.

4.4.2. Passive condition

The results of the passive experiment demonstrate that affective deviants evoke a mismatch response even when subjects do not attend the auditory stimuli akin to the mismatch negativity that was seen for pitch and instrumental deviants. The ERP response evoked by the pitch deviant preceded the other two by about 80 ms. In addition, it did not display distinguishable N1 and MMN responses. It is possible that because the physical difference between standards and deviants was large in the pitch condition, the N1 and MMN components overlapped, as was described by Scherg, Vajsar, and Picton (1989). While the peak latency of the mismatch effects to the affective and instrumental deviants was delayed compared to the pitch deviant, the scalp distribution of the three mismatch effects was virtually identical on visual inspection (Fig. 4.6) and was statistically indistinguishable. The question arises then, what aspect of the emotionally deviant stimuli triggers the mismatch response in the current study. The finding of a highly similar distribution of all three deviant stimuli suggests that all of these engage the same generators, which are known to reside in the supratemporal plane with additional contribution by frontal cortex (Picton, Alain, et al., 2000; Sams, Kaukoranta, Hamalainen, & Näätänen, 1991). This further indicates that it is not the emotional quality per se but rather the physical differences between the stimuli of different emotional quality that give rise to the mismatch response. While the finding that tones which differ in physical structure evoke a mismatch negativity is trivial and has been shown repeatedly (Näätänen, 1992; Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001; Picton, Alain, et al., 2000, for reviews), the current study shows that the subtle physical differences used to convey emotional expression in single musical notes are sufficient to trigger the brain‘s automatic mismatch response. This automatic detection early in the auditory processing stream at least allows the rapid classification of stimuli according to their emotional quality during further and more detailed auditory analysis that then could be restricted to the emotionally deviant stimulus. The present study does not allow to determine whether the mismatch detection system indexed by the MMN component to emotional and instrumental deviants would be capable to extract physical invariants from

a series of different tone stimuli that are characteristic for particular (standard) emotions. To answer that question, a study using many different happy tones and respectively sad tones as standards was set up and will be presented in the next chapter.

5. MMN-Exp II: Are single tones categorized by the brain based on their emotional expression?

5.1. Introduction

5.1.1. Aim of the study

It was demonstrated in the first MMN-Experiment (see chapter 4) that subtle physical differences encoding the emotional expression in single violin tones are pre-attentively registered by the brain. However, because the MMN is known to be sensitive to all kinds of physical deviances (see review in section 4.1.2) it cannot be ruled out that the mismatch was triggered mainly by the physical difference of the happy and sad stimulus rather than affective valence per se. This notion, though, quite touches the crux of the matter. As reviewed in the introduction, affective expressions of a certain emotion indeed seem to share (psycho)acoustical features which also distinguish them from other emotional categories. However, what makes the study of acoustical emotion difficult is, that the set of features decoding the same emotion does not seem to be very well cut and that there is a great variance of feature combinations found within individual emotion categories. The question that was addressed in a subsequent MMN-experiment thus was, whether affective expressions are pre-attentively categorized even when their acoustical structure differs.

To test whether the brain automatically builds up categories of basic emotions across tones of different (psycho)acoustical structure, it was necessary to create two sets of tones, where tones within one set could clearly be categorized as happy and sad re-

spectively but differed with respect to their acoustical structure. Importantly, there has been evidence that the mismatch negativity is sensitive to perceptual differences rather than purely physical differences in sound structure (Winkler et al., 1995; Winkler, Tervaniemi, & Näätänen, 1997). Winkler et al. (1995) presented complex tones in which the fundamental frequency was missing. Usually, the perceived pitch of a tone equals its fundamental frequency. However, in a curious phenomenon, known as 'virtual' or 'residue' pitch perception (Moore, 2004), the pitch of missing-fundamental-tones does not differ from the fundamental frequency, although it is not present in the physical spectrum, nor can it be detected in the cochlea of the inner ear. In the experiment by Winkler et al. (1995) a MMN (recorded with magnetoencephalography) was found in response to variance of the virtual pitch. Thus the MMN proved to be sensitive to the subjective dimension of perception rather than the acoustical structure. To account for this finding, tones in the present experiment were selected based on subjective dissimilarity ratings instead of purely physical differences.

Two types of criteria were set for tones to be used as standards in the the planned Mismatch-Negativity-Study:

1. each tone needed consistently be categorized as happy or sad
2. tones within one set as well as across sets needed to be perceived as different.

The first point was addressed by performing affect-ratings on a set of violin tones which only differed in emotional expression but not in pitch or instrumental timbre. To tackle point 2, pairwise same-different-comparisons were collected for all tones and fed into a Fechnerian scaling procedure to assess the perceived similarity among the tones.

The following section describes the scaling experiment and the rating procedures that were applied to generate the stimulus set. The MMN-experiment will be described in section 5.3.

5.2. Scaling experiment

5.2.1. About scaling

Multidimensional Fechnerian scaling (E. N. Dzhafarov & Colonius, 1999, 2001) is a tool for studying the perceptual relationship among stimuli. The general aim of multidimensional scaling (MDS) is to arrange a set of stimuli in a low-dimensional (typically euclidean) space in a way that the distances among the stimuli represent their subjective (dis)similarity as it was perceived by a group of judges. Judges generally perform their ratings in pairwise comparisons of all stimuli in question. Based on the dissimilarity data a multidimensional scaling procedure finds the best fitting spatial constellation by use of a function minimization algorithm that evaluates different configurations with the goal of maximizing the goodness-of-fit (Kruskal, 1964a, 1964b). Though the dimensions found to span the scaling space can be interpreted as psychologically meaningful attributes that underlie the judgment, no a priori assumptions have to be made about the nature of the dimensions. Thus, with MDS perceptual similarity can be studied without the need to introduce predefined feature concepts (as labels for the dimensions) which might bias people's judgments.

Fechnerian scaling is a development of classical multidimensional scaling which is more suitable to be used with psychophysical data. Dzhafarov and Colonius (E. Dzhafarov & Colonius, 2006) point out that certain requirements for data to be used with classical MDS are usually violated in empirical data, namely the property of symmetry and the property of constant self-dissimilarity. The property of symmetry assumes that discrimination probability is independent of presentation order, thus that the probability to judge a stimulus x as different from a stimulus y is same no matter whether x or y is presented first ($p(x;y) = p(y;x)$). It has been known since Fechner (1860) that this is not true. The property of constant self-dissimilarity expects that any given stimulus is never perceived as different from itself, thus that the probability to judge stimulus x as different from itself is 0 ($p(x;x) = p(y;y)$. However, it has been shown repeatedly that this is not the case in psychophysical data (Rothkopf, 1957, for example). The only

requirement made by Fechnerian scaling is that of regular minimality, requesting that the probability to judge a stimulus as different from itself needs to be lower than any other discrimination probability.
In the present experiment Fechnerian scaling is used to establish subjective distances for a set of tones where tones differ only with respect to their emotional expression.

5.2.2. Materials and methods

5.2.2.1. Participants

Participants were 10 students (mean age=25.4 yrs, 5 female) with no musical expertise. Each was paid 20 Euro for participation.

5.2.2.2. Stimulus material

Stimulus material consisted of 10 individual violin tones of approximate length (mean=1.600 ms) and frequency (mean=558.97, see table 5.1 for details).

Table 5.1.: Features of the stimulus material.

Code	length in ms	mean pitch (SD) in Hz	mean level dB(A)
tone01	1676	559.69 (2.41)	64.5
tone02	1526	558.99 (2.046)	66.2
tone03	1658	559.98 (4.459)	72.2
tone04	1628	554.39 (3.557)	71.6
tone05	1506	555.86 (1.133)	68.8
tone06	1534	561.86 (4.352)	68.5
tone07	1660	563.00 (4.588)	66.6
tone08	1630	561.31 (3.613)	67.8
tone09	1570	556.96 (1.254)	72.4
tone10	1608	557.64 (0.353)	68.8

Generation of stimulus material To generate stimulus material 9 female violinists (all students of the Hanover University for Music and Drama) were asked to play brief melodic phrases all ending on c-sharp. Melodies were to be played several times with

either a happy, a neutral or a sad expression. To give all musicians the same idea as to what was meant by happy, neutral or sad expression, before each musician started with a new expression she was shown a sequence of pictures from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1995) which depicted happy, neutral or sad scenes, respectively.[1] All violinists were recorded on the same day in the same room using the same recording technique: stereo (2 Neumann-microphones TLM127), 44.1 kHz sampling rate, 24 bit, distance from the instrument to the microphones was always 50 cm. After playing each musician filled out a form describing the changes in technique they applied to perform different expressions (see Appendix). From 200 melodic phrases the last tone (always c-sharp) was extracted using Adobe Audition 1.0. Only those tones were selected which were between 1450 and 1700 ms in length and had a pitch between 550 and 570 Hz. Tones from two violinists had to be discarded altogether because they were consistently below pitch level. The resulting pre-selection consisted of 35 tones by 7 different violinists. To soften the tone onset a smooth-fade-in-envelope was created from 0 to 100 ms post tone onset. The pre-selection was rated on a 5-point scale from very sad (1) to very happy (5) by 9 participants (mean age=25.9 yrs, 5 men) who did not participate in the actual scaling experiment and were naïve to the purpose of the study. Each tone was rated twice by each participant to test the raters' consistency. Tones were not amplitude-normalized because it was found that differences in affective expression could hardly be differentiated in a normalized version. Based on the affect ratings and the rating consistence 10 tones were selected for the final stimulus set (see Appendix for complete rating results). No significant difference was found between first and second judgment[2] ($Z = -0.884$, $p = 0.377$).

5.2.2.3. Design

Participants took part in two separate sessions. In session 1 they performed a same-different-forced-choice-task on the violin tones to provide data for the multidimensional scaling. In session 2 (approximately one week later) they were asked to rate the emotional

[1] IAPS-numbers of the used pictures were: 1440, 1610, 8540, 7004, 7233, 7950, 2800, 9220, and 9421.
[2] Signed Rank on 8 raters (one had to be excluded because the data for the second testing was missing)

expression of the tones on a five-point-scale.
Stimulus material was rated according to valence and arousal by two additional groups of participants in different sessions.

5.2.2.4. Procedure – Same-different-judgment

Participants were tested individually while sitting in a comfortable chair 120 cm away from a 20-inch-computer screen. All auditory stimuli were presented via closed headphones (Beyerdynamic DT 770 M) with a level ranging from 64 dB to 73 dB. The 'Presentation' software (Neurobehavioral Systems) was used to present trials and to record responses. All 10 tones were combined with each other including themselves, resulting in 10x10=100 pairs; all 100 pairs were presented ten times, each time in a differently randomized order (resulting in 1000 trials altogether). The stimulus onset asynchrony (SOA) between two tones of one pair was 3500 ms. Participants had to strike one of two keys to respond same or different (forced choice). To make sure participants judged the psycho-acoustical similarity of the tones unbiased, they were kept uninformed that the purpose of the experiment was to study emotional expression. Trial duration was about 6000 ms. The next trial was automatically started when one of the two buttons was pressed. Participants performed a short training (3 trials, tones presented in the training session were not used in the main experiment) to familiarize them with the procedure. In the main experiment the first ten trials of each participant were excluded from the analysis. Participants were allowed to pause after each block of 25 trials. There were 40 blocks altogether. Participants could end the pause by pressing a button on the keyboard. Thus pause length was controlled by the participant. The duration of one whole experiment was about 2 hours. Participants were verbally instructed to decide whether the two tones of one pair were same or different. For the data analysis same-different-responses were recorded as 0 (same) and 1 (different). Mean values (discrimination probabilities) per pair of tones were calculated over all participants and all responses. Minimum number of responses per pair was 90. The resulting discrimination probabilities were transformed into Fechnerian distances using FSDOS (Fechnerian Analysis of Discrete Object Sets by

E.N. Dzhafarov and H. Colonius, see http://www.psych.purdue.edu/∼ehtibar/).

5.2.2.5. Procedure – Affect Rating

In session 2 each participant performed an affect-rating of each individual violin tone. All stimuli were presented twice but the order was randomized for each participant. Participants were asked to rate each tone on a 5-point-scale ranging from very sad (1) to very happy (5) by pressing one of the keys from F1 to F5 on the keyboard. Emblematic faces illustrated the sad and the happy end of the scale.

5.2.2.6. Procedure – Valence and Arousal Ratings

Since valence and arousal have been considered as separate underlying dimensions of emotion, the stimulus material was characterized accordingly in two separate ratings. Again, all stimuli were presented twice but the order was randomized for each participant. To give participants an idea what was meant by the terms valence and arousal they performed a short test trial on pictures taken from the IAPS. Group A (valence) (5 male, 5 female, mean age=27.6) was asked to rate all 10 tones on a 5-point-scale ranging from very negative (1) to very positive (5). Group B (5 male, 5 female, mean age=24.4) was asked to rate the 10 tones from very relaxed (German = 'sehr entspannt') (1) to highly aroused (German = 'sehr erregt') (5).

5.2.3. Results

5.2.3.1. Results of the same-different-judgment

Discrimination probabilities for each pair of tones based on participants' same-different-judgments are shown in table 5.2. Fechnerian distances for each pair of tones calculated from discrimination probabilities are shown in table 5.3. Given values reflect the relative distances between pairs of tones as perceived by the mean participant. For example, tone04 (abbreviated t.04 in the row), is perceived about 1.5 times more distant from tone05 than from tone07.

5.2.3.2. Results of the affect rating

Results of the affect-rating are shown in table 5.4, separated for first and second response. Since responses did not differ significantly between first and second presentation [$Z = -0.230$, Asymp. Sig. (2-tailed)[3]$= 0.818$] mean values of first and second rating (column 6) were used for subsequent analysis.

5.2.3.3. Results of the arousal rating and the valence rating

For results of the arousal and the valence ratings see table 5.5 and table 5.6. Though stemming from different groups of participants, there was a high correlation between the affect and the arousal ratings [$r=0.937$, $p<0.001$]. In contrast, the correlation between valence and affect ratings was rather low [$r=0.651$, $p=0.042$]. This is surprising for it was expected that valence and affect are closely related. It has to be noted, though, that during the testing it became apparent that participants used different concepts for the valence dimension. While some understood positive – negative in the sense of pleasant – unpleasant, others linked the two ends of the dimension to happy and sad. The problem is paralleled by a heterogeneous use of the valence-term in the literature (see Russell and Barrett, 1999, for a discussion) and might serve as an explanation for the incongruous pattern. In the current experiment the valence ratings will be interpreted with caution.

[3]Wilcoxon Signed Ranks Test; based on negative ranks

Table 5.2.: Discrimination probabilities for the 10 tones; given are probabilities with which the mean perceiver judges the row tones to be different from the column tones (t.=tone).

	t.01	t.02	t.03	t.04	t.05	t.06	t.07	t.08	t.09	t.10
t.01	0.06	0.12	1	0.89	0.74	0.81	0.86	0.94	0.88	0.89
t.02	0.16	0.08	0.98	0.91	0.69	0.72	0.85	0.89	0.88	0.93
t.03	0.99	0.97	0.04	0.93	0.97	0.93	0.85	0.88	0.98	0.95
t.04	0.9	0.93	0.96	0.08	0.82	0.42	0.51	0.64	0.6	0.96
t.05	0.7	0.77	1	0.84	0.08	0.79	0.85	0.91	0.78	0.74
t.06	0.89	0.8	0.94	0.62	0.93	0.07	0.3	0.35	0.74	0.79
t.07	0.92	0.91	0.97	0.69	0.86	0.41	0.09	0.2	0.89	0.93
t.08	0.9	0.91	0.94	0.75	0.9	0.31	0.16	0.1	0.86	0.83
t.09	0.88	0.95	0.96	0.66	0.82	0.77	0.8	0.76	0.08	0.26
t.10	0.91	0.94	1	0.91	0.65	0.77	0.89	0.82	0.34	0.06

Table 5.3.: Fechnerian distances between the tones (t.=tone) as calculated by FSDOS; reflected are the degrees of subjective dissimilarity among the tones as perceived by the participants (the larger the value the more distant the tones).

	t.01	t.02	t.03	t.04	t.05	t.06	t.07	t.08	t.09	t.10
t.01	0.000	0.140	1.890	1.650	1.290	1.510	1.630	1.670	1.620	1.680
t.02	0.140	0.000	1.830	1.680	1.290	1.370	1.590	1.620	1.660	1.730
t.03	1.890	1.830	0.000	1.770	1.850	1.760	1.690	1.680	1.820	1.850
t.04	1.650	1.680	1.770	0.000	1.500	0.890	1.030	1.190	1.100	1.550
t.05	1.290	1.290	1.850	1.500	0.000	1.570	1.540	1.630	1.440	1.250
t.06	1.510	1.370	1.760	0.890	1.570	0.000	0.550	0.490	1.360	1.430
t.07	1.630	1.590	1.690	1.030	1.540	0.550	0.000	0.170	1.520	1.660
t.08	1.670	1.620	1.680	1.190	1.630	0.490	0.170	0.000	1.440	1.490
t.09	1.620	1.660	1.820	1.100	1.440	1.360	1.520	1.440	0.000	0.460
t.10	1.680	1.730	1.850	1.550	1.250	1.430	1.660	1.490	0.460	0.000

Table 5.4.: Results of the affect-rating. Given are the results of the first (columns 2 and 3) and the second rating (columns 4 and 5) though for subsequent analyses mean values (column 6) were used. Labels in column 7 indicate which tones were chosen for the MMN-experiment.

Code	1st resp. Mean (SD)	1st resp. Median	2nd resp. Mean (SD)	2nd resp. Median	Mean resp. Mean (SD)	labels for MMN-exp.
tone01	2.0 (0.82)	2.00	1.80 (0.63)	2.00	1.90 (0.61)	sad01
tone02	1.90 (0.74)	2.00	2.00 (0.67)	2.00	1.95 (0.61)	sad02
tone03	4.30 (1.06)	5.00	4.50 (0.97)	5.00	4.40 (0.94)	
tone04	2.70 (0.67)	3.00	3.10 (0.57)	3.00	2.90 (0.39)	
tone05	2.50 (1.08)	2.50	1.90 (0.57)	2.00	2.20 (0.71)	sad03
tone06	2.50 (0.97)	3.00	2.90 (0.88)	3.00	2.70 (0.59)	
tone07	3.70 (1.06)	3.50	3.20 (1.14)	3.50	3.45 (0.98)	hap01
tone08	3.60 (0.97)	4.00	3.60 (0.70)	4.00	3.60 (0.77)	hap02
tone09	3.40 (0.70)	3.00	3.30 (1.16)	3.50	3.35 (0.71)	hap03
tone10	2.20 (0.79)	2.00	2.90 (0.57)	3.00	2.55 (0.55)	

Table 5.5.: Results of the arousal-rating. Given are the results of the first (columns 2 and 3) and the second rating (columns 4 and 5)as well as the mean values (column 6). Labels in column 7 indicate which tones were chosen for the MMN-experiment.

Code	1st resp. Mean (SD)	1st resp. Median	2nd resp. Mean (SD)	2nd resp. Median	Mean resp. Mean (SD)	labels for MMN-exp.
tone01	1.70 (0.82)	1.50	1.80 (0.42)	2.00	1.75 (0.42)	sad01
tone02	1.90 (0.74)	2.00	1.90 (0.74)	2.00	1.90 (0.66)	sad02
tone03	4.40 (0.52)	4.00	4.70 (0.48)	5.00	4.55 (0.44)	
tone04	3.00 (1.15)	3.50	3.30 (1.06)	3.00	3.15 (1.00)	
tone05	1.90 (0.57)	2.00	1.70 (0.67)	2.00	1.80 (0.54)	sad03
tone06	3.10 (0.88)	3.00	2.90 (0.99)	3.00	3.00 (0.62)	
tone07	3.00 (0.82)	3.00	2.90 (0.99)	3.00	2.95 (0.55)	hap01
tone08	3.20 (0.63)	3.00	3.20 (0.92)	3.00	3.20 (0.71)	hap02
tone09	3.40 (0.97)	3.50	3.40 (1.07)	3.00	3.40 (0.81)	hap03
tone10	2.70 (0.82)	2.50	2.90 (0.88)	3.00	2.80 (0.63)	

Table 5.6.: Results of the valence-rating. Given are the results of the first (columns 2 and 3) and the second rating (columns 4 and 5) as well as the mean values (column 6). Labels in column 7 indicate which tones were chosen for the MMN-experiment.

Code	1st resp. Mean (SD)	1st resp. Median	2nd resp. Mean (SD)	2nd resp. Median	Mean resp. Mean (SD)	labels for MMN-exp.
tone01	2.80 (1.40)	2.00	2.80 (1.48)	2.50	2.80 (1.40)	sad01
tone02	3.40 (1.07)	4.00	3.00 (1.49)	3.00	3.20 (0.98)	sad02
tone03	3.60 (1.17)	3.50	3.50 (1.08)	3.50	3.55 (0.90)	
tone04	3.40 (0.97)	3.50	3.30 (0.95)	4.00	3.35 (0.67)	
tone05	2.80 (1.14)	2.50	2.60 (0.84)	3.00	2.70 (0.63)	sad03
tone06	3.60 (1.07)	4.00	2.90 (0.74)	3.00	3.25 (0.49)	
tone07	3.20 (0.92)	3.00	2.70 (0.95)	3.00	2.95 (0.44)	hap01
tone08	3.50 (0.97)	4.00	3.10 (0.74)	3.00	3.30 (0.63)	hap02
tone09	3.20 (0.92)	3.00	3.30 (1.42)	4.00	3.25 (1.03)	hap03
tone10	2.50 (1.08)	2.50	2.90 (1.10)	3.00	2.70 (1.01)	

5.2.4. Selection for follow-up experiment

Following the criteria defined in 5.1.1, three sad tones [tone01 (sad01), tone02 (sad02), tone05 (sad03)] and three happy tones [tone07 (hap01), tone08 (hap02), tone09 (hap03)] were chosen from the data set based on their affect ratings. The happy tones had mean affective ratings of 3.45, 3.60 and 3.35, sad tones were rated 1.90, 1.95, and 2.20, respectively[4]. Affect ratings of happy and sad tones were significantly different (ANOVA-result: $F(9, 90) = 12.889$ $p < .000$; see Table 5.7 for pairwise post-hoc comparisons) and scaling procedures demonstrated that tones were perceived as different even when belonging to the same emotion category. Fechnerian distances between happy and sad tones lay between 1.44 and 1.67. Distances were 0.17, 1.52 and 1.44 among happy tones and 0.14, 1.29, and 1.29 among sad tones.

Table 5.7.: Post-Hoc-Tests (Bonferroni) for pairwise comparison between all six tones. Given are also the mean (m) affect ratings. Abbreviation: n.s.=not significant.

	Hap01 (m=3.45)	Hap02 (m=3.6)	Hap03 (m=3.35)	Sad01 (m=1.9)	Sad02 (m=1.95)	Sad03 (m=2.2)
Hap01	–	n.s.	n.s.	.000***	.000**	.007**
Hap02	n.s.	–	n.s.	.000***	.000**	.001**
Hap03	n.s.	n.s.	–	.001**	.001**	.021*
Sad01	.000***	.000***	.001**	–	n.s.	n.s.
Sad02	.000***	.000***	.001**	n.s.	–	n.s.
Sad03	.007**	.001**	.021*	n.s.	n.s.	–

[4]Despite a highly 'happy' mean rating tone03 was not included because it was described as aggressive rather than happy by a number of participants

5.3. The MMN-study

In the current experiment several different tones of a certain affective expression were included which had previously been rated as perceptually different (see above). It was to be tested if a stable memory trace is established for the standards despite variances between the tones of one emotional category. It has already been demonstrated that a MMN can be evoked by a deviant tone even in settings where the standards are not identical tones but differ among each other (Picton, Alain, et al., 2000, for a review). For example, in the study on timbre perception by Tervaniemi et al. (1997) described in section 4.1.2, nine different standard tones had been presented. In that example, the deviant was perceived as deviant because it lacked a feature that all other tones shared - the existence of any partial harmonics. It was concluded that the MMN-system found the constant feature of all standards and included the invariance among them in its representation of the standard. In a different study by Tervaniemi (Tervaniemi, Maury, & Näätänen, 1994) a MMN was evoked by tones in a series of decreasing frequencies whenever a tone did not follow the descending direction of the sequence (i.e. when its frequency was same or higher than the preceding stimulus). Thus, the standards were grouped together based on an abstract rule.

The hypothesis for the current experiment was, that if the expressive tones are pre-attentively categorized as happy or sad based on prototypical, psychological and/or psycho-acoustical similarities, a MMN will be found in response to the emotional deviant. No assumption on the nature of the grouping rule for the standards was made prior to the experiment but the question will be discussed on the basis of the results.

5.3.1. Materials and methods

5.3.1.1. Participants

Of a total of 19 participants three had to be excluded because of technical error (two) or too many blink artifacts in the ERP data (one). The remaining 16 participants (8 female) were aged between 21 and 29 years (mean=24.94 yrs). None was a professional musician.

5.3.1.2. Stimulus material

Stimuli were the 6 different single violin tones chosen on the basis of the scaling experiment (section 5.2).

5.3.1.3. Design

Two conditions were set up in a classical oddball-design. In condition A three sad tones were randomly presented (standards) with one happy tone (deviant) randomly interspersed. In condition B three happy tones were randomly presented as standards with one sad tone randomly interspersed as deviant tone (table 5.8 gives details on the design). The probability of occurrence was 25% for each of the three standard and the deviant tone, resulting in an overall probability of 75% for the standard and 25% for the affective deviant. In both conditions each tone was presented 340 times resulting in a total of 1360 tones per condition. A pseudo-randomization algorithm was applied to guarantee that identical tones were never presented back-to-back. Both conditions were divided in two blocks of 680 tones. The order of blocks was ABAB or BABA. All four blocks were presented in one session with one pause between block 2 and 3. The total duration of the experiment was about 90 minutes.

Table 5.8.: Design for the MMN-study: happy and sad tones were arranged such that one happy and one sad tone served as both, standard (std.) and deviant (dev.), respectively. Abbreviation: pres.=presentations.

	HAP01	HAP02	HAP03	SAD01	SAD02	SAD03	Total
number of pres. in Condition A	0	340 (dev.)	0	340 (std.)	340 (std.)	340 (std.)	1360
number of pres. in Condition B	340 (std.)	340 (std.)	340 (std.)	340 (dev.)	0	0	1360

5.3.1.4. Procedure

Participants were tested individually in the EEG-lab. Tones were presented via insert ear phones (EAR tone ABR; used with 'Earlink' ear-tips, Aearo Comp.). Stimulus on-

set asynchrony between two tones was 2000 ms. The mean sound pressure level of the presentation of all tones was 70 dB. To realize a non-attentive listening paradigm participants were instructed to pay attention to the cartoons[5] presented on a computer screen in front of them. The film was shown with no sound. To control how well participants had attended the film a post-test was performed after the experiment requiring participants to recognize selected scenes[6].

5.3.1.5. ERP-recording

The electroencephalogram (EEG) was recorded from 32 tin electrodes mounted in an elastic cap following the 10-20-system (Klem et al., 1999; see figure 5.1 for exact setup). Electrode impedance was kept below 5 kΩ. The EEG was processed through amplifiers set at a bandpass of 0.1 to 40 Hz and digitized continuously at 250 Hz. Electrodes were referenced on-line to the left mastoid (electrode A1 in figure 5.1). To allow for later off-line re-referencing to linked mastoids (the mean of the right and left mastoid electrodes) or nose, additional electrodes were positioned accordingly. Electrodes placed at the outer canthus of each eye were used to monitor horizontal eye movements. Vertical eye movements and blinks were monitored by electrodes above and below the right eye. Averages were obtained for 1024 ms epochs including a 100 ms pre-stimulus baseline period. Trials contaminated by eye movements or amplifier blocking or other artifacts within the critical time window were rejected prior to averaging. ERPs were calculated for time domain averaging for each subject. Separate averages were calculated for standards and deviants in both conditions. However, four separate averages were calculated for the two tones that served as deviant in one condition and as standard in the other condition: SAD01 as standard, SAD01 as deviant, HAP02 as standard, and HAP02 as deviant. ERPs were quantified by mean amplitude measures using the mean voltage of the 100 ms period preceding the onset of the stimulus as a reference. Time windows and electrode sites are specified at the appropriate places of the result section. Effects were tested for significance in separate ANOVAs, with function (standard or deviant) and

[5]Tom and Jerry - The classical collection 1

[6]Recognition rate was above 85% in all participants.

electrode site as factors. Before computing the statistics, the amplitudes were vector normalized according to the method described by McCarthy and Wood (McCarthy & Wood, 1985). Again, the Huynh-Feldt epsilon correction (Huynh & Feldt, 1980) was used to correct for violations of the sphericity assumption. Reported are the original degrees of freedom and the corrected p-values.

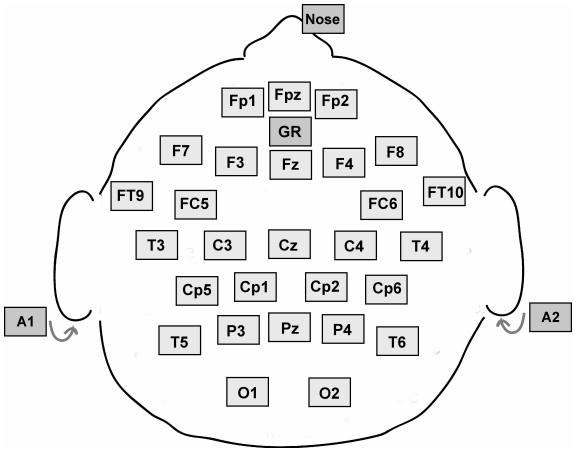

Figure 5.1.: Electrode setup according to the 10-20-System.

5.3.2. Results

The grand average waveforms to the standard and deviant tones (see Fig. 5.2) are characterized by a N1-P2-complex as typically found in auditory stimulation (Näätänen et al., 1988), followed by a long-duration negative component with a frontal maximum and a peak around 400 to 500 ms.

As in the first MMN-study, the current design allows two different ways to compare emotional deviants. Firstly, deviants and standards collected in the same experimental blocks can be compared (i.e. happy standard vs. sad deviant or sad standard vs. happy deviant). These stimulus classes are emotionally as well as physically different. Secondly, the ERP to the deviant can be compared with the same tone when it was presented as standard in the other condition (see table 5.8), so the compared stimuli are physically and emotionally same but differ in their functional significance as standard and deviant (i.e. sad standard vs. sad deviant and happy standard vs. happy deviant). Time windows for the the statistical analysis were set as follows: 100-200ms (N1), 200-300 ms (P2), and 380-600 ms. Electrode sites included in the analysis of the ERP-time courses were F3, Fz, F4, FC5, Fz, FC6, C3, Cz, C4, see Fig. 5.2. In condition A (Fig. 5.2, top), emotional

Table 5.9.: Comparison of standard vs. deviant stimuli; given are the F-values (df=1,15). Abbreviations: cond.=conditions.

comparison	Standard	Deviant	100-200 ms	200-300 ms	380-600 ms
condition A	sad standards	HAP02	0.93	2.40	7.32*
condition B	happy standards	SAD01	0.06	10.94**	0.00
across cond.	HAP02 as std.	HAP02	0.27	0.55	9.20**
across cond.	SAD01 as std.	SAD01	3.04	0.00	0.01

*** $p < 0.001$
** $p < 0.01$
* $p < 0.05$

(happy) deviants elicited a more negative waveform in a late latency range (from 380 ms), regardless of the comparison (see Table 5.9 for results of the statistical analysis).

Thus, the mismatch response cannot be explained by the fact that physically different tones elicited the different ERP waveforms. In condition B (Fig. 5.2, bottom), emotional (sad) deviants, too, elicited a more negative waveform than the happy standards, though in an earlier latency range (P2, 200-300ms) (see Table 5.9). However, no difference was found when the ERPs to the sad tone were compared across conditions, suggesting that this effect was triggered by the structural difference of happy and sad tones rather than their functional significance as standard and deviant.

To summarize the result: presenting a happy tone in a series of sad tones resulted in a late (possibly mismatch) negativity that was larger in amplitude than the ERP to the same happy tone functioning as standard in the opposite condition. In contrast, no difference that could be related to its functional significance was found for the sad tone presented in a train of differing happy tones.

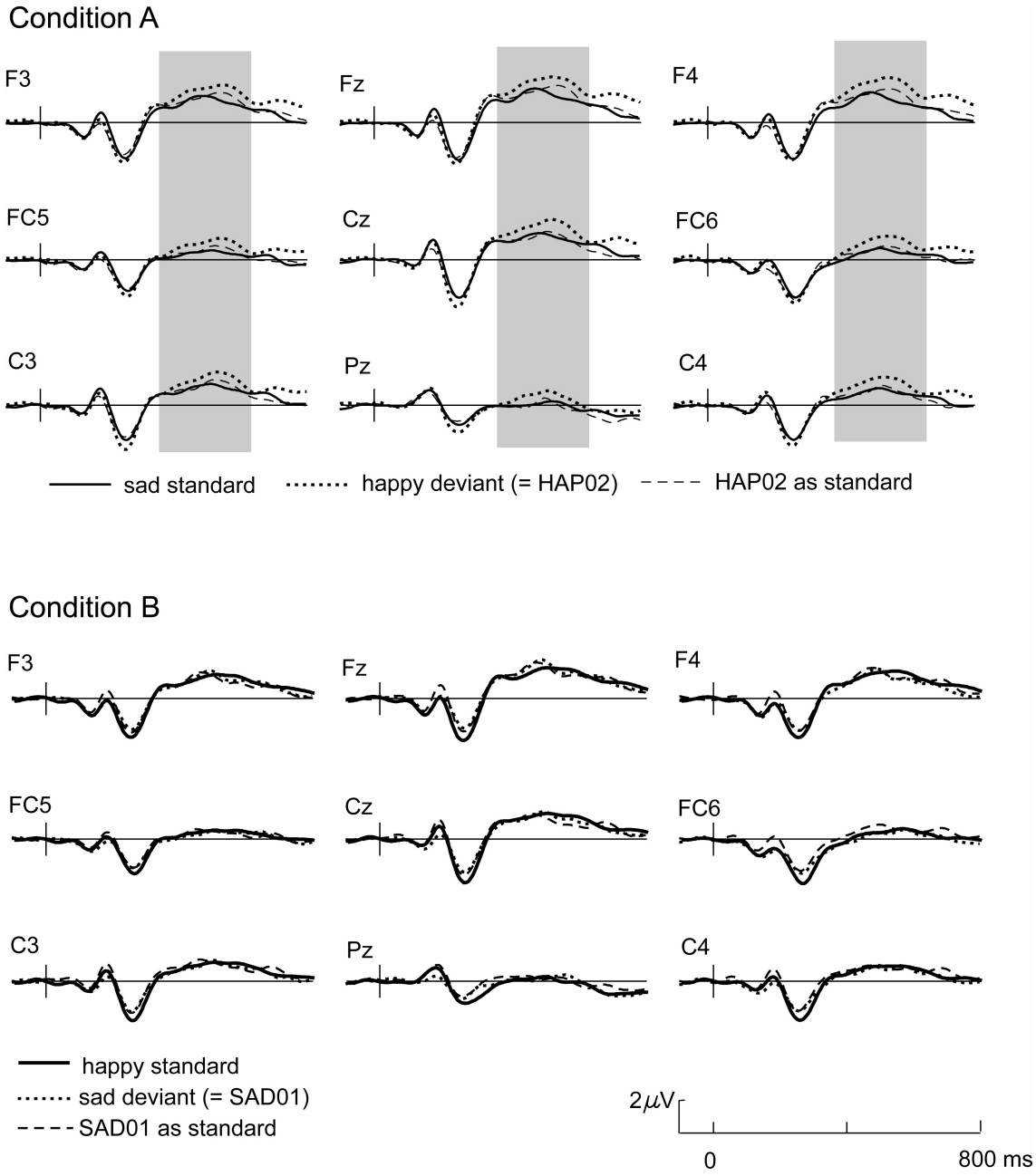

Figure 5.2.: Grand average ERPs for condition A (top) and B (bottom); the respective standard-ERP (bold line) is superposed by the ERP to the emotionally deviating tone when it was presented as deviant in the same (dotted line) or as standard in the concurrent condition (dashed line). Timewindows are highlighted where the ERP amplitude was significantly different in both standard-deviant and standard-standard comparison.

5.4. Discussion

The affective deviant in condition A evoked a clear mismatch reaction. Though the latency was unusually long, its topographic distribution as well as the fact that it was evoked by a deviant tone in a classical oddball paradigm indicate it as belonging to the MMN-family. Indeed, it is a known fact (as reviewed in section 4.1.2), that MMN-latency increases with discrimination difficulty. No doubt, discrimination was particularly difficult in the present experiment because the difference in timbre was reduced to subtle changes in the expression of same-pitch and same-instrument tones. However, a mismatch reaction reflected that a happy tone was pre-attentively categorized as different from a group of different sad tones. Generally, a MMN stands for change detection in a previously established context (Näätänen, 1992). Thus, for it to occur, a context needs to be set up first. Consequently, the important question in the present experiment is not, what is so particular about the happy tone? The question is, what has led to grouping the standard (sad) tones into one mutual category, so that the single happy tone was perceived as standing out? For the happy tone to be categorized as deviant it was required that the sad tones – though different in structure – were perceived as belonging to the same context, i.e. category. The question thus arises: what has led to grouping of the sad tones? Three possibilities seem plausible:

-perceptual similarity

-emotional similarity or

-emotion-specific perceptual similarity.

Perceptual similarity

From the result of the scaling-experiment it can be derived, that tones within the sad category were perceived quite as different from each other on a perceptual basis (e.g. sad01 and sad03: Fechnerian distance=1.290) as was the deviant from the standards (e.g. sad03 vs. happy deviant: Fechnerian distance=1.440). Relative distances are visualized in Fig. 5.3. The arrangement of tones in a three dimensional space results from feeding Fechnerian distance values into a multidimensional scaling procedure (Alscal

in SPSS 12.0 for windows)[7]. Though the positions of SAD01 and SAD02 are relatively close, both are rather distant from SAD03. Grouping, thus, cannot be explained by perceptual similarity alone.

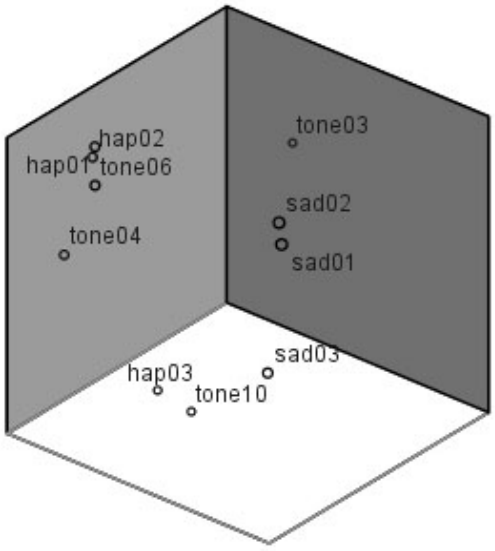

Figure 5.3.: Arrangement of tones in a three dimensional space based on a multidimensional scaling procedure. Note that orientation of dimensions is arbritary.

Emotional similarity

Affect ratings (1.90, 1.95, and 2.20) indicate that the tones were perceived as equally sad in expression. There thus is some support for the hypothesis that the tones were grouped together based on their emotional category. However, if it was the emotional expression that has led to the automatic categorization why did it not work in condition B? No index was found for a mismatch reaction in response to a sad tone randomly interspersed in a train of different happy tones. Arguing along the same line as before, this (non)finding implies that no mutual standard memory trace was derived from the happy tones. Since the affect ratings of the happy tones had been just as homogeneous (3.35, 3.45, and 3.60) as those of the sad tones, the question arises, if the affect ratings gave a good enough representation of the emotion as it was decoded by the listeners.

[7]Alscal finds the optimal constellation of stimuli in a n-dimensional space based on dissimilarity data; 3 dimensions were found to explain 99% of variance. Note that the orientation of the dimensions is arbritary.

Against the background that decoding accuracy of acoustical emotion expressions has repeatedly been reported to be better for sadness than for happiness (Juslin & Laukka, 2003; Elfenbein & Ambady, 2002; Johnstone & Scherer, 2000), it might be necessary to take a second look at the stimulus material. Banse and Scherer (1996) found that if participants had the option to choose among many different emotional labels to rate an example of vocal expression, happiness was often confused with other emotions. In the present experiment participants had given their rating on bipolar dimensions ranging from happy to sad. It cannot be ruled out that the response format biased the outcome. It is for example possible that in some cases participants chose to rate happy because the tone was found to be definitely not-sad even if it was not perceived as being really happy either. In an attempt to examine the perceived similarity of the tones with respect to the expressed emotion without pre-selected response categories, a similarity rating on emotional expression was performed post-hoc. For that purpose, the students who had participated in the first scaling-experiment already (see section 5.2.2.4), were asked to perform another same-different-judgment on the same stimulus material though this time with regard to the emotion expressed in the tone. The results are depicted in table 5.10. As can be read off of table 5.10, sad tones (t.01, t.02, and t.05) were perceived considerably more similar to each other with respect to the emotion expressed than the happy tones (t.07, t.08, and t.09). In fact, sad tones were judged half as dissimilar from each other than the happy tones (0.503 vs. 1.02). Fig. 5.4 shows the relation of same and different responses given for happy and sad tone pairs, respectively. Sad tones were considerably more often considered to belong to the same emotional category than happy tones (80% vs. 57% 'same'-responses). It can be assumed that in the MMN-experiment, too, sad tones (in condition A) were perceived as belonging into one emotional category while happy tones (in condition B) were not. The difficulty to attribute the happy tones to the same 'standard' category can serve as explanation why the sad tone did not evoke a MMN. It was not registered as deviant against a happy context because no such context existed. Nevertheless, the hypothesis that the MMN reflects deviance detection based on emotional categorization can at least be held up for condition A.

Table 5.10.: Fechnerian distances as calculated from same-different-judgments of emotional expression for the 10 tones; given are perceived distances of row tones and column tones with respect to their emotional expression; sad tones were t.01, t.02, and t.05, happy tones were t.07, t.08, and t.09.

	t.01	t.02	t.03	t.04	t.05	t.06	t.07	t.08	t.09	t.10
t.01	0.000	0.012	1.763	1.003	0.491	0.943	1.103	1.003	1.072	0.983
t.02	0.012	0.000	1.751	0.991	0.503	0.931	1.091	0.991	1.072	0.971
t.03	1.763	1.751	0.000	1.390	1.700	1.040	0.880	0.990	1.420	1.560
t.04	1.003	0.991	1.390	0.000	0.820	0.580	0.630	0.620	0.600	0.750
t.05	0.491	0.503	1.700	0.820	0.000	1.020	1.170	1.080	0.730	0.650
t.06	0.943	0.931	1.040	0.580	1.020	0.000	0.160	0.060	0.860	0.850
t.07	1.103	1.091	0.880	0.630	1.170	0.160	0.000	0.110	1.020	1.010
t.08	1.003	0.991	0.990	0.620	1.080	0.060	0.110	0.000	0.920	0.910
t.09	1.072	1.072	1.420	0.600	0.730	0.860	1.020	0.920	0.000	0.150
t.10	0.983	0.971	1.560	0.750	0.650	0.850	1.010	0.910	0.150	0.000

Figure 5.4.: Same and different responses for tone pairs in the categories sad (left) and happy (right), respectively. Responses are give in percent (1,0=100%).

Emotion-specific perceptual similarity

It was presupposed that emotion recognition in acoustical stimuli is based on certain acoustical cues coding the emotion intended to be expressed by the sender. To test whether the sad tones in the present experiment were similar with regard to prototypical cues for sadness an acoustical analysis was performed on the stimulus set. Tones were analyzed on the parameters found to be relevant in the expression of emotion on single tones (see section 2.3.3), namely timbre, attack, pitch, vibrato, and sound level. The analysis was performed with the help of 'dbSonic' (01dB GmbH) and 'PRAAT' (Boersma & Weenink, University of Amsterdam; see also appendix). Table 5.11 summarizes the results. To be able to categorize parameter values as low, medium, and high, values were set into relation with the tones of the 'happy'-set. The acoustical analysis revealed

Table 5.11.: Results of the acoustical analysis of the sad tones. Tested were parameters expected to be relevant cues to express emotion on single tones (compare table 2.2 in section 2.3.3). Categorization as low, medium, and high was based on comparison with the 'happy' tones. Parameter values meeting expectations are printed bold.

	SAD01	SAD02	SAD03
timbre (high frequ. energy)	**low**	**low**	**low**
attack	medium	medium	medium
mean pitch	**low**	medium	medium
pitch contour	normal	**down**	**down**
vibrato amplitude	medium	medium	**low**
vibrato rate	**slow**	medium	**slow**
sound level	**low**	medium	medium

that some though not all parameters were manipulated the way it would have been expected based on previous findings (printed in bold). However, the summary in table 5.11 reflects that the cues were not used homogeneously. For example, mean pitch level was no reliable cue. Moreover, vibrato was manipulated in individual ways by the musicians, as had been predicted in section 2.3.3. Timbre, however, was well in line with expectations. All sad tones were characterized by little energy in the high frequency spectrum. In contrast, more energy in high frequencies was found in the spectrum of

the deviant happy tone (see Appendix). Based on the findings by Tervaniemi et al. (1994), outlined in section 5.3, the possibility needs to be discussed that the difference in spectral structure alone triggered the MMN. That would mean that the sad tones were grouped together as standards based on their mutual feature of attenuated higher partials. It has to be noted though that the high-frequency energy parameter is a very coarse means to describe timbre. Especially in natural tones (compared to synthesized tones as used by Tervaniemi et al., 1994) the spectrum comprises a large number of frequencies with different relative intensities. As a consequence, the tones still have very individual spectra (and consequently sounds), even if they all display a relatively low high-frequency energy level. This fact is also reflected in the low perceptual similarity ratings. Moreover, if the spectral structure really was the major grouping principle, it should also have applied to the happy tones in condition B. Here, all happy tones were characterized by a high amount of energy in high frequencies, while the sad deviant was not. Nevertheless, no MMN was triggered. To conclude, though the possibility cannot be completely ruled out, it is not very likely that the grouping of the sad tones was based solely on similarities of timbre structure.

Instead, the heterogeneity of parameters in table 5.11 provides support for Juslin's idea of redundant code usage in emotion communication (Juslin 1997b, 2001, see also section 2.3.4). Obviously, expressive cues were combined differently in different sad tones. Thus, though the sad tones did not display homogeneous patterns of emotion-specific cues, each tone was characterized by at least two prototypical cues for sadness expression. Based on the model assumption of redundant code usage, it seems likely that tones were grouped together because they were identified as belonging to one emotional category based on emotion specific-cues.

What implication does this consideration have for the question of grouping principles in the MMN-experiment? From what is known about the principles of the MMN, the results imply that the representation of the standard in the working memory included invariances of several different physical features. The pattern of features, however, needed to be in

line with a certain template on how sadness is acoustically encoded. Several researchers have suggested the existence of such hard-wired templates for the rapid processing of emotional signals (Lazarus, 1991, LeDoux, 1991, Ekman, 1999a, Scherer, 2001a, see also section 2.2.2). It is assumed that to allow for quick adaptational behavior, stimulus evaluation happens fast and automatic. Incoming stimuli are expected to run through a matching process in which comparison with a number of schemes or templates takes place. Templates can be innate and/or formed by social learning (Ekman, 1999a). The present study cannot give any information on the origin of the template. But it gives some idea how such a matching process might be performed on a pre-attentive level. Given the long latency of the MMN in the present experiment, it can be assumed that basic sensory processing has already taken place before the mismatch reaction occurs. It is thought likely that the MMN instead reflects the mismatch between the pattern of acoustical cues and the template for sad stimuli activated by the preceding standard tones. Our data is thus in line with considerations that the MMN does not only occur in response to basic acoustical feature processing. Several authors have suggested that the MMN can also reflect 'holistic' (Gomes, Bernstein, Ritter, Vaughan, & Miller, 1997; Sussman, Gomes, Nousak, Ritter, & Vaughan, 1998) or 'gestalt-like' (Lattner et al., 2003) perception. They assume that the representation of the 'standard' in the auditory memory system is not merely built up based on the just presented standard-stimuli, but that it can be influenced by prototypical long-term representations stored in other areas of the brain (Phillips et al., 2000). Evidence for this notion comes from speech-specific phoneme processing. Phillips et al. (2000) presented syllables ranging on the /dæ/ - /tæ/ continuum, which acoustically only differed with respect to voice onset time (VOT). With increasing VOT a categorical perception shift from 'd' to 't' takes place. The authors used several different standards (i.e. different VOTs) and several different deviants (again different VOTs) to test if the different stimuli were grouped together based on phonetic categories (i.e. /dæ/ vs. /tæ/). A MMN (in MEG) was found if the low vs. high ratio of occurrence for deviants and standards followed the perceptual boundary (i.e. few /tæs/ vs. many /dæs/). However, in a control condition the VOT

was increased by 20 ms in all samples, thus equalizing the proportion of /dæs/ and /tæs/ without changing the relative range of VOT. The proportion of the perceived 't's and 'd's was now equal. The authors assumption was that if the MMN in the first experiment would have been triggered by the acoustical difference (short vs. long VOT), a similar effect would have to be expected in the second experiment. But no MMN was evoked. The results provide strong evidence that the MMN-response did not only rely on matching processes in the transient memory store but that long-term representations for prototypical stimuli (in this case phonemes) were accessed already at a pre-attentive level. For phonemes, Näätänen (1999) indeed assumed the existence of long-term memory traces serving as recognition patterns or templates in speech perception. He expects that they can also be activated by sounds *"nearly matching with the phoneme-specific invariant codes"* (p. 14). He points out though (Näätänen, Jacobsen, & Winkler, 2005) that the *"mechanisms of generation of these more cognitive kinds of MMNs of course involve other, obviously higher-order, neural populations than those activated by a mere frequency change"* (p. 27).

The results of this thesis also fit well into the 3-stage-model of emotional processing by Schirmer & Kotz (2006), introduced in section 2.3.5. In the model emotional-prosodic processing is conceptualized as a hierarchical process. Stage 1 comprises initial sensory processing of the auditory information before emotionally significant cues are integrated (stage 2) and cognitive evaluation processes (stage 3) take place. Based on their own data (Schirmer, Striano, & Friederici, 2005), the authors, too, suppose that the MMN in response to emotional auditory stimuli reflects the stage of integrating emotionally significant cues. The present data extends the model to the area of nonverbal auditory emotion processing. However, because of the similarities between emotional coding in segmental features of music and paralinguistic features of speech, it can be assumed that the recognition of at least certain aspects of affective speech prosody are based on similar mechanisms. The current data contributes to disentangling the processes underlying emotion recognition in the auditory domain. It has to be pointed out though that the present results can only give a first glimpse on the mechanisms underlying

processing of emotionally expressive tones. More studies with a larger set of tones characterized by different cues are needed to systematically examine the nature of the stimulus evaluation process.

Part II.

Processing of Vocal Emotion Expression

6. Experiment II-01: Timbre as a code for emotion and identity

6.1. Introduction

Listening to a speaking voice provides the listener not only with the semantic information conveyed by the linguistic content of the utterance. It is also possible to make inferences on the speaker's gender, age, health, or emotional state. The voice may be regarded as the *"auditory face"* (Belin, Fecteau, & Bedard, 2004) of a speaker. The features of the auditory face are commonly referred to as the 'paralinguistic' aspects of speech. However, paralinguistic aspects may further be divided into 'expressive' and 'organic' aspects (Traunmüller, 1997). Expressive aspects can be intentionally used by the speaker (e.g. emotional expression). In contrast, organic aspects cannot be modulated intentionally because they are based on physiological conditions related to age, gender or state of health (Roach, Stibbard, Osborne, Arnfield, & Setter, 1998). As an example for an organic paralinguistic aspect, gender can be derived from the F0-frequency (Gelfer & Mikos, 2005), which is generally lower in men than in women or children as a consequence of the length of the vocal tract (i.e. the distance from the vocal folds to the mouth) (Sundberg, 1999).

Recent years have seen an increase of studies addressing the question how paralinguistic features are processed in the brain. Though the majority focused on expressive aspects (see review in section 2.3.5), organic features have gained particular interest for their function to encode speaker identity (Traunmüller, 1997; Fecteau, Armony, Joanette, & Belin, 2004; Lattner, Meyer, & Friederici, 2005). Brain imaging results indicate involvement of anterior regions of the temporal lobes in speaker identification (Imaizumi et al.,

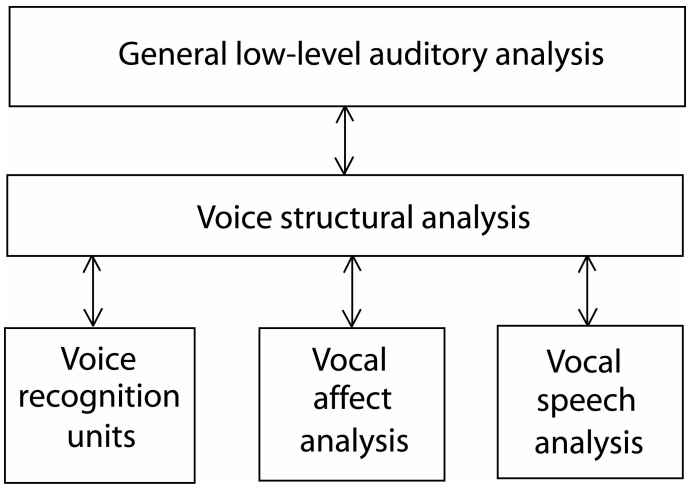

Figure 6.1.: Model of voice perception (adapted from Belin et al., 2004, p. 131).

1997; Nakamura et al., 2001; Kriegstein, Egera, Kleinschmidt, & Giraud, 2003). The finding is in line with brain lesion studies. Deficits to recognize or discriminate speakers are often found after temporal lobe damage (D. R. van Lancker & Canter, 1982; Neuner & Schweinberger, 2000). Belin et al. (2004) introduced a model of voice perception which postulates separate processing units for vocal speech analysis, vocal affect analysis, and voice recognition (see Fig. 6.1). There is evidence from a brain imaging study that vocal affect analysis and speaker recognition are mediated by different structures in the brain. Imaizumi et al. (1997) found a stronger activation of the right than the left parahippocampal gyrus when subjects were required to identify the emotional expression of spoken words, but a stronger left activation when the task was to identify the speaker. A dissociation of the two functions would parallel the assumed independent processing of expression and identity recognition in face perception (Young, Newcombe, Haan, Small, & Hay, 1993; Calder & Young, 2005; Hefter, Manoach, & Barton, 2005). Matching faces for identity was found to happen faster than matching emotional expression (Münte et al., 1998). Münte et al. (1998) recorded event-related potentials (ERP) while presenting pairs of faces belonging either to the same or to different persons in one condition (identity task) and pairs of faces with either the same or different emotional expressions in the other condition (emotion task). Differences in ERP-amplitude between the congruent and the incongruent condition occurred at 200 ms in the identity

task but about 250 ms later in the emotion task.

Timing differences between recognition of vocal affect and speaker identification have not yet been under study. However, examining the temporal dynamics of these cognitive functions is important to get a comprehensive picture of the mechanisms underlying voice processing. Thus, following the design of Münte et al. (1998) a priming experiment was set up presenting pairs of voice stimuli, which were either congruent or incongruent with respect to expressed emotion or with respect to identity of the singer. Sung tones were used to avoid the linguistic and structural influence a verbal utterance might have on the decoding process (Banse & Scherer, 1996).

Aim of the study Priming effects are generally assumed to reflect facilitated identification of perceptual material that was preceded by a stimulus triggering a similar brain response. Whether priming has taken place can be inferred from reduced response latency, reduced error rates, or altered neural responses. In the present experiment, it was expected that presenting a congruent voice pair would modulate the ERP amplitude to the second tone compared to when the preceding tone was incongruent. Occurrence of a priming effect was taken as an index that processing of the task relevant features was completed. It was hypothesized that, if the processing of identity-relevant cues happened earlier than that of emotion-relevant cues an earlier difference would occur between congruent and incongruent pairs in the identity than in the emotion matching task.

6.2. Materials and methods

6.2.1. Stimulus material

Tone pairs were created from individual notes sung on the syllable 'ha' by five different female opera singers and advanced singing students. Singers had been instructed to give each individual tone a happy or sad expression. To select the tones that could be categorized according to affect regardless of their briefness, a group of 10 subjects (21-30 yrs, 5 female) naïve to the purpose of the experiment rated the tones on a 7-point-scale ranging from 1 (very sad) to 7 (very happy). For the EEG experiment only those tones were used that had consistently been categorized as happy (5, 6 or 7 on the rating-scale) or sad (1, 2 or 3 on the rating-scale) by at least 7 of 10 raters, who did not participate in the main experiment. From this pre-selection two sets á 10 stimuli were created, one consisting of tones categorized as happy the other one consisting of tones categorized as sad. The sets were matched for arousal [mean=2.73 (SD=0.39) for happy, mean=2.39 (SD=0.28) for sad on a 1-to-7-scale] and intensity of valence [mean=5.33 (SD=0.45) for happy, mean=2.88 (SD=0.48) for sad on a 1-to-7-scale]. The mean length was 373 ms (SD=63 ms) for happy tones and 414 ms (SD=53 ms) for the sad tones. The happy tones ranged between C4 ($\sim$262 Hz) and A#4 ($\sim$466 Hz) in pitch, the sad tones between B3 ($\sim$247 Hz) and A#4 ($\sim$466 Hz).

6.2.2. Participants

Sixteen undergraduates of the University of California San Diego (UCSD) (8 males, 8 females, age 18-25 years) participated in the experiment for cash and/or credit. The study was approved by the UCSD Human Subjects Committee and participants gave informed, written consent. All participants were right-handed. None of the students was enrolled in music classes though some reported that they had learned to play an musical instrument as children but did not play anymore. For the final analysis 1 subject was excluded because of too many blink artifacts. Another one was excluded because the performance was at chance level.

6.2.3. Design

Two tasks were alternated blockwise. In task A (emotion matching) participants were asked to decide whether the emotional expression in pairwise presented tones was same or different. Task B (identity matching) required to decide whether the identity of the singer was same or different in tone 1 and 2 of a pair. For each task two blocks of 54 pairs were presented with pauses between blocks. Order of blocks was counterbalanced across participants (either ABAB or BABA). The tone pairs for both tasks were created from the two sets of tones described in section 6.2.1. A total of 120 pairs were constructed for each task consisting of 50% pairs congruent in emotional valence and 50% incongruent in emotional valence, and holding equally many happy and sad tones. The same stimuli were used as targets (S2) in both tasks. In the emotion matching task the first (S1) and the second tone (S2) were always sung by different singers regardless of the expressed emotion. In the identity matching task the expressed emotion between S1 and S2 was always same even if the singer was not[1]. The target (S2) set was identical for the emotion and the identity matching task as well as for the congruent and the incongruent condition. Pairs were kept stable for all participants but the order of presentation was randomized each time. Table 6.1 gives an overview of the stimulus setup. Because pretests revealed that some pairs were too hard to match they were taken out of the stimulus set. To maintain an equal number of pairs per task equally many pairs (3 per condition) were taken out of the whole set, resulting in 27 pairs per condition. In the experiment the whole set of pairs was presented twice per task resulting in the presentation of a total of 216 pairs.

Reaction time experiment

For methodological reasons participants' responses in the ERP experiment were delayed. Thus, no reaction times were recorded. However, to get a behavioral timing correlate, a reaction time experiment was run in parallel presenting the same stimulus set to another

[1] Unfortunately, the stimulus material did not allow for a complete factorial design (cross-over manipulation of factors) because it was not always possible to use happy and sad tones sung by the same singer.

Table 6.1.: Overview of the experimental setup. Note that in the emotion matching task the first (S1) and the second tone (S2) were always sung by different singers regardless of the expressed emotion. In the congruent pairs of the identity matching task though the singer was same, the actual tone was not (as marked by *). However, the expressed emotion between S1 and S2 was always same even if the singer was not. The target (S2) set was identical for the emotion and the identity matching task as well as for the congruent and the incongruent condition. Abbreviations: nr.=number.

Emotion matching task	first tone (S1)		second tone (S2)		nr. of pairs
congruent	happy	(singer A)	happy	(singer B)	27
congruent	sad	(singer A)	sad	(singer B)	27
incongruent	happy	(singer A)	sad	(singer B)	27
incongruent	sad	(singer A)	happy	(singer B)	27
Identity matching task					
congruent	singer A	(happy)	singer A*	(happy)	27
congruent	singer A	(sad)	singer A*	(sad)	27
incongruent	singer A	(happy)	singer B	(happy)	27
incongruent	singer A	(sad)	singer B	(sad)	27

group of ten subjects (mean age=28.2 (SD=4.85) yrs, 5 male).

6.2.4. Experimental procedure

Participants of the ERP-experiment were tested in a soundproof, electrically shielded chamber. They were seated in a comfortable chair in front of a 21-inch-monitor (distance to the monitor screen was 127 cm). Before starting the experimental session all participants were tested in a hearing test to determine their individual auditory threshold. The volume of the stimuli was then adjusted via an attenuator (Hewlett Packard 350 D) to guarantee the same relative loudness for all participants. The experimental session started with a short practice trial to familiarize the participants with the procedure. The presentation of the stimulus pair was preceded by the appearance of a fixation cross in the center of the screen. Participants were asked to fixate the cross and refrain from blinking throughout one trial. The presentation time of the initial fixation cross was jittered between 800 to 1300 ms to prevent expectation. The stimuli were presented via loudspeakers suspended from the ceiling of the testing chamber approximately 2 m in

front of the subjects, 0.5 m above and 1.5 m apart. The inter stimulus interval (ISI) was 2050 ms. After the offset of the second tone the fixation cross remained on the screen for another 1200 ms. The screen went black for 400 ms before a prompt to respond appeared on the screen. Participants were instructed to wait for the prompt before pressing one of two buttons to indicate if they thought the two tones they just heard were same or different. The 'same' and the 'different'-button were assigned to the right and the left hand of the participant, respectively. Assignment of hands was counterbalanced across subjects. The button press was followed by a black screen for 1500 ms before the next trial started. Small pauses between the alternating blocks allowed the participants to stretch and to rest their eyes.

6.2.5. Apparatus and recording

6.2.5.1. ERP-recording

The electroencephalogram (EEG) was recorded from 26 tin electrodes mounted in an elastic cap with reference electrodes at the left and right mastoid. A different electrode setup than that described for MMN-exp. I and II was used in the present experiment. See Fig. 6.2 for details. For orientation, positions of the international 10-20-system used in previous experiments are marked by black dots. Electrode impedance was kept below 5 kΩ. The EEG was digitized continuously at 250Hz. Electrodes were referenced on-line to the left mastoid and re-referenced off-line to the mean of the right and left mastoid electrodes. Electrodes placed at the outer canthus of each eye were used to monitor horizontal eye movements. Vertical eye movements and blinks were monitored by an electrode below the right eye referenced to the right lateral prefrontal electrode. Averages were obtained for 2048 ms epochs (including a 500 ms pre-stimulus baseline period) for both, first and second tone of a pair. Trials contaminated by eye movements or amplifier blocking or other artifacts within the critical time window were rejected prior to averaging. ERPs were calculated by time domain averaging correct trials of each participant in different conditions. The average ERPs were quantified by mean amplitude measures using the mean voltage of the 500 ms time-period preceding the onset of the

stimulus as a baseline reference. Time windows to calculate mean amplitudes for the statistical analyses were set as follows: 50-150 ms (N1), 150-250 ms (P2), 300-400 ms and 400-1000 ms. Electrode sites used for the analysis (Fig 6.2, bold prints) were midline prefrontal (MiPf), left and right lateral prefrontal (LLPf and RLPf) and medial prefrontal (LMPf and RMPf), left and right medial frontal (LMFr and RMFr), and medial central (LMCe and RMCe), midline central (MiCe), midline parietal (MiPa), left and right mediolateral parietal (LDPa and RDPa) and medial occipital (LMOc and RMOc). The resulting data were entered into an analysis of variance (ANOVA) on repeated measures with factors 'task' (emotion matching vs. identity matching), 'congruence' (=congruence vs. incongruence between 1st and 2nd tone), 'emotion' [=emotional category of the 2nd tone (happy or sad)], 'laterality' (left-lateral, left-medial, midline, right-medial and right-lateral) and 'caudality' (prefrontal, fronto-central and parieto-occipital). Separate ANOVAS were performed on data of the 4 time-windows followed by comparisons between pairs of conditions. Whenever there were two or more degrees of freedom in the numerator, the Huynh-Feldt epsilon correction was employed. Here the original degrees of freedom and the corrected p-values are reported.

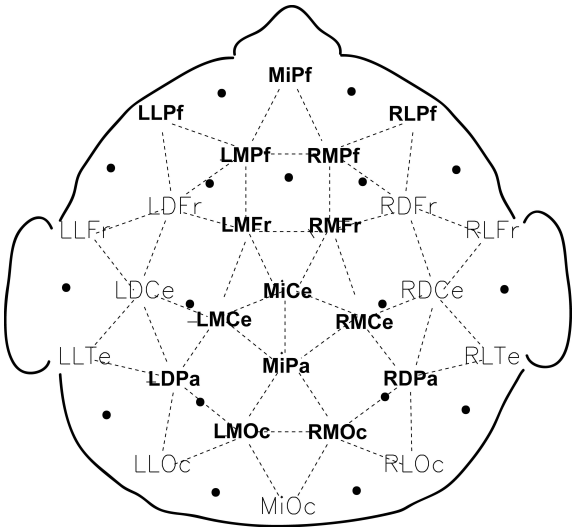

Figure 6.2.: Electrode setup used in the experiment. Dots symbolize positions of electrodes in the 10-20-system. Electrodes used in the analysis are printed in bold.

6.2.5.2. Reaction time experiment

In the additional reaction time experiment, stimuli were presented via loudspeakers (Yamaha). Emotion matching and identity matching task were alternating blockwise resulting in 8 blocks à 15 pairs per task. Order of pairs was randomized. After four blocks of both tasks were completed, hand assignment to response keys was switched in all participants. The first ten trials of the very first block as well as of the first block after switching hand assignment were not included in the analysis.

6.3. Results

6.3.1. ERP-experiment

6.3.1.1. Behavioral data

The level of performance in the emotion matching task was low but well above chance. The mean percentage of correctly matched pairs was 62.9% (SD=9.1%). Of the congruent pairs 63.9% (SD=9.8%) were correctly matched; 61.8% (SD=8.5%) of the incongruent pairs were correctly recognized as different. The mean level of performance in the identity matching task lay at 68.5% (SD=13.1%). The performance was not equally well for congruent and incongruent pairs. While the performance level reached 77.8% (SD=10.7%) for congruent pairs, incongruent pairs were only recognized correctly in 59.2% (SD=7.5%) of the presented pairs.

6.3.1.2. ERP data

The grand average waveforms to the second tone in both tasks (see Fig. 6.3) were characterized by a N1-P2-complex as typically found in auditory stimulation (Näätänen et al., 1988), followed by a long-duration component which was negative at frontal electrodes but became more and more positive toward the back of the head (main effect of caudality between 300 and 1000 ms, $F(2,26)=12.13$ and 11.59, both $p<0.001$).

Main effect of task

ERPs in the late time window were relatively more negative in the emotion matching task than in the identity matching task (see Fig. 6.3, left). Statistical analysis in the covering time windows (300-400 ms and 400-1000 ms) confirmed a main effect of task ($F(1,13)=6.08$ and 6.19, $p<0.05$).

Main effect of emotion

P2 amplitude was modulated by emotion in both task (see Fig 6.3, center). Across tasks, ERPs were more positive going for happy tones than for sad tones ($F(1,13)=9.76$, both $p<0.01$).

Main effect of congruence

A main effect of congruence in the P2 time window (150-250 ms) reflected a reduction

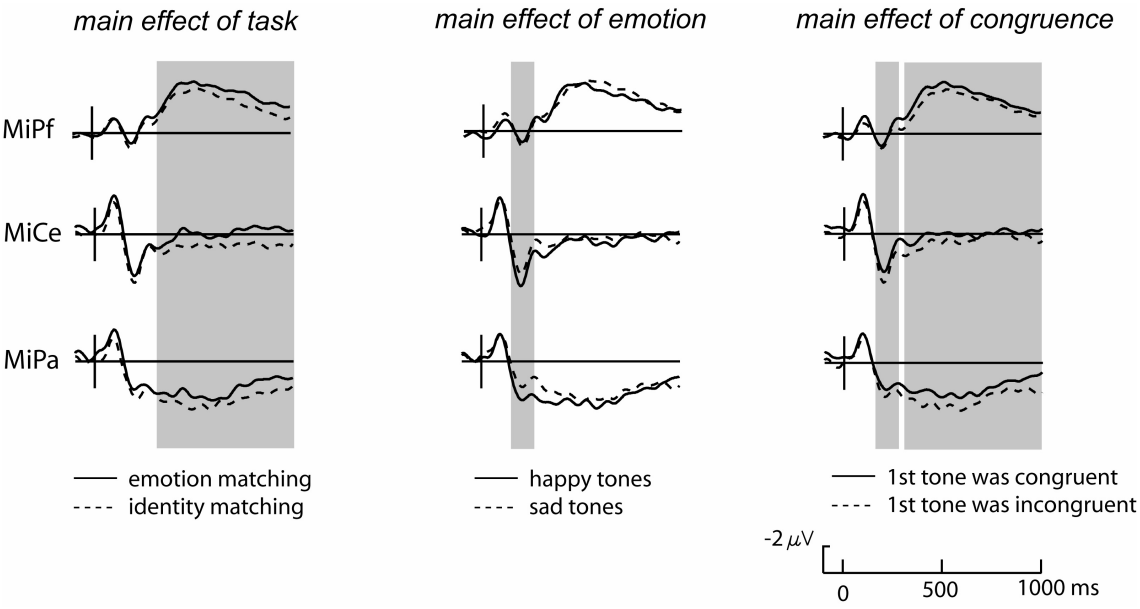

Figure 6.3.: Grand average ERPs are depicted at the midline electrodes to visualize main effects of task (left), emotion (center), and congruence (right). Time windows where effects reached significance ($p<0.05$) are highlighted. Abbreviations: MiPf=midline prefrontral, MiCe=midline central, MiPa=midline parietal.

of P2-amplitude if the tone was preceded by a congruent stimulus compared to when it was incongruent in both tasks (see Fig. 6.3, right, $F(1,13)=5.55$, $p<0.05$). The congruence effect persisted up to 1000 ms ($F(1,13)=11.07$ and 9.21, both $p<0.01$).

It has to be noted that, as a consequence of the experimental design, tones in the incongruent condition of the emotion matching task were incongruent on two levels (different emotion and different singer) while in the identity matching task incongruence was limited to the level of identity (different singer). To disentangle emotion congruence and identity congruence effects, separate ANOVAs with factors emotion of the 2nd tone, congruence, caudality, and laterality were calculated for the emotion matching and the identity matching task. Fig. 6.4 and Fig. 6.5 show ERPs separated for the emotion and the identity matching task.

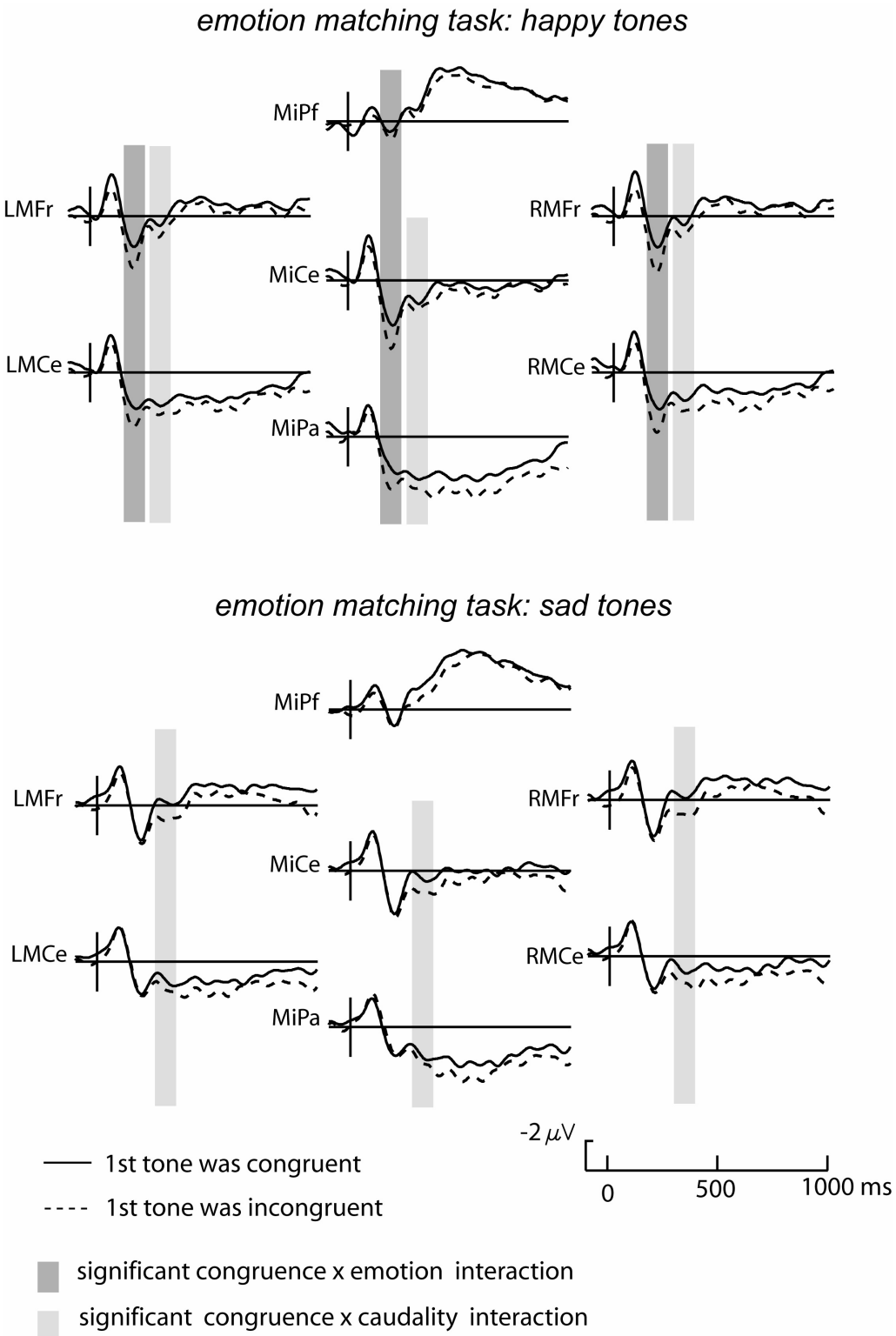

Figure 6.4.: Grand average ERPs to the 2nd tone of a pair in the emotion matching task. Conditions are depicted separately for happy tones (top) and sad tones (bottom). ERPs to emotionally congruent pairs (bold line) are superposed by ERPs to emotionally incongruent pairs (dashed line). Time windows where effects reached significance ($p<0.05$) are highlighted differently for different effects.

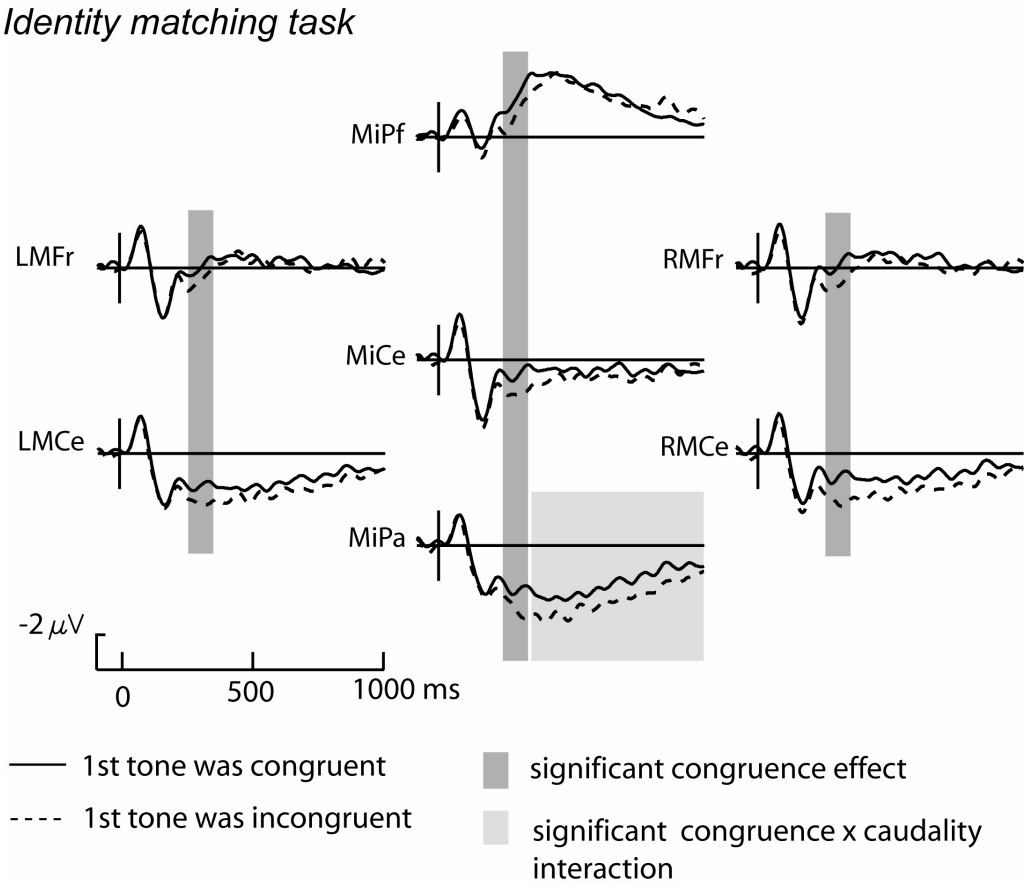

Figure 6.5.: Grand average ERPs to the 2nd tone of a pair in the identity matching task. ERPs to pairs of congruent identity (bold line) are superposed by ERPs to pairs of incongruent identity (dashed line). Since no interaction with emotion was found, ERPs are depicted collapsed for happy and sad tones. Time windows where effects reached significance ($p<0.05$) are highlighted.

In the emotion matching task, irrespective of the emotional category of the 2nd tone of the pair, an effect of congruence was found between 300 and 400 ms, though only at fronto-central and parieto-occipital electrodes (interaction congruence x caudality, see table 6.2 for F-values). However, if the emotional expression (of the 2nd tone) was happy, the P2-amplitude was reduced in the congruent condition compared to the incongruent condition (interaction congruence x emotion, see Fig. 6.4, top). No such early effect of congruence was found in the sad condition (Fig. 6.4, bottom).

In the identity matching task (see Fig. 6.5), an effect of stimulus congruence started later than in the emotion matching task (at 300 ms). Tones which were preceded by a tone sung by a different singer evoked a more positive going ERP than if the preceding tone was sung by the same singer. The congruence effect continued in the consecutive time window but was restricted to electrodes at the back of the head (interaction congruence x caudality). Congruence and emotion did not interact, though emotion modulated the P2 amplitude (main effect of emotion, see table 6.2). The P2-amplitude was larger to happy than to sad tones.

Table 6.2.: Results of the separate ANOVAs on ERP-data in the emotion matching (top) and the identity matching task (bottom); only significant interactions are reported; given are the F-values. [$^{***}p< 0.001$, $^{**}p< 0.01$, $^{*}p< 0.05$], abbreviations: congr=congruence, emo=emotion, caud=caudality.

emotion matching					
		50-150ms	150-250ms	300-400ms	400-1000ms
emotion	F(1,13)	x	x	x	x
congruency	F(1,13)	x	x	x	x
caudality	F(2,26)	x	x	11.18^{***}	8.25^{***}
laterality	F(4,52)	x	x	x	x
congr x emo	F(1,13)	x	5.65^{*}	x	x
congr x caud	F(2,26)	x	x	3.66^{*}	x
identity matching					
		50-150ms	150-250ms	300-400ms	400-1000ms
emotion	F(1,13)	x	5.54^{*}	x	x
congruency	F(1,13)	x	x	4.72^{*}	x
caudality	F(2,26)	x	x	12.64^{***}	11.67^{***}
laterality	F(4,52)	x	x	x	x
congr x emo	F(1,13)	x	x	x	x
congr x caud	F(2,26)	x	x	x	4.31^{*}

To summarize, an earlier effect of congruence was found in the emotion matching than in the identity matching task, though only for happy tones. P2-amplitude was found reduced if the happy voice stimulus had been preceded by another happy voice stimulus. No such early priming effect was found in the identity matching task.

It was thought possible that the P2-priming effect was a consequence of the acoustical similarity of the first and the second tone in the congruent happy condition, thus reflecting a mere physical priming effect (Wiggs & Martin, 1998). If this was the case, the same kind of amplitude reduction should be found in the identical stimulus set of the identity matching task. Though an incomplete factorial design had been used, the pairing of tones in the incongruent condition of the identity matching task equaled that of the congruent condition in the emotion matching task (see schema 6.6). In both conditions the singer was different between tone 1 and 2 while the emotion was kept stable. The emotion-specific acoustical structure of tone 1 and 2 should thus be equally the same in both tasks. ERPs for the two conditions are superposed in Fig. 6.7 and compared to the incongruent emotion condition (dashed line). No effect of priming (as reflected by reduced P2-amplitude) was found for the identity matching task. The statistical comparison of mean amplitudes in both tasks corroberated the difference ($F(1,13)=14.49$, $p=0.0022$). The difference waves were calculated for both tasks by subtracting the ERP to the incongruent condition from the ERP to the congruent condition (see Fig. 6.8). It demonstrates the timing difference of the congruence effects in both tasks.

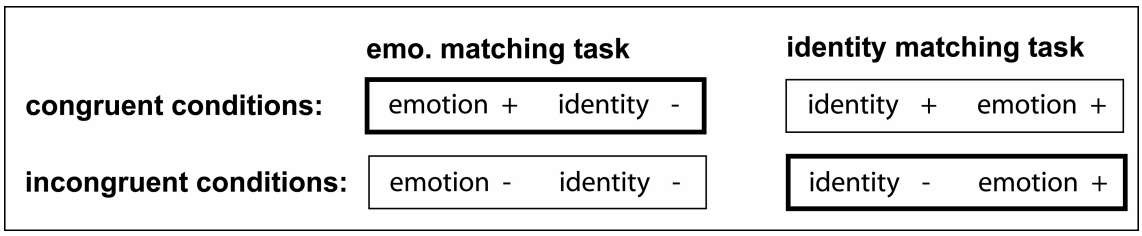

Figure 6.6.: Schema of the experimental design. In the emotion matching task, the identity of the singer was different (-) in both, the congruent and the incongruent condition. In contrast, in the identity matching task, the emotion was same (+) in the congruent and the incongruent condition. As a consequence, the way stimulus pairs were constructed was identical in the congruent condition of the emotion matching task and the incongruent condition of the identity matching task.

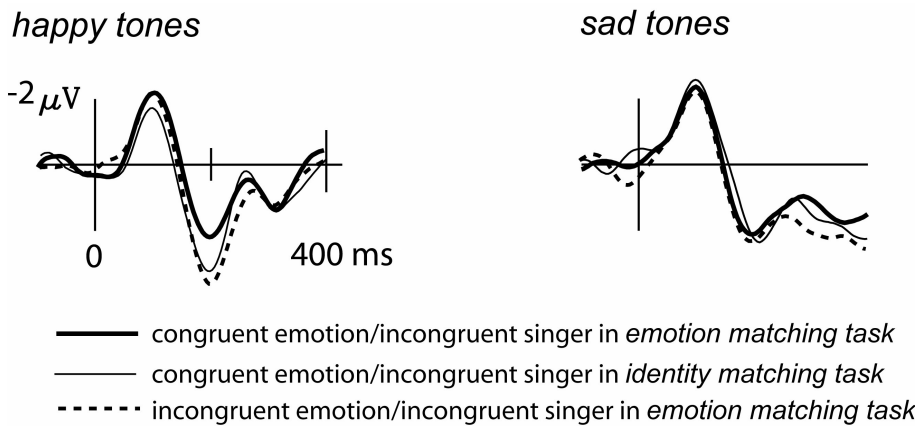

Figure 6.7.: P2-congruence effect in the emotion matching task. Grand average ERPs in the P2-time window are overlapped for three different conditions.

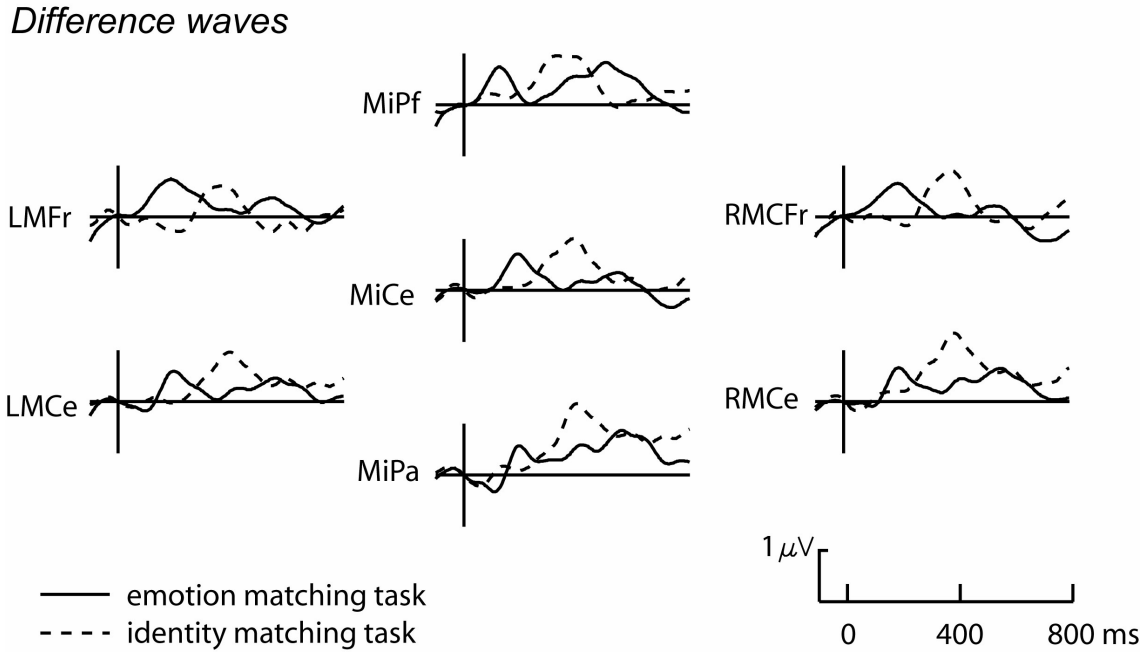

Figure 6.8.: Difference waves for the emotion matching (bold line) and the identity matching (dashed line) task were calculated by subtracting the ERP to the incongruent condition from the ERP to the congruent condition.

Results of the reaction time experiment Reaction times were fed into an ANOVA of repeated measures with factors task (emotion, identity), congruency (same, different) and emotional expression of the second tone of pairs (happy, sad). No main effect of task was found for the emotion matching vs. the identity matching task (1311 vs. 1284 ms, $F(1,9)=0.145$, $p=0.712$). However, an interaction of task and emotion ($F(1,9)=9.4$, $p=0.013$) reflected the fact that reaction was faster in the emotion than in the identity matching task if the second tone of a pair was happy ($p<0.01$). This difference also caused a main effect of emotion (1256 for happy vs. 1339 ms for sad items, $F(1,9)=7.007$, $p=0.027$). Reaction times separated for all conditions are given in table 6.3 and graphically depicted in Fig. 6.9.

Table 6.3.: Results of the reaction time experiment; given are means and (in brackets) standard deviations.

Emotion same [Identity different]		Emotion different [Identity different]		Identity same [Emotion same]		Identity different [Emotion same]	
happy	sad	happy	sad	happy	sad	happy	sad
1270 (341)	1374 (411)	1192 (292)	1407 (410)	1293 (504)	1261 (412)	1268 (321)	1315 (406)

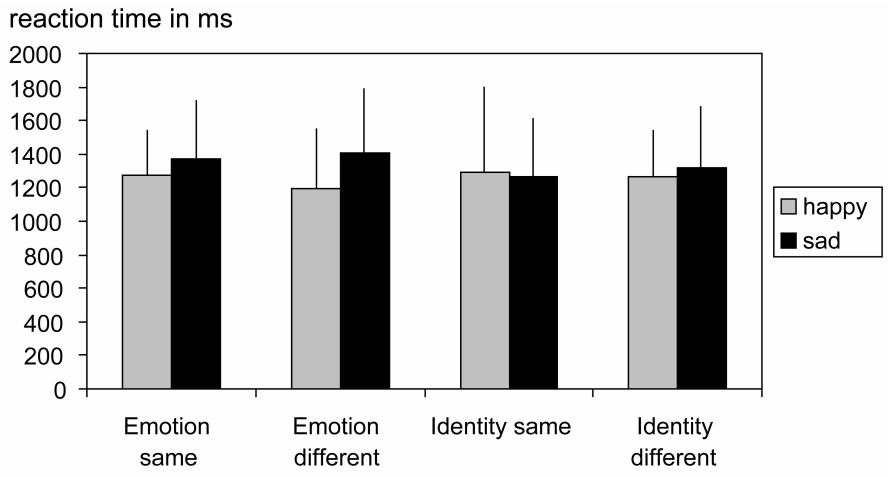

Figure 6.9.: Mean reaction times are depicted separated for conditions (emotion same and different, identity same and different) and emotion (happy=grey, sad=black).

6.4. Discussion

The ERP-data imply that extracting emotional information from the voice happens faster than voice recognition. Thus the results do not parallel the finding of Münte et al. (1998) that identity recognition of faces precedes recognition of emotional expression. It is however possible that emotion and identity recognition are quite different matters in face and voice perception. It is assumed that human vocalization evolved to facilitate interaction in larger groups (Dunbar, 1996), including the necessity to communicate via large distances. Acoustic waves travel far and are a more reliable signal over distances than visual signs. Some species dispose of a sophisticated vocal warning system. The acoustical structure of distinct calls gives the peers information on the exact nature of an upcoming threat, e.g. whether the predator is a leopard or an eagle (Hauser, 1997). Since humans cannot rely on such a refined mechanism to interpret voice quality, it might be safer for them to run before taking the time to figure out who expressed the warning call.

Though the notion that the timing differences found in the present experiment reflect phylogenetically old mechanisms for survival has to remain speculation, the data provides support for Belin et al.'s model of voice processing (Belin et al., 2004). Together with fMRI findings that both functions activate different areas of the brain (Imaizumi et al., 1997), the differences in timing confirm the assumption that different processes underlie the recognition of vocal identity and vocal affect.

However, the effect of emotional priming was earlier in response to happy than to sad voices. Repeated presentation of happy voice stimuli reduced the P2 amplitude. Interestingly, an emotion-specific sensitivity of the P2 amplitude has also been reported by others (Schapkin, Gusev, & Kuhl, 2000; Alter et al., 2003; Holmes, Kiss, & Eimer, 2006). In the present experiment the earlier ERP effect for happy voices was paralleled by a shorter mean reaction time in the recognition of happy compared to sad tones in the emotion matching task. It seems that happiness could more easily be extracted from the tones than sadness. Shorter reaction times to positive than to negative prosody have been reported by Schirmer et al. (2004) though their negative prosody was angry

not sad. Shorter reaction times were found in response to happy than to sad targets in experiment MMN-I described in chapter 4 of this thesis. The difference can most likely be linked to the generally faster tone attack found in happy tones. Attack has been identified as important cue in musical emotion coding (see section 2.3.3). Because sad tones are characterized by a longer rise-time than happy tones (Gabrielsson & Juslin, 1996) it might take longer until they are recognized as emotionally significant.

The present experiment, for the first time, delineated the time course of two cognitive functions, highly important for social interaction: to recognize a person from the voice, and to make inferences on a speaker's emotional state. Both could be performed on very short samples. Given that the listeners could not rely on suprasegmental acoustical aspects such as tempo or rhythm, the results underline the important role timbre or voice quality plays for recognizing a person's 'auditory face'.

7. Integration of visual and auditory emotional stimuli

7.1. Introduction

Judging the emotional content of a situation is a daily occurrence that typically necessitates the integration of inputs from different sensory modalities - especially, vision and audition. Although the combined perception of auditory and visual inputs has been studied for some years (McGurk & MacDonald, 1976, Welch & Warren, 1986, Stein & Meredith, 1993, see also Calvert, 2001, and Thesen, Vibell, Calvert, and Österbauer, 2004, for reviews), the multisensory perception of emotion has only relatively recently come into focus. Those studies investigating the integration of affective information have typically used emotional faces paired with emotionally spoken words (Massaro & Egan, 1996; Gelder, Bocker, Tuomainen, Hensen, & Vroomen, 1999; Gelder & Vroomen, 2000; Pourtois, Gelder, Vroomen, Rossion, & Crommelinck, 2000; Balconi & Carrera, 2005). Behaviorally, face-voice-pairs with congruent emotional expressions have been found to be associated with increased accuracy and faster responses for emotion judgments compared to incongruent pairs. Massaro and Egan (1996) , for example, used a computer-generated "talking-head" with a male actor's voice saying 'please' in a happy, neutral or angry way while the head's face displayed either a happy, neutral or angry expression. Participants made two-alternative forced choice judgments (happy or angry) on the audio-visual percept. Reaction times increased with the degree of ambiguity between the facial and vocal expressions. The probability of judging the audio-visual performance as angry was calculated for all conditions based on participants' responses. Overall, facial expression had a larger effect on judgments than the voice. However,

when the facial expression was neutral, the combined percept was influenced considerably by the expression of the voice. The authors concluded that the influence of one modality on the emotion perception depended to a large extent on how ambiguous or undefined affective information in that modality was. De Gelder and Vroomen (2000), found an overall larger effect of voice on the ratings of voice-face presentations than that reported by Massaro and Egan (1996). Besides a possible difference between angry and sad faces with respect to salience, the different visual presentation formats may help account for the somewhat different results. Specifically, the use of moving faces by Massaro and Egan may have led to visual dominance as in the ventriloquism effect[1] (Stein & Meredith, 1993). This possibility is supported by de Gelder and Vroomen's observation that the effect of voice was reduced, although not completely eliminated when participants were instructed to selectively attend the face and ignore the voice. They also confirmed Massaro and Egan's finding that voice information had a greater impact when facial expressions were ambiguous.

Of particular interest in the realm of audio-visual integration is the question of timing, namely, when in the processing stream does the integration actually take place? Using event-related brain potentials (ERP) to examine the time-course of integrating emotion information from facial and vocal stimuli Pourtois et al. (2000) found a sensitivity of the auditory N1 ($\sim$110 ms) and P2 ($\sim$200 ms) components to the multisensory input: N1 amplitudes were increased in response to attended angry or sad faces that were accompanied by voices expressing the same emotion, while P2 amplitudes were smaller for congruent face-voice pairs than for incongruent pairs. By presenting congruent and incongruent affective face-voice pairs with unequal probabilities, de Gelder et al. (1999) evoked auditory mismatch negativities (MMN) in response to incongruent pairs as early as 178 ms after voice onset. Both these results suggest that interactions between affective information from the voice and the face take place before either input has been fully processed.

[1] The ventriloquism effect refers to the perception of sounds as coming from a direction other than their true direction (e.g. from an apparently speaking puppet), due to the influence of visual stimuli from an apparent sound source (the puppet's moving mouth).

Considerably less effort has been directed toward the integration of emotional information from more abstractly related inputs as they typically occur in movies, commercials or music videos (but see de Gelder, Vroomen, and Pourtois, 2004, for discussion). Though music has been found to be suitable to alter a film's meaning (Marshall & Cohen, 1988; Bolivar, Cohen, & Fentress, 1994), few attempts have been made to study the mechanisms involved in the integration of emotion conveyed by music and visually complex material. We assume that integration of complex affective scenes and affective auditory input takes place later than integration of emotional faces and voices because the affective content of the former is less explicit and less salient and thereby requires more semantic analysis before their affective meaning can begin to be evaluated. Although earlier components such as the N2 have been reported to be sensitive to emotional picture valence (e.g. Palomba & Angrilli, 1997), the most commonly reported ERP effect is modulation of the visual P3 amplitude: pictures of pleasant or unpleasant content typically elicit a larger P3 (300-400 ms) and subsequent late positive potential (LPP) than neutral pictures (Johnston et al., 1986; Diedrich, Naumann, Maier, Becker, & Bartussek, 1997; Palomba, Angrilli, & A, 1997; Schupp et al., 2000). LPP amplitude also has been found to vary with the degree of arousal; both, pleasant and unpleasant pictures with highly arousing contents elicit larger LPP-amplitudes than affective pictures with low arousal (Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000). The finding that affective (compared to non-affective) pictures elicit a pronounced late positive potential which is enlarged by increasing arousal has been taken to reflect intensified processing of emotional information that has been categorized as significant to survival (Lang et al., 1997). The P3 in such studies has been taken to reflect the evaluative categorization of the stimulus (Kayser, Bruder, Tenke, Stewart, & Quitkin, 2000).

Support for the notion that an integration of affective pictures of complex scenes and affective voices takes place later than integration of affective faces and voices comes from the demonstration that the auditory N1 to fearful voices is modulated by facial expressions even in patients with striate cortex damage who cannot consciously perceive the facial expression (Gelder, Pourtois, & Weiskrantz, 2002). In contrast, pictures of

emotional scenes did not modulate early ERP components even though the patients' behavioral performance indicated that the picture content had, though unconsciously, been processed. The authors suggested that while non-striate neural circuits alone might be able to mediate the combined evaluation of face-voice pairs, integrating the affective content from voices and pictures is likely to require that cortico-cortical connections with extrastriate areas needed for higher order semantic processing of the picture content be intact.

To examine the time-course of integrating affective scene-voice pairs in healthy subjects, event-related brain potentials (ERP) were recorded while simultaneously presenting affective and neutral pictures with musical tones sung with emotional or neutral expression. The aim of the study was to assess when and to what extent the processing of affective pictures is influenced by affective information from the voice modality. In addition, the relative importance of attention to this interaction was examined by directing participants' attention to either the picture modality or the voice modality.

It was hypothesized that affective information in the auditory modality can facilitate as well as impede processing of affective information in the visual modality depending on whether the emotion expressed in the voice matches the picture valence or not. Presumably congruent information enhances stimulus salience, while incongruent information leads to an ambiguous percept, thereby reducing stimulus salience. Given what is known from investigations of affective picture processing as well as from picture-voice integration in patients with striate damage, it was expected that integration would not become manifest in ERP-components before 300 ms post stimulus onset. Rather, it was thought more likely that the simultaneously presented auditory information would have a modulating effect on the P3 and the subsequent late positive potential, assuming that significance of the pictures would be influenced by related additional information. There was no clear hypothesis for what to expect when participants attended to the voice instead of the picture. The amplitude of the P3 to auditory (non-affective) oddball target stimuli co-occurring with visual stimuli is smaller in conjunction with affective faces (Morita, Morita, Yamamoto, Waseda, & Maeda, 2001) and affective pictures (Schupp,

Cuthbert, Bradley, Birbaumer, & Lang, 1997) than with neutral visual stimuli. Such results have been interpreted as reflecting a re-allocation of attentional resources away from the auditory input to the affective pictures. Thus, it was considered possible that the ERP pattern obtained in the attend-voice-task would differ significantly from that in the attend-picture-task.

7.2. Materials and Methods

7.2.1. Stimuli

7.2.1.1. Picture stimuli

Picture stimuli were 22 happy, 22 neutral, and 22 sad pictures from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1995)[2].

Because the experimental setup required that the pictures be presented for very short durations (300-515 ms), a pre-experiment was conducted to assure that the pictures could still be recognized and evaluated similarly to the reported ratings (Lang, Bradley, & Cuthbert, 1995) even with presentation times as short as 300 ms. In the pre-experiment a larger pool of IAPS pictures (30 per emotion category) was presented to 5 volunteers from the lab (all PhD students, age 25 to 30 yrs, 4 female) with duration times randomized between 302 and 515 ms. Participants were asked to rate the pictures on their emotional valence and on their arousal on a 7-point scale. Participants were additionally asked to note whenever they thought the picture was too hard to recognize or too shocking. Pictures were excluded whenever any one participant's valence rating did not match Lang et al.'s rating (e.g. happy instead of sad or vice versa) or whenever anyone noted that a picture was too difficult to recognize or repulsive. The mean valence ratings of the remaining 22 pictures per category were 5.90 (SD= 0.39) for happy pictures, 4.02 (SD=0.36) for neutral pictures, and 1.80 (SD=0.58) for sad pictures. Valence-ratings

[2]Pictures used from the IAPS were 1463, 1610, 1710, 1920, 2040, 2057, 2080, 2150, 2160, 2311, 2340, 2530, 2550, 2660, 4220, 5480, 5760, 5910, 7580, 8190, 8470, 8540, 2840, 2880, 2890, 7160, 4561, 5510, 5531, 6150, 7000, 5920, 7002, 7004, 7009, 7010, 7020, 7035, 7050, 7185, 7233, 7235, 7950, 8160, 2205, 2710, 2750, 2800, 2900, 3180, 3220, 3230, 3350, 6560, 6570, 9040, 9050, 9181, 9220, 9340, 9421, 9433, 9560, 2590, 2661, 3300.

among the three categories differed significantly as tested with an one-way ANOVA ($F(2,63)=447.27$, $p<0.001$) and post-hoc Scheffé tests ($p<0.001$ for all comparisons). Analogous to Lang et al. (1995), arousal ratings were larger for both happy and sad than for neutral pictures (4.29 (SD=0.82), and 4.07 (SD=0.84) vs. 2.15 (SD=1.21); $F(2,63)=31.78$, $p<0.001$; post-hoc (Scheffé): $p<0.001$ for sad vs. neutral and for happy vs. neutral).

7.2.1.2. Voice stimuli

Voice stimuli were generated from 10 professional opera singers and advanced singing students (5 women) asked to sing the syllable 'ha' with a happy, sad or neutral tone (see experiment II-01). From 200 different tones, twenty-two were selected for each emotional category based on the valence ratings of 10 raters (age 21-30, 5 female) on a 7-point scale (1 = extremely sad to 7 = extremely happy). The selected stimuli met the following criteria: their mean ratings were within the category boundaries (rating <3 sad, >5 happy, between 3 and 5 neutral) and they were consistently rated as happy (responses had to be 5,6 or 7), neutral (responses had to be 3,4 or 5), or sad (responses had to be 1,2 or 3) by at least 7 of 10 raters. All tones were also rated by these same participants for arousal on a 7-point scale (1 = 'not arousing at all' to 7 = 'extremely arousing'). Mean valence ratings by category, were 5.23 (SD=0.35) for happy, 3.91 (SD=0.28) for neutral, and 2.81 (SD=0.44) for sad notes. Mean ratings between all three categories were significantly different as tested with an one-way ANOVA ($F(2,63)=247.03$, $p<0.000$) and post-hoc Scheffé tests ($p<0.000$ for all comparisons). Mean arousal ratings for happy, neutral and sad notes on a 7-point scale were 2.62 (SD=0.37), 2.18 (SD=0.28), and 2.51 (SD=0.27), respectively. As for pictures, arousal ratings were higher for both happy and sad than for neutral tones ($F(2,63)=12.07$, $p<0.00$; post-hoc (Scheffé): $p<0.01$ for sad vs. neutral and for happy vs. neutral). Between valence categories, notes were matched for length (mean=392 ms, SD=60 ms) and pitch level (range: A^2 to A^4). A total of 66 voice stimuli was digitized with a 44.1 kHz sampling

rate and 16 bit resolution. The amplitude of all sounds was normalized to 90% so the maximum peak of a waveform was equally loud across all the tones.

7.2.1.3. Picture-voice-pairings

Picture and voice stimuli were combined such that each picture was paired once with a happy, once with a neutral, and once with a sad voice. Likewise, each voice stimulus was paired with a happy, a neutral and a sad picture. Thus, all pictures and all sung tones were presented three times, each time in a different combination. Picture-voice-pairs were created randomly for each participant. To increase the overall number of trials, the resulting set of 198 pairs was presented twice in the experiment, each time in a different randomized order.

7.2.2. Participants

Fourteen right-handed students of the University of San Diego (UCSD; age range 18-27 yrs, mean=21 yrs (SD=2.75), 8 women) received either money or course credit for their participation in the experiment. None of the participants considered him- or herself a musician, though some reported having learned to play a musical instrument at some point of their life. Participants gave informed consent and the study was approved by the UCSD Human Subjects' Internal Review Board. Prior to the experiment participants were given a hearing test to allow for an individual adjustment of audio volume.

7.2.3. Task procedure

Participants were tested in a sound attenuating, electrically-shielded chamber. They were seated 127 cm in front of a 21-inch computer monitor. Auditory and visual stimuli were presented under computer control. Each trial started with a black screen for 1600 ms. Picture and voice pairs were presented simultaneously following the presentation of a crosshair, orienting participants toward the centre of the screen. The interval between cross onset and stimulus onset was jittered between 800 and 1300 ms to reduce temporal predictability. Voice stimuli were presented via two loudspeakers suspended from the

ceiling of the testing chamber approximately 2 m in front of the subjects, 0.5 m above and 1.5 m apart. Each picture remained on screen as long as the concomitant auditory stimulus (ranging from 302-515 ms) lasted. Pictures subtended 3.6 x 6.3 degrees of visual angle (width x height).

Two different tasks were alternated between blocks. In the *attend picture task*, participants were asked to rate picture valence on a 7-point scale (ranging from 1 = very sad to 7 = very happy) while ignoring the voice stimulus. In the *attend voice task*, participants were asked to rate the emotional expression of the voice (sung tone) on the same scale while ignoring the picture stimulus. Participants gave their rating orally after a prompt to do so appeared on the screen 1500 ms after stimulus offset. After their response had been registered, the next trial was started manually by the experimenter. Trial durations ranged between 4102 and 4815 ms. Order of task blocks was counterbalanced. Prior to the experiment, participants took part in a short practice block to familiarize them with the experimental procedures.

7.2.4. ERP recording

The electroencephalogram (EEG) was recorded from 26 tin electrodes mounted in an elastic cap (see setup in the previous chapter, Fig. 6.2) with reference electrodes at the left and right mastoid. Electrode impedance was kept below 5 kΩ. The EEG was processed through amplifiers set at a bandpass of 0.016-100 Hz and digitized continuously at 250Hz. Electrodes were referenced on-line to the left mastoid and re-referenced off-line to the mean of the right and left mastoid electrodes. Electrodes placed at the outer canthus of each eye were used to monitor horizontal eye movements. Vertical eye movements and blinks were monitored by an electrode below the right eye referenced to the right lateral prefrontal electrode. Averages were obtained for 2048 ms epochs including a 500 ms pre-stimulus baseline period. Trials contaminated by eye movements or amplifier blocking or other artifacts within the critical time window were rejected prior to averaging.

ERPs were calculated by time domain averaging for each subject and each valence combination (picture-voice: happy-happy, happy-neutral, happy-sad, neutral-happy, neutral-neutral, neutral-sad, sad-happy, sad-neutral, and sad-sad) in both tasks (voice rating, picture rating).

These average ERPs were quantified by mean amplitude measures using the mean voltage of the 500 ms time-period preceding the onset of the stimulus as a baseline reference. Time windows for the statistical analyses were set as follows: N1 (50-150 ms), P2 (150-250 ms), N2 (250-350 ms), P3 (380-420 ms) and N2b (420-500 ms), followed by a sustained late positive potential (LPP, 500-1400 ms). Electrode sites used for the analysis (Fig. 6.2, bold prints) were midline prefrontal (MiPf), left and right lateral prefrontal (LLPf and RLPf) and medial prefrontal (LMPf and RMPf), left and right medial frontal (LMFr and RMFr), and medial central (LMCe and RMCe), midline central (MiCe), midline parietal (MiPa), left and right mediolateral parietal (LDPa and RDPa) and medial occipital (LMOc and RMOc).

The resulting data were entered into ANOVAs. Separate ANOVAs on 4 repeated measures with within factors $\text{valence}_{\text{att}}$ [= valence in the attended modality (happy, neutral, sad)], $\text{valence}_{\text{unatt}}$ [= valence in the unattended modality (happy, neutral, sad)], 'laterality' (left-lateral, left-medial, midline, right-medial and right-lateral) and 'caudality' (prefrontal, fronto-central and parieto-occipital) were conducted on data from each task, followed by comparisons between pairs of conditions. To test for effects of task an additional ANOVA on 3 repeated measures [two levels of task (picture rating, voice rating), 5 levels of laterality (left-lateral, left-medial, central, right-medial and right-lateral) and 3 levels of caudality (prefrontal, fronto-central and parieto-occipital)] was performed.

Whenever there were two or more degrees of freedom in the numerator, the Huynh-Feldt epsilon correction was employed. Here the original degrees of freedom and the corrected p-values are reported.

7.3. Results

7.3.1. Behavioral results

Separate ANOVAs on two repeated measures (factor $\text{valence}_{\text{att}}$ [=valence in the attended modality (happy, neutral, sad)] and factor $\text{valence}_{\text{unatt}}$ [= valence in the unattended modality (happy, neutral, sad)]) were conducted for both rating tasks (for mean ratings and standard deviations in the 9 different conditions per task see table 7.1). In the *attend picture task* a significant main effect of valence of the attended modality was found with mean ratings for happy, neutral and sad pictures being 5.71, 3.94 and 2.19, respectively ($\text{valence}_{\text{att}}$ $F(2,26)=356.4$, $p<0.001$). Posthoc analysis (Scheffé) revealed all categories differed significantly from each other (all $p<0.01$). There was no main effect of the emotion expressed by the unattended voice stimuli on picture valence ratings ($\text{valence}_{\text{unatt}}$ $F(2,26)=2.14$, $p=0.15$) and picture valence and voice valence did not interact ($F(4,52)=0.58$, $p=0.64$).

Table 7.1.: Mean valence ratings for pictures in the attend-picture-task (left) and for voices in the attend-voice-task (right) for all possible picture-voice-combinations.

Attend-picture-task			Attend-voice-task		
picture valence	voice valence	picture rating mean (SD)	voice valence	picture valence	voice rating mean (SD)
happy	happy	5.77 (0.42)	happy	happy	5.07 (0.38)
happy	neutral	5.72 (0.45)	happy	neutral	4.79 (0.44)
happy	sad	5.65 (0.55)	happy	sad	4.63 (0.53)
neutral	happy	3.92 (0.21)	neutral	happy	4.06 (0.33)
neutral	neutral	3.92 (0.15)	neutral	neutral	4.01 (0.31)
neutral	sad	3.90 (0.20)	neutral	sad	3.65 (0.47)
sad	happy	2.19 (0.41)	sad	happy	3.79 (0.45)
sad	neutral	2.20 (0.38)	sad	neutral	3.61 (0.32)
sad	sad	2.18 (0.33)	sad	sad	3.42 (0.42)

In the *attend voice task* mean ratings for happy, neutral and sad voice stimuli also differed as expected (4.83, 3.91, and 3.61, respectively; $\text{valence}_{\text{att}}$ $F(2,26)=68.5$, $p<0.001$).

Posthoc-analysis (Scheffé) revealed significant differences between all three categories (all $p<0.001$). In contrast to the picture valence ratings, however, there was a significant main effect of the valence of the concurrently presented unattended picture on voice valence ratings ($valence_{unatt}$ $F(2,26)=14.0$, $p<0.001$). Happy voice stimuli were rated more positive when paired with a happy picture than when paired with a sad picture (5.07 vs. 4.63; $t(13)=4.77$, $p<0.01$). The same was true for neutral voice stimuli (4.06 vs. 3.65; $t(13)=2.72$, $p<0.05$). No reliable influence of picture valence was observed for sad voice stimuli. Nevertheless, voice valence and picture valence did not interact ($F(4,52)=1.10$, $p=0.36$).

7.3.2. ERP data

7.3.2.1. Valence effect

7.3.2.1.1. Attend-picture-task

Effect of (attended) picture valence ERPs recorded in the attend-picture-task are depicted in Fig. 7.1. Responses to neutral, happy and sad pictures collapsed across voice valence are superimposed. Picture valence affected the amplitude of P2, P3 and N2b ($valence_{att}$ $F(2,26)=8.86$, 4.76, 7.23, all $p<0.05$) as well as the LPP ($F(2,26)=18.78$, $p<0.00$). Pair wise comparisons revealed that P2 was more pronounced for happy pictures than for neutral ($F(1,13)=36.64$, $p=0.000$) and sad ($F(1,13)=5.42$, $p=0.037$) pictures. Since P3, N2b and LPP effect interacted with caudality ($F(4,52)=6.86$, 3.75, and 3.53, all $p<0.01$), pair wise comparisons were conducted separately at prefrontal, fronto-central and parieto-occipital sites (see table 7.2 for F-values). Starting at 380 ms, the ERP was more positive going for happy pictures than for neutral and sad pictures at prefrontal sites. The pattern changed towards the back of the head and at parieto-occipital electrodes where both happy and sad pictures elicited equally greater positivities than did neutral pictures.

Effect of (unattended) voice valence To determine what effect(s) the valence of the unattended voice stimuli had on the brain response to picture stimuli, ERPs elicited

Table 7.2.: Pairwise comparison of ERP-averages to pictures of different valence in the attend-picture- (top) and the attend-voice-task (bottom). Given are significant F-values (df=1,13) for comparison of mean amplitudes in the P3 (380-420 ms), N2b (420-500 ms) and LPP (500-1400 ms) time window at three levels of caudality (prefrontal, fronto-central and parieto-occipital), *p<0.05, **p<0.01, ***p<0.001, n.s.=not significant.

Attend-picture-task				
		380-420ms	420-500ms	500-1400ms
prefrontal	happy-neutral	7.15^{*}	14.17^{**}	8.86^{*}
	happy-sad	9.69^{**}	8.27^{*}	6.88^{*}
	neutral-sad	n.s.	n.s.	n.s.
fronto-central	happy-neutral	n.s.	22.17^{***}	81.23^{***}
	happy-sad	11.16^{**}	n.s.	n.s.
	neutral-sad	n.s.	n.s.	29.59^{***}
parieto-occ.	happy-neutral	8.45^{*}	23.96^{***}	18.00^{**}
	happy-sad	n.s.	n.s.	n.s.
	neutral-sad	6.41^{*}	7.56^{*}	21.19^{**}
Attend-voice-task				
		380-420ms	420-500ms	500-1400ms
prefrontal	happy-neutral	n.s.	n.s.	n.s.
	happy-sad	10.10^{**}	12.96^{**}	n.s.
	neutral-sad	25.54^{***}	19.49^{**}	6.29^{*}
fronto-central	happy-neutral	7.33^{*}	n.s.	n.s.
	happy-sad	n.s.	n.s.	n.s.
	neutral-sad	22.53^{***}	8.00^{*}	n.s.
parieto-occ.	happy-neutral	4.98^{*}	n.s.	n.s.
	happy-sad	n.s.	n.s.	9.65^{**}
	neutral-sad	5.11^{*}	n.s.	14.69^{**}

by pictures paired with different valence voices were superimposed separately for happy, neutral and sad pictures (shown for 3 midline sites in Fig. 7.2). A valence effect of the unattended voice modality was found for the N1 component; this effect varied with electrode location ($valence_{unatt}$ x caudality $F(4,52)=3.90$, $p<0.01$). At parieto-occipital sites pairing with sad voices led to reduction of the N1 amplitude compared to pairing with neutral ($F(1,13)=11.43$, $p=0.005$) or happy voices ($F(1,13)=8.86$, $p=0.011$). A main effect of voice valence was found for the P2 component ($valence_{unatt}$ $F(2,26)=3.56$, $p=0.043$). P2 amplitudes were larger for all pictures paired with happy than sad voices

($F(1,13)=5.72$, $p=0.033$) or with neutral voices (although this difference was marginally significant: $F(1,13)=3.93$, $p=0.069$). At fronto-central electrodes congruent pairings of happy pictures with happy voices yielded the largest P2 amplitude overall (compared to sad picture/happy voice: $F(1,13)=10.05$, $p=0.007$, and neutral picture/happy voice $F(1,13)=36.02$, $p<0.000$; interaction $valence_{att}$ x $valence_{unatt}$ x caudality $F(8,104)=2.08$, $p=0.044$). Finally, attended picture modality interacted with unattended voice modality between 500 and 1400 ms ($valence_{att}$ x $valence_{unatt}$ $F(4,52)=2.72$, $p=0.040$). This LPP was only affected by voice valence when sad pictures were presented. It was more pronounced in combination with a sad than with a neutral voice stimulus ($F(1,13)=22.40$, $p=0.000$). At prefrontal electrodes, sad pictures paired with happy voices, also led to a more pronounced LPP than when paired with neutral voices (interaction with caudality ($F(2,26)=3.54$, $p<0.05$), but pair-wise comparison at pre-frontal electrodes did not reach significance ($F(1,13)= 3.43$, $p=0.087$).

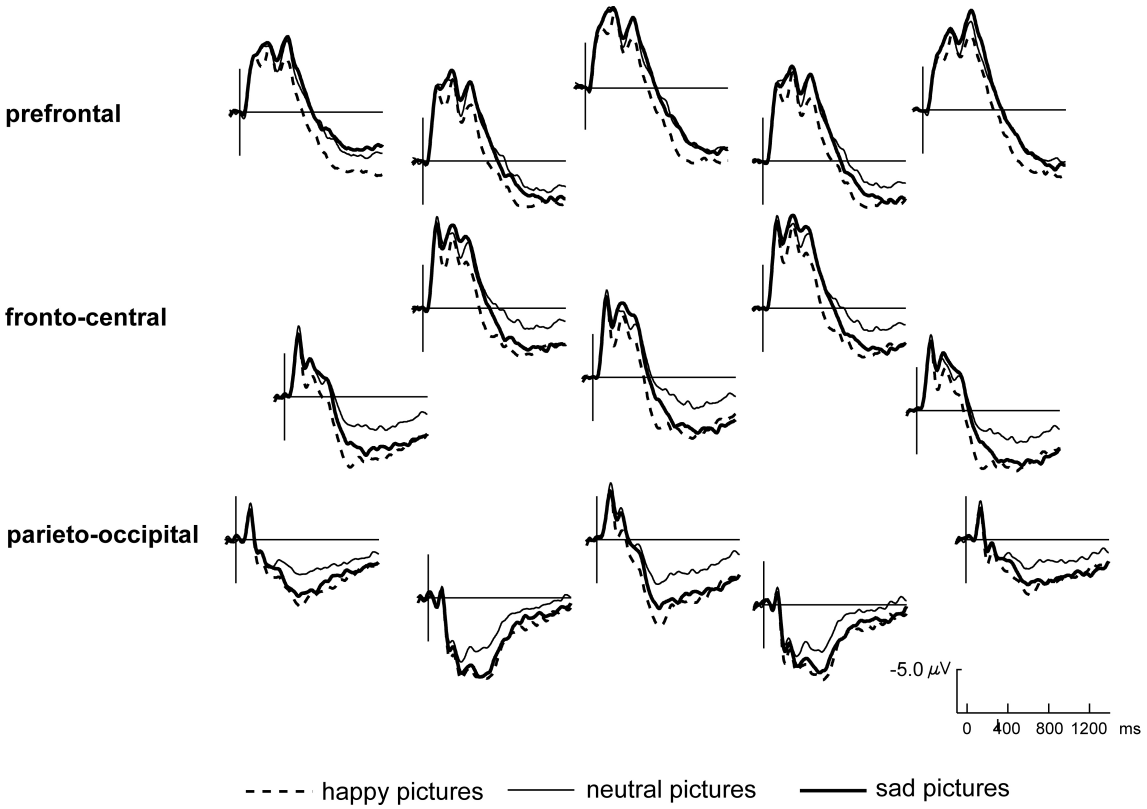

Figure 7.1.: Effects of (attended) picture valence in the attend-picture-task: depicted are grand average ERPs to the three different categories of picture valence (happy, neutral, sad) at prefrontal (top two rows), fronto-central (middle two rows) and parieto-occipital electrodes (bottom two rows).

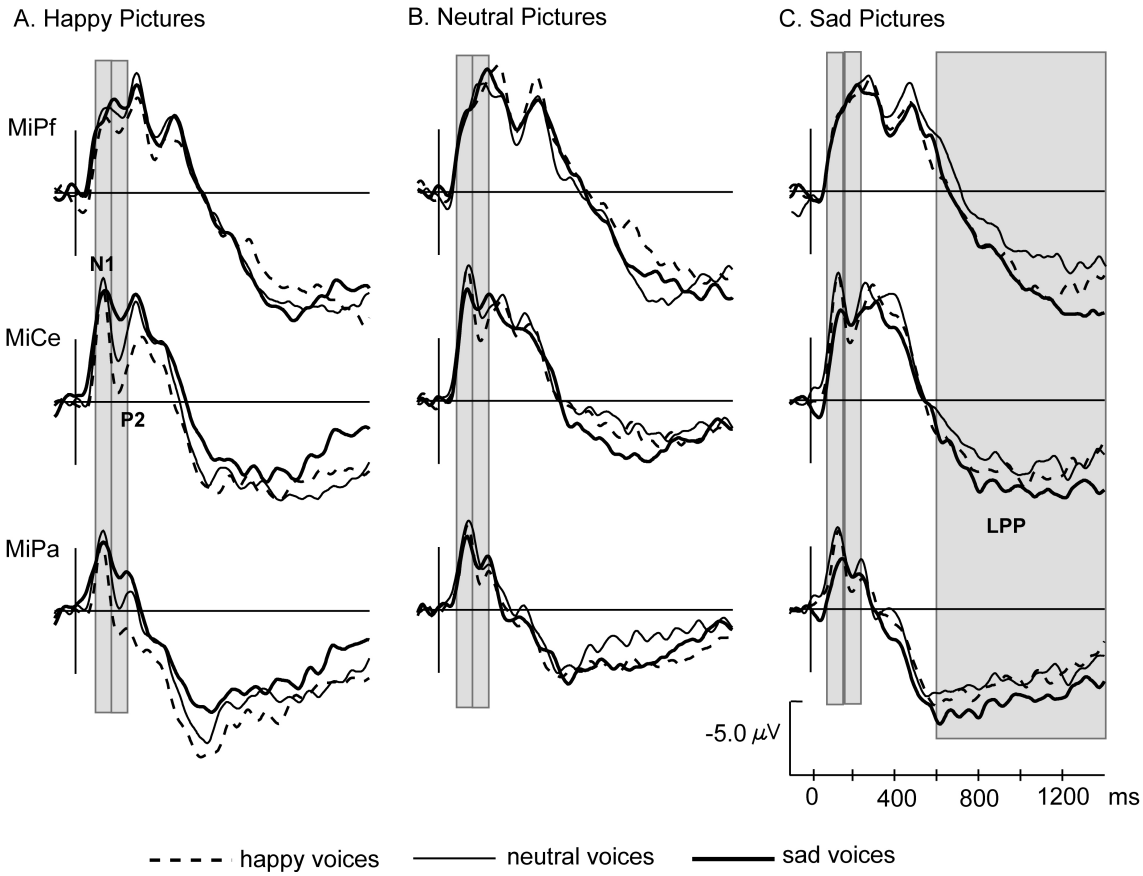

Figure 7.2.: Effects of (unattended) voice valence in the attend-picture-task: Grand average ERPs to the three different categories of voice valence (happy, neutral, sad), separately depicted for happy (A), neutral (B) and sad (C) pictures at three midline electrodes (MiPf=Midline Prefrontal, MiCe=Midline Central, MiPa=Midline Parietal). Time-windows with significant effects of affective $valence_{unatt}$ or $valence_{att}$ x $valence_{unatt}$ - interaction are highlighted.

7.3.2.1.2. Voice rating task

Effect of (unattended) picture valence When participants were asked to attend the voice instead of the picture, picture valence affected P3 ($F(2,26)=10.01$, $p<0.001$) and N2b amplitudes ($F(2,26)=2.16$, $p<0.05$) (see Fig. 7.3): P3 was greater for neutral pictures than for sad ($F(1,13)=28.79$, $p=0.000$) or happy ($F(1,13)=5.62$, $p=0.034$) pictures. The effect was largest over fronto-central electrodes (interaction with caudality ($F(4,52)= 5.32$, $p<0.001$; see table 7.2, bottom, for details). Sad pictures led to a

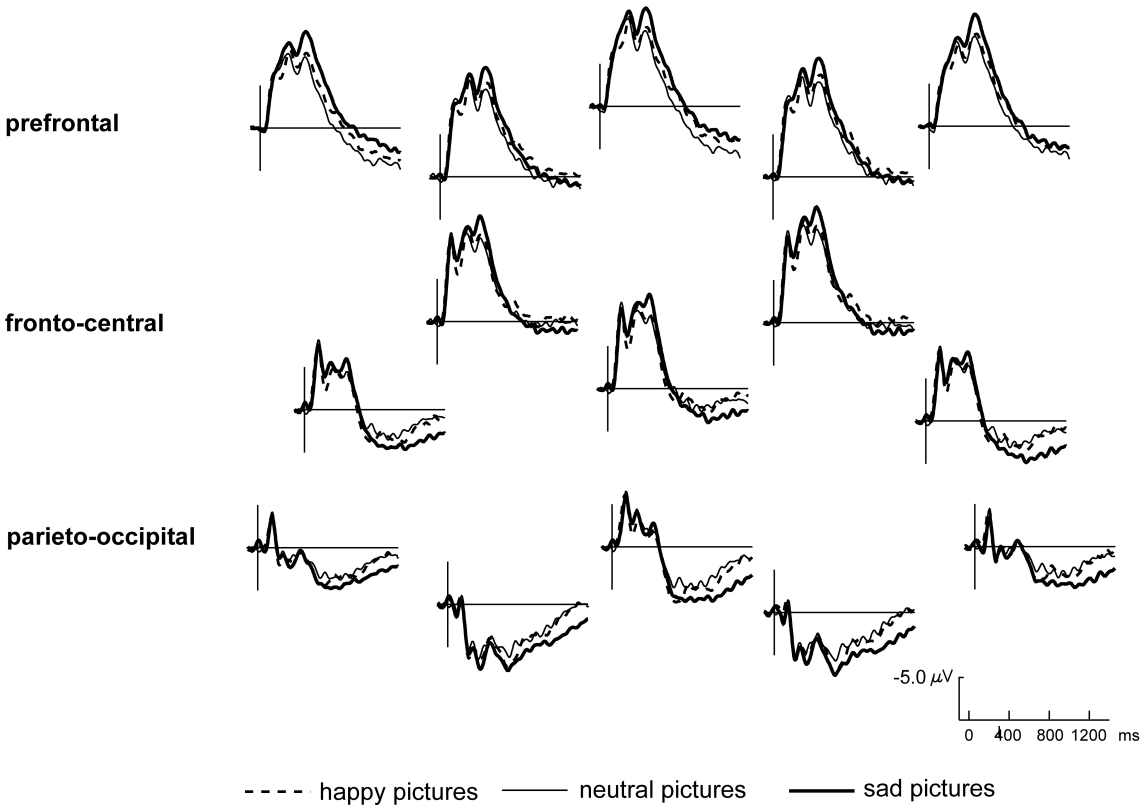

Figure 7.3.: Effects of (unattended) picture valence in the attend-voice-task: depicted are grand average ERPs to the three different categories of picture valence (happy, neutral, sad) at prefrontal (top two rows), fronto-central (middle two rows) and parieto-occipital electrodes (bottom two rows).

larger N2b than happy and neutral pictures. This effect also interacted with caudality ($F(4,52)=10.23$, $p<0.000$), reflecting a larger effect at prefrontal sites than at any other sites (see table 7.2 for details). The LPP effect seen in the attended-picture-condition was reduced and interacted with caudality ($F(4,52)=8.62$, $p<0.000$). Prefrontally, neutral pictures led to a greater positive deflection than sad pictures, while parieto-occipitally, sad pictures led to a greater positivity than happy and neutral pictures (see table 7.2 for details). No effect of picture valence was found for the P2 ($F(2,26)=2.31$, n.s.).

Effect of (attended) voice valence The N1 effect of voice valence reported for the attend-picture-task did not reach significance ($F(2,26)=2.53$, $p=0.099$) in the attend-voice-task (Fig. 7.4). However, valence of the voice stimulus, now attended, had a

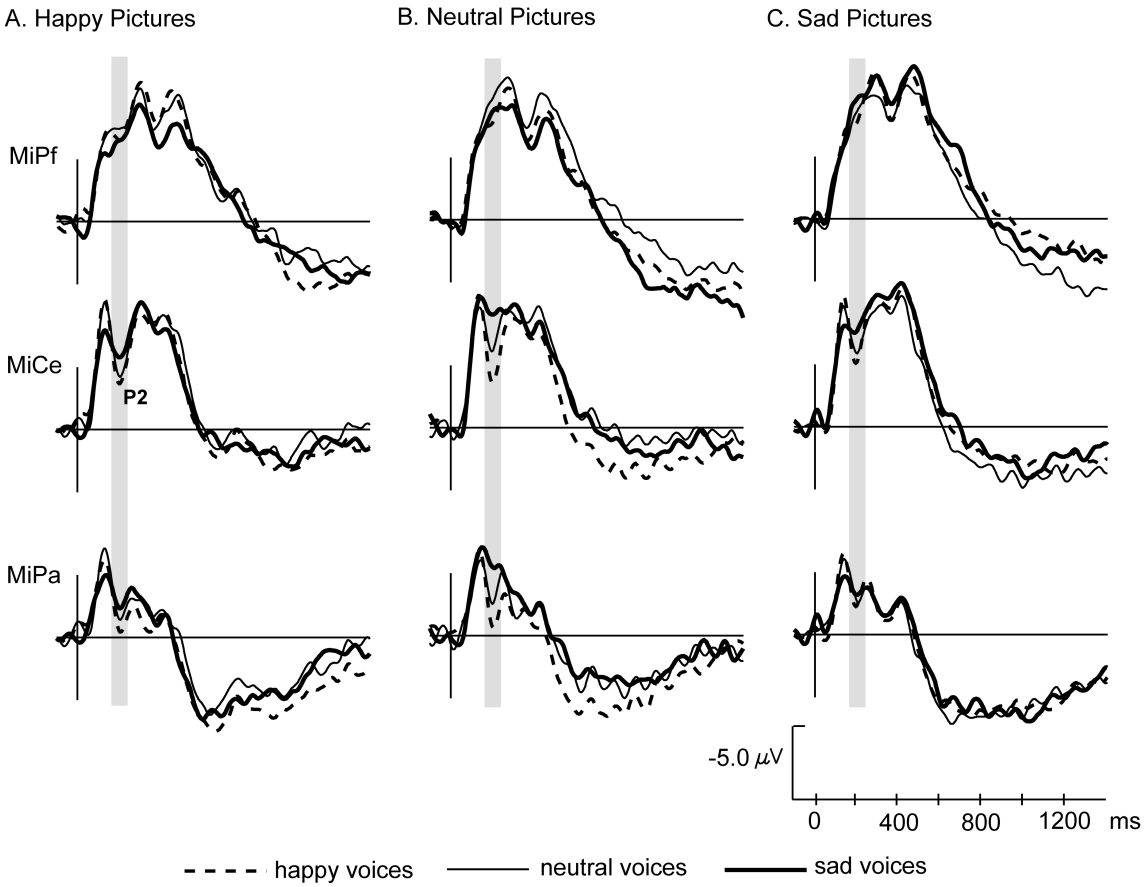

Figure 7.4.: Effects of (attended) voice valence in the attend-voice-task: Grand average ERPs to the three different categories of voice valence, separately depicted for happy (A), neutral (B) and sad (C) pictures at three midline electrodes (MiPf=Midline Prefrontal, MiCe=Midline Central, MiPa=Midline Parietal). Time-windows with significant effects of affective valence$_{unatt}$ or valence$_{att}$ x valence$_{unatt}$ - interaction are highlighted.

significant main effect on P2 amplitude (valence$_{att}$ $F(2,26)=6.19$, $p<0.01$). Again, the P2 was more pronounced when happy voice stimuli were presented than when neutral ($F(1,13)=7.29$, $p=0.018$) or sad ($F(1,13)=12.09$, $p=0.004$) voices were presented. No effect of voice valence was found for the LPP ($F(2,26)=1.84$, p=n.s.).

7.3.2.2. Task effect

ERPs were affected by the task manipulation. From 250 ms onwards ERPs took a relatively more positive course when the picture was being rated than when the voice was being rated (F-values for consecutive time-windows starting at 250 ms (1,13): 18.93,

76.19, 148.38, 20.83, all p<0.000). Between 250 and 500 ms, a main effect of caudality reflected greater positivity at parieto-occipital than at prefrontal and fronto-central leads in both tasks (F(2,26)= 48.55, 46.08, 63.81, all p<0.000) (see Fig 7.5). During the LPP, the caudality pattern interacted with task (interaction task x caudality F(2,26)=18.67, p<0.000), reflecting equipotential LPPs across the head in the voice rating task and a more frontally distributed positivity in the picture rating task.

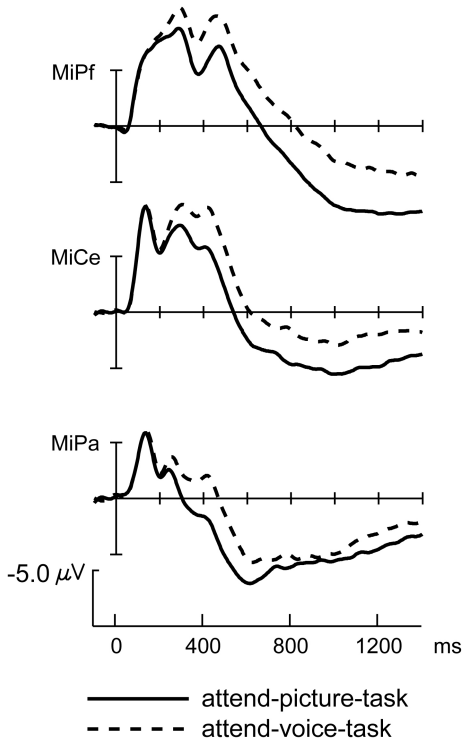

Figure 7.5.: Task effect: comparison of grand average ERPs of attend-picture-task (dotted line) and attend-voice-task (solid line) at three midline electrodes (MiPf=Midline Prefrontal, MiCe=Midline Central, MiPa=Midline Parietal) collapsed over all conditions.

7.4. Discussion

While it may not be surprising that people combine facial expressions with voice tones to gauge others' emotional states, it does not necessarily follow that people's affective ratings or processing of pictures would be influenced in any way by the affective content of a concurrent, but irrelevant sung note or vice versa. The current study, however, provides both behavioral and electrophysiological evidence for some interaction at the level of affect between simultaneously presented pictures and voices, even when only one of these modalities is actively attended (by instruction).

It was hypothesized that additional affective information in an unattended modality would have a certain potential to intensify or reduce affective impact of an emotional picture stimulus depending on whether its valence is congruent or incongruent with the picture valence. Although the rating of the pictures did not show a bias towards the valence of the concurrently presented voices, ERP responses indicate modified processing of picture-voice-pairs with matching affective valence. Sad pictures evoked a more positive-going LPP when the accompanying voice was also sad. Congruent pairing of happy pictures and happy voices led to enlargement of the P2-component.

P2-effect While it seemed likely to find modulation of ERP components known to reflect stimulus significance such as P3 and LPP, it was surprising to find such an early effect of affective coherence as the P2-effect for happy picture-voice-pairs. P2 is known to be an early sensory component that can be modulated by acoustical features of an auditory stimulus such as loudness or pitch (Picton, Goodman, & Bryce, 1970; Antinoro, Skinner, & Jones, 1969). In fact, the main effects of voice valence on the early components N1 and P2, found in both tasks, can be linked to differences in the acoustic structure of the voice stimuli. Musical notes expressing sadness tend to have a slower tone attack, also described as longer rise time, than happy notes (see Juslin03 for a review and discussion in previous chapters). However, increasing rise times are known to reduce the amplitude of auditory onset potentials (Kodera, Hink, Yamada, & Suzuki, 1979; Elfner, Gustafson, & Williams, 1976). This explanation cannot, however, account

for the striking asymmetry in P2 amplitude between congruent and incongruent happy picture-voice pairs. Obviously, the simultaneous presentation of the happy picture has led to enhanced processing of the happy voice, clearly indicating an early integration of the two modalities. Modulation of the P2-component has already been reported in audio-visual object recognition tasks. In designs comparing the ERP to simultaneous audio-visual presentation with the 'sum'-ERP of the uni-modally presented stimuli, P2 is larger in the 'simultaneous'-ERPs (Molholm et al., 2002; Giard & Peronnet, 1999). The functional significance of this effect, however, remains unclear. Pourtois et al. (2000) reported modulation of P2 in response to emotional congruent face-voice-pairs. However, the question arises: Why is there such an early effect for happy pictures but not for sad ones? It is possible that due to their specific physical structure (loud tone onset) happy voice stimuli are harder to ignore than sad or neutral voice stimuli and thus more likely to be integrated early in the visual perception process. Moreover, it is conceivable that happy pictures, too, are characterized by certain physical features such as a greater brightness and luminance than e.g. sad pictures. It is known that certain sensory dimensions correspond across modalities and that dimensional congruency enhances performance even when task irrelevant. For example, pitch and loudness in audition have been shown to parallel brightness in vision (Marks, Ben-Artzi, & Lakatos, 2003). Thus, loud and high pitched sounds that are paired with bright lights result in a better performance than incongruent pairing with dim lights. Findings that such cross-modal perceptual matches can already be made by small children has led to assume similarity of neural codes for pitch, loudness and visual brightness (Mondloch & Maurer, 2004; Marks, 2004). However, the notion that P2 reflects such loudness-brightness correspondence would need to be studied in future experiments. The picture-voice-valence interaction vanished when the attention was shifted from pictures to voices in the attend-voice-task indicating that whatever caused the effect of picture valence on the auditory component, was not an automatic process but required attending the picture.

LPP-effect In line with the hypothesis the LPP in the attend-picture-task was enhanced for sad pictures that were paired with sad voice stimuli. Based on the assumption that LPP-amplitude increases with stimulus significance and reflects enhanced processing it can be inferred that the additional congruent affective information has intensified the perceived sadness or at least made it less ambiguous. Happy pictures, too, gained enhanced processing when paired with happy voices, though only over visual areas at the back of the head. However, the latter effect did not become significant. Perhaps if the valence in the voices would have been more salient it would have been more easily extracted automatically and had a greater influence on the ERPs to pictures. Nevertheless, our data implies that even affective information that is less naturalistically associated than faces and voices is integrated across channels. Thus, our results underline the role of emotional coherence as a binding factor.

Effect of task The change of attentional focus from pictures to voices in the attend-voice-task had a considerable effect on the ERP with amplitude and topographical differences starting at around 250 ms. Both tasks elicited a late positivity starting at $\sim$400 ms with a maximum at about 600 ms at parietal sites. Only at prefrontal and fronto-central electrodes the positivity continued to the end of the time window (1400 ms). A frontal effect with a similar time course has previously been described in response to emotional stimuli when the task specifically calls for attention to the emotional content (Johnston et al., 1986; Johnston & Wang, 1991; Naumann, Bartussek, Diedrich, & Laufer, 1992) and has been taken to reflect engagement of the frontal cortex in emotional processing (Bechara, Damasio, & Damasio, 2000). However, shifting the attention away from the pictures in the voice rating task resulted in an overall more negative going ERP. Particularly at prefrontal and frontal electrodes P3 and LPP were largely reduced in the voice rating task compared to the picture rating task. Naumann et al. (1992) reported a similar pattern after presenting affective words and asking two groups of participants to either rate the affective valence (emotion group) or to count the letters of the words (structure group). The resulting pronounced frontal late positive potential only present

in the emotion group was interpreted as reflecting emotion specific processes. It thus seems that rating the voice valence was a suitable task to shift participants' attention away from the emotional content of the pictures. It also indicates that the frontal cortex is less involved in the evaluation of the affective voice stimuli than in evaluation of the picture. In the next paragraph the effects of picture and voice valence when attention was drawn off the pictures will be discussed.

The rating of the voices was considerably biased by the valence of the pictures. It seemed to have been much more difficult to fight off the impression of the picture than ignoring the voice. The bias of affective ratings of faces and voices has been reported to be stronger if the expression of the to be rated item was neutral (Massaro & Egan, 1996). The behavioral data of the present study confirm this notion though a bias of the unattended picture valence was also found if the voice was happy. Interestingly, the ERP recording revealed larger P3 amplitudes for neutral than for happy or sad pictures. It is possible that this pattern reflects a shift of attentional resources. As has been suggested by others (Schupp et al., 1997; Morita et al., 2001) more attentional resources were available for the auditory stimulus (resulting in an enhanced P3) when the concurrently presented picture was not affective and/or arousing than when it was. As an additional effect of picture valence, sad pictures elicited a larger N2b than happy and neutral pictures over the front of the head. Enhanced N2b-components over fronto-central electrode sites are typically observed when response preparation needs to be interrupted as in response to NoGo-items in Go/NoGo-tasks (Pfefferbaum & Ford, 1988; Jodo & Kayama, 1992; Eimer, 1993). Based on the finding that negative items are more likely than positive items to bias a multi-sensory percept (Ito, Larsen, Smith, & Cacioppo, 1998; Ito & Cacioppo, 2000; Windmann & Kutas, 2001), it can be speculated that sad pictures are more difficult to ignore and thus lead to a greater NoGo-response.

The greater LPP-amplitude for affective versus non-affective pictures that is characteristic for affective picture processing (Palomba et al., 1997; Ito et al., 1998; Schupp et al., 2000; Cuthbert et al., 2000) and which had been observed in the attend-picture-task, appeared to be largely reduced if attention was directed away from the visual toward

the auditory modality. Diedrich et al. (1997), likewise, did not find a difference between affective and neutral pictures when participants' were distracted from attending to the emotional content of the pictures by a structural processing task. In the present study, however, the effect of valence on the LPP while reduced was not completely eliminated. Prefrontally, neutral pictures were associated with a greater positive deflection than sad pictures, while parieto-occipitally, sad pictures were associated with a greater positivity than happy and neutral pictures. Against the theoretical background that LPP-amplitudes to affective stimuli reflect their intrinsic motivational relevance (Lang et al., 1997; Cuthbert et al., 2000), both, the parietal as well as the prefrontal effect seem to be related to the perceived valence of the multi-sensory presentation. However, perceived valence was not always dominated by the valence of the to-be-attended voice modality. The prefrontal effect bears some similarity to the P3 effect of picture valence discussed earlier. The valence of the voices could only be adequately processed if the evaluation was not disturbed by arousing content of affective pictures. While the dominant (sad) picture valence influences neural responses mainly over primary visual areas at the back of the head, detection of happy and sad voice tones is accompanied by enhanced positivities over prefrontal sites which, if taken at face value, reflect activity of brain areas known to be involved in the processing of emotional vocalizations as well as emotion in music (see review in section 2.3.5). The different topographies, thus, implicate at least two separate processes, each related to modality-specific processing of affect.

To conclude, the present study delineated the time-course of integration of affective information from different sensory channels extracted from stimuli that are only abstractly related. The data indicate that integration of affective picture-voice-pairs can occur as early as 150 ms if the valence information is salient enough. Congruent auditory information evokes enhanced picture processing. It was thus demonstrated that audio-visual integration of affect is not reduced to face-voice pairs but also occurs between voices and pictures of complex scenes. Probably because the human voice is a particularly strong emotional stimulus affective information is automatically extracted from it even if it is

not task relevant. The data further highlights the role of attention in the multisensory integration of affective information (Gelder, Vroomen, & Pourtois, 2004) indicating that integration of picture and voice valence require that pictures are attended.

8. Conclusions

8.1. Summary of key findings

- The first experiment demonstrated that even subtle changes in tones of different emotional expression can be detected, both actively and passively. Reaction time and P3b latency were shorter if the target was happy than if it was sad, possibly as a consequence of more prominent acoustical features of the happy tone. In the passive condition emotional deviants triggered a classical MMN.

- Results of MMN-Exp. II demonstrated that auditory stimuli are pre-attentively grouped into one emotional category irrespective of differences in acoustical structure, as long as enough stimulus features meet with certain criteria of a prototypical emotional representation in the brain.

- The experiments presented in part II addressed processing of emotional information conveyed by the voice. Experiment II-01 provided evidence that emotional information can be extracted from the voice within the first 200 ms of listening. Emotion recognition precedes identity recognition though only if the acoustical structure allows for rapid decoding of the emotionally significant cues. The result supports recent models of voice perception suggesting separate processing pathways for emotion and identity recognition (Belin et al., 2004).

- It was shown in experiment II-02 that emotional information from voices and pictures are integrated early in the processing stream. Again, emotion-specific timing differences were found. While congruent happy voice-picture pairs increased the amplitude of the ERP-component P2, congruent sad voice-picture pairs modulated the late positive potential (LPP) from 500 ms onwards.

8.2. General discussion

It is commonly agreed that emotion expression in the auditory domain is coded by acoustical features. As a consequence it is difficult to disentangle effects which are triggered by the functional significance of the stimulus and those which merely reflect structural processing. In addition, code usage in emotional communication is known to be redundant and ambiguous. Both issues were addressed in part I of this thesis. The results support Juslin's adaptation of the Brunswikian lens model (Juslin, 1997b) which suggests that the sender of an emotional expression (speaker or musician) can use several expressive cues in different combinations which can nonetheless be decoded by the listener.

Integrating the findings of the experiments into previous models on auditory perception provides a rough outline of how the fast recognition of auditory emotional information might be performed by the brain:

After the basic sensory features of the acoustical stimulus have been analyzed in the auditory cortex (stage 1 of the 3-stage-model by Schirmer & Kotz, 2006), integration of the separate acoustic features takes place to form an 'auditory object' (stage 2). Näätänen & Winkler (1999) have introduced the term 'central sound representation' (CSR) to describe the earliest form of an integrated feature pattern. The CSR is supposed to be built immediately after the acoustical features of a sound have been separately processed in the brain and mapped into a common sensory memory. It is expected to be completed within 200 ms after sound onset (Näätänen, 2001) and provides the basis of what is perceived as sound by the listener. The mechanisms underlying the 'binding' of separate features into one auditory object (i.e. the CSR) are, however, still unclear. Nevertheless, it can be assumed that the CSR is matched against a number of auditory object representations already existent in the brain in consecutive processing steps. The MMN evoked by the emotional deviant in experiment MMN-II of this thesis is thought to reflect such a matching process. It implies that emotional categorization is based on prior feature integration. Moreover, it provides evidence that the matching process of emotionally significant input can take place pre-consciously.

The conscious processing of emotional auditory input was addressed in part II. The re-

sults of experiment II-01 support considerations that the recognition of vocally expressed emotions is provided by different processes than speaker recognition (Belin et al., 2004). In the experiment, identical stimuli evoked different brain responses as a consequence of different tasks. In line with the considerations outlined above, it may be assumed that the prime in the emotion matching task triggered the activation of different neural representations than if the same sound served as prime in the identity matching task. However, it can only be speculated why the emotional prime facilitated target processing more than did the identity prime, as reflected in the earlier effect of emotional priming. One possible explanation is that the matching process underlying emotion recognition is more tolerant to feature variance than the identity matching process. Based on the findings of the MMN-Exp. II it can be assumed, that the emotional prime activated a large number of representations of different feature combinations, all encoding the same emotion. In contrast, no such sets of prototypical feature patterns are likely to exist for vocal identity (at least not for unfamiliar voices). It is likely that the pre-activation of representations in the same emotional category facilitated the processing of the emotionally congruent target sound.

Another point touched upon by the present results concerns the differences in behavioral as well as brain activity measures found for different emotions. Though different stimulus material was used in part I and II, unequal result patterns for happy and sad tones were found in both. Accumulated results from violin tones (active condition of experiment MMN-I) and sung notes (experiment II-01 and II-02) imply that happiness can be registered faster than sadness.

The advantage of happy tones can most likely be linked to a higher salience resulting from typical acoustic features such as loudness and attack. The question then arises what functional purpose the greater salience of happy stimuli might serve. It does not seem plausible that happiness requires fast adaptational behavior in the sense of fight or flight, because it does not signal potential danger. It seems, however, possible that the early reaction to happy stimuli can be linked to the potential of certain acoustical fea-

tures to code more than one emotion. Especially the fast tone attack has been identified as a cue for panic and anger. Both are certainly likely to trigger adaptive behavior in the listener. Thus, it is possible that the early effect for happy tones seen in the present data, reflects a general orientation reaction to stimuli signaling potential danger. The fact that the multidimensional scaling in experiment MMN-II (see section 5.4) revealed that the happy tones were perceived ambiguously, supports this notion.

8.3. Implications for future research

The present data can only contribute to the understanding of basic categorization of simple emotional stimuli. However, it is believed that the complexity of the phenomenon of vocal affect communication requires that more studies focus on very circumscribed aspects of the evaluation process. The results point to an important role of voice quality and timbre in emotion coding. As a consequence, future studies on emotional prosody and musical performance need to take into account the effect of this and other segmental features, which considerably add to the effect of large-scale manipulations of speech melody or tempo.

Differing results for happy and sad stimuli could be linked to insufficient functional validity of the cues used to encode happiness. This finding, however, has important implications for future studies. It underlines the importance of being very thorough when selecting affective stimulus material. Mere testing on rating scales might not be sufficient to evaluate valence. It was demonstrated in the present study that multidimensional scaling provides an interesting means to assess the perceived emotional similarity of stimulus sets. From the results it can further be derived that it might be necessary to address expression and processing of different emotions separately for different emotions. So far most studies treated expressions of different emotions as if they were the same. However, because of the strong dependence of auditory expressed emotions on acoustical features it might be that processing of different expressed emotions (or the dimensions underlying them) relies on different brain patterns because different acoustic features dominate their expression. For example, it is possible that the processing of sad expressions relies to

a larger extent on the analysis of spectral aspects related to timbre than happiness. Consequently, other brain structures are likely to be involved than during the analysis of the tempo-variations found to be an important cue in recognition of happiness. The considerations about emotion-specific physical differences constitute the main difference between affect communication in audition and vision and need to be kept in mind when trying to draw parallels.

8.4. Concluding remark

The aim of this thesis was to contribute to the understanding of auditory emotion perception. By presenting brief and relatively simple auditory stimuli, it was possible to demonstrate that the brain is in possession of refined tools to interpret even the smallest nuances of tonal variation. These functions are the core of human verbal communication and are indispensable for appreciation of music. The reason for this can best be expressed by a quote from Korean composer Isang Yun (1917-1995):

> *"A tone is made of many small movements, which build up to a whole cosmos. The tone itself is life."*

References

Adolphs, R., Tranel, D., & Damasio, A. R. (2003). Dissociable neural systems for recognizing emotions. *Brain Cogn*, *52*(1), 61-9.

Adolphs, R., Tranel, D., & Damasio, H. (2001). Emotion recognition from faces and prosody following temporal lobectomy. *Neuropsychology*, *15*(3), 396-404.

Altenmüller, E. (2003). How many music centers are in the brain? In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (p. 346-353). Oxford: Oxford University Press.

Altenmüller, E., Münte, T. F., & Gerloff, C. (2005). Neurocognitive functions and the EEG. In E. Niedermeyer & F. Lopes da Silva (Eds.), *Electroencepahlography: basic principles, clinical applications, and related fields* (p. 661-682). Philadelphia: Lippincott Williams & Wilkins.

Altenmüller, E., Schürmann, K., Lim, V. K., & Parlitz, D. (2002). Hits to the left, flops to the right: different emotions during listening to music are reflected in cortical lateralisation patterns. *Neuropsychologia*, *40*(13), 2242-56.

Alter, K., Rank, E., Kotz, S. A., Toepel, U., Besson, M., Schirmer, A., et al. (2003). Affective encoding in the speech signal and in event-related brain potentials. *Speech communication*, *40*, 61-70.

American-Standards-Association. (1960). *Acoustical terminology SI, 1-1960*. Washington D.C.: American Standards Association/American National Standards Institute.

Anagnostaras, S., Maren, S., & Fanselow, M. (1999). Temporally graded retrograde amnesia of contextual fear after hippocampal damage in rats: Within-subjects examination. *The Journal of Neuroscience*, *19*(3), 1106-1114.

Anderson, A. K., & Phelps, E. A. (1998). Intact recognition of vocal expressions of fear following bilateral lesions of the human amygdala. *Neuroreport*, *9*(16), 3607-13.

Antinoro, F., Skinner, P. H., & Jones, J. J. (1969). Relation between sound intensity and amplitude of the AER at different stimulus frequencies. *J Acoust Soc Am*, *46*(6), 1433-6.

Balconi, M., & Carrera, A. (2005). Cross-modal perception of emotion by face and voice: an ERP study. In *Proceedings of the 6th international multisensory research forum.* Trento, Italy.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614-636.

Barrett, A. M., Crucian, G. P., Raymer, A. M., & Heilman, K. M. (1999). Spared comprehension of emotional prosody in a patient with global aphasia. *Neuropsychiatry Neuropsychol Behav Neurol*, *12*(2), 117-20.

Barrett, J. F., Pike, G. B., & Paus, T. (2004). The role of the anterior cingulate cortex in pitch variation during sad affect. *Eur J Neurosci*, *19*(2), 458-64.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cereb Cortex*, *10*(3), 295-307.

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci*, *8*(3), 129-35.

Belin, P., & Zatorre, R. J. (2003). Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport*, *14*(16), 2105-9.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*(6767), 309-12.

Bezooijen, R. van. (1984). *Characteristics and recognizability of vocal expressions of emotion*. Dordrecht: ICG Printing.

Blonder, L. X., Bowers, D., & Heilman, K. M. (1991). The role of the right hemisphere in emotional communication. *Brain*, *114 (Pt 3)*, 1115-27.

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proc Natl Acad Sci U S A*, *98*(20), 11818-23.

Blood, A. J., Zatorre, R. J., Bermudez, P., & Evans, A. C. (1999). Emotional responses to pleasant and unpleasant music correlate with activity in paralimbic brain regions. *Nat Neurosci*, *2*(4), 382-7.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci*, *8*(3), 389-95.

Boettcher-Gandor, C., & Ullsperger, P. (1992). Mismatch negativity in event-related potentials to auditory stimuli as a function of varying interstimulus interval. *Psychophysiology*, *29*(5), 546-50.

Bolivar, V., Cohen, A., & Fentress, J. (1994). Semantic and formal congruency in music and motion pictures: effects on the interpretation of visual action. *Psychomusicology*, *13*, 28-59.

Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology*, *41*(2), 259-68.

Bradley, M. M., Greenwald, M. K., & Hamm, A. O. (1993). Affective picture processing. In N. Birbaumer & A. Öhman (Eds.), *The structure of emotion* (p. 48-65). Seattle: Hogrefe & Huber.

Breitenstein, C., Lancker, D. van, Daum, I., & Waters, C. H. (2001). Impaired perception of vocal emotions in parkinson's disease: influence of speech time processing and executive functioning. *Brain Cogn*, *45*(2), 277-314.

Brown, S. (2000). The "musilanguage" model of music evolution. In N. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (p. 271-300). Cambridge: The MIT Press.

Brown, S., Merker, B., & Wallin, N. (2000). An introduction to evolutionary musicology. In N. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (p. 3-24). Cambridge: The MIT Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(193-217).

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., et al. (2000). Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Brain Res Cogn Brain Res*, *9*(3), 227-38.

Buck, R. (1984). *The communication of emotion*. New York: Guilford Press.

Cacioppo, J. T., Berntson, G., Larsen, J., Poehlmann, K., & Ito, T. (2000). The psychophysiology of emotion. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of emotions* (p. 173-191). New York: The Guilford Press.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nat Rev Neurosci*, *6*(8), 641-51.

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex*, *11*(12), 1110-23.

Cannon, W. (1927). The James-Lange theory of emotions: A critical examination and an alternative theory. *American Journal of Psychology*, *39*, 106-124.

Celesia, G. G., & Brigell, M. G. (2005). Auditory evoked potentials. In E. Niedermeyer & F. Lopes da Silva (Eds.), *Electroencepahlography: basic principles, clinical applications, and related fields* (p. 1045-1066). Philadelphia: Lippincott Williams & Wilkins.

Coles, M. G. H., & Rugg, M. D. (1996). Event-related brain potentials: an introduction. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind - event-related brain potentials and cognition* (25 ed., p. 1-26). Oxford: Oxford University Press.

Crummer, G., Walton, J., Wayman, J., Hantz, E., & Frisina, R. (1994). Neural processing of musical timbre by musicians, nonmusicians, and musicians possessing absolute pitch. *Journal of the acoustical society of america*, *95*(5(1)), 2720-2727.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biol Psychol*, *52*(2), 95-111.

Damasio, A. R. (1999). *The feeling of what happens*. San Diego: Harcourt Inc.

Darwin, C. (1998/1872). *The expression of the emotions in man and animals*. London: Harper-Collins.

Davidson, R. J. (1992). Anterior cerebral asymmetry and the nature of emotion. *Brain Cogn*, *20*(1), 125-51.

Davidson, R. J. (2003). Darwin and the neural basis of emotion and affective style. *Annals of the New York Acadademy of Sciences*, *1000*, 316-336.

Davis, K. D., Taylor, K. S., Hutchison, W. D., Dostrovsky, J. O., McAndrews, M. P., Richter, E. O., et al. (2005). Human anterior cingulate cortex neurons encode cognitive and emotional demands. *J Neurosci*, *25*(37), 8402-6.

Davitz, J. (1964). Personality, perceptual, and cognitive correlates of emotional sensitivity. In J. Davitz (Ed.), *The communication of emotional meaning* (p. 57-68). New York: McGraw-Hill.

Demaree, H. A., Everhart, D. E., Youngstrom, E. A., & Harrison, D. W. (2005). Brain lateralization of emotional processing: historical roots and a future incorporating "dominance". *Behav Cogn Neurosci Rev*, *4*(1), 3-20.

Deouell, L. Y., & Bentin, S. (1998). Variable cerebral responses to equally distinct deviance in four auditory dimensions: a mismatch negativity study. *Psychophysiology*, *35*(6), 745-54.

Diedrich, O., Naumann, E., Maier, S., Becker, G., & Bartussek, D. (1997). A frontal positive slow wave in the ERP associated with emotional slides. *Journal of Psychophysiology*, *11*, 71-84.

Dien, J. (1998). Issues in the application of the average reference: Review, critiques and recommendations. *Behavior Research Methods, Instruments and Computers*, *30*, 34-43.

Dissanayake, E. (2000). Antecedents of the temporal arts in early mother-infant interaction. In N. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (p. 389-410). Cambridge: The MIT Press.

Donchin, E. (1981). Surprise!...surprise? *Psychophysiology*, *18*(5), 493-513.

Donchin, E., & Coles, M. G. H. (1988a). Is the P300 component a manifestation of context updating? *The Behavioral and Brain Sciences*, *11*, 355-372.

Donchin, E., & Coles, M. G. H. (1988b). On the conceptual foundations of cognitive psychophysiology. *The Behavioral and Brain Sciences*, *11*, 406-417.

Donchin, E., Heffley, E., Hillyard, S. A., Loveless, N., Maltzman, I., Öhman, A., et al. (1984). Cognition and event-related potentials. ii. the orienting reflex and p300. *Ann N Y Acad Sci*, *425*, 39-57.

Dunbar, R. (1996). *Grooming, gossip and the evolution of language*. London: Faber and Faber.

Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: the variation of event-related potentials with subjective probability. *Psychophysiology*, *14*(5), 456-67.

Duncan-Johnson, C. C., & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. *Biol Psychol*, *14*(1-2), 1-52.

Dzhafarov, E., & Colonius, H. (2006). Generalized fechnerian scaling. In H. Colonius & E. Dzhafarov (Eds.), *Measurement and representation of sensations* (p. 47-87). Mahwah, NJ: Erlbaum.

Dzhafarov, E. N., & Colonius, H. (1999). Fechnerian metrics in unidimensional and multidimensional stimulus spaces. *Psychon Bull Rev*, *6*(2), 239-68.

Dzhafarov, E. N., & Colonius, H. (2001). Multidimensional fechnerian scaling: Basics. *J Math Psychol*, *45*(5), 670-719.

Eibl-Eibesfeldt, I. (1973). The expressive behaviour of the deaf-and-blind born. In M. Von Cranach & I. Vine (Eds.), *Social communication and movement* (p. 163-194). London: Academic Press.

Eimer, M. (1993). Effects of attention and stimulus probability on ERPs in a go/nogo task. *Biol Psychol*, *35*(2), 123-38.

Ekman, P. (1957). A methodological discussion of nonverbal behavior. *Journal of Psychology*, *43*, 141-149.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*, 169-200.

Ekman, P. (1999a). Basic emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (p. 45-60). Sussex: John Wiley and Sons Ltd.

Ekman, P. (1999b). Facial expressions. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (p. 301-320). Sussex: John Wiley and Sons, Ltd.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*, *221*(4616), 1208-10.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. *Psychol Bull*, *128*(2), 203-35.

Elfner, L. F., Gustafson, D. J., & Williams, K. N. (1976). Signal onset and task variables in auditory evoked potentials. *Biol Psychol*, *4*(3), 197-206.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, *44*(3), 227-40.

Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., et al. (2006). Effects of prosodic emotional intensity on activation of associative auditory cortex. *Neuroreport*, *17*(3), 249-53.

Fechner, G. (1860). *Elemente der Psychophysik [elements of psychophysics]*. Leipzig: Breitkopf und Härtel.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Priming of non-speech vocalizations in male adults: the influence of the speaker's gender. *Brain Cogn*, *55*(2), 300-2.

Feldman Barrett, L., & Russell, J. (1999). The structure of current affect: controversies and emerging consensus. *Current directions in psychological science*, *8*(1), 10-14.

Fonagy, I. (1962). Mimik auf glottaler Ebene. *Phonetica*, *8*, 209-219.

Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: a dynamic dual pathway model. *Brain and Language*, *89*, 267-276.

Friend, M., & Farrar, M. J. (1994). A comparison of content-masking procedures for obtaining judgments of discrete affective states. *J Acoust Soc Am*, *96*(3), 1283-90.

Gabrielsson, A., & Juslin, P. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*, 68-91.

Gabrielsson, A., & Juslin, P. (2003). Emotional expression in music. In R. J. Davidson, H. H. Goldsmith, & K. R. Scherer (Eds.), *Handbook of affective sciences* (p. 503-534). New York: Oxford University Press.

Gärtner, J. (1974). *Das Vibrato unter besonderer Berücksichtigung der Verhältnisse bei Flötisten : historische Entwicklung, neue physiologische Erkenntniss sowie Vorstellungen über ein integrierendes Lernverfahren*. Regensburg: Bosse.

Geissmann, T. (2000). Gibbon songs and human music from an evolutionary perspective. In N. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (p. 103-124). Cambridge: The MIT Press.

Gelder, B. de, Bocker, K. B., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. *Neurosci Lett*, *260*(2), 133-6.

Gelder, B. de, Pourtois, G., & Weiskrantz, L. (2002). Fear recognition in the voice is

modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proc Natl Acad Sci U S A*, *99*(6), 4121-6.

Gelder, B. de, & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and emotion*, *14*(3), 289-311.

Gelder, B. de, Vroomen, J., & Pourtois, G. (2004). Multisensory perception of affect, its time course and its neural basis. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processes* (p. 581-596). Cambridge, MA: MIT.

Gelfer, M. P., & Mikos, V. A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *J Voice*, *19*(4), 544-54.

George, M. S., Parekh, P. I., Rosinsky, N., Ketter, T. A., Kimbrell, T. A., Heilman, K. M., et al. (1996). Understanding emotional prosody activates right hemisphere regions. *Arch Neurol*, *53*(7), 665-70.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci*, *11*(5), 473-90.

Glaser, E., & Ruchkin, D. (1996). *Principles of neurobiological signal analysis*. New York: Academic Press.

Gomes, H., Bernstein, R., Ritter, W., Vaughan, J., H. G., & Miller, J. (1997). Storage of feature conjunctions in transient auditory memory. *Psychophysiology*, *34*(6), 712-6.

Gosselin, N., Peretz, I., Noulhiane, M., Hasboun, D., Beckett, C., Baulac, M., et al. (2005). Impaired recognition of scary music following unilateral temporal lobe excision. *Brain*, *128*(Pt 3), 628-40.

Gouzoules, H., & Gouzoules, S. (1989). Design features and developmental modification of pigtail macaque, macaca nemestrina, agonistic screams. *Animal Behaviour*, *37*, 383-401.

Gratton, G. (1998). Dealing with artifacts: The EOG contamination of the event-related brain potential. *Behavior Research Methods, Instruments and Computers*, *30*(1), 44-53.

Grey, J. (1977). Multidimensional perceptual scaling. *Journal of the acoustical society of america*, *61*, 1270-1277.

Halberstadt, J. (2005). Featural shift in explanation-biased memory for emotional faces. *J Pers Soc Psychol*, *88*(1), 38-49.

Hammerschmidt, K., Freudenstein, T., & Jürgens, U. (2001). Vocal development in squirrel monkeys. *Behaviour*, *138*, 1179-1204.

Hauser, M. (1997). *The evolution of communication*. Cambridge: MIT Press.

Hefter, R. L., Manoach, D. S., & Barton, J. J. (2005). Perception of facial expression and facial identity in subjects with social developmental disorders. *Neurology*, *65*(10), 1620-5.

Helmholtz, H. (1885/1954). *On the sensations of tone*. New York: Dover Publications.

Hevner, K. (1935). The affective character of the major and minor modes in music. *American Journal of Psychology*, *47*, 103-118.

Hevner, K. (1936). Experimental studies of the elements of expression in music. *American Journal of Psychology*, *48*, 246-268.

Hevner, K. (1937). The affective value of pitch and tempo in music. *American Journal of Psychology*, 621-630.

Holmes, A., Kiss, M., & Eimer, M. (2006). Attention modulates the processing of emotional expression triggered by foveal faces. *Neurosci Lett*, *394*(1), 48-52.

Huynh, H., & Feldt, L. (1980). Conditions under which mean square ratios in repeated measure designs have excact f-distributions. *Journal of The American Statistical Association*, *65*, 1582-1589.

Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., et al. (1997). Vocal identification of speaker and emotion activates different brain regions. *Neuroreport*, *8*(12), 2809-12.

Ito, T., & Cacioppo, J. T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *Journal of Experimental Social Psychology*, *36*, 660-676.

Ito, T., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *J Pers Soc Psychol*, *75*(4), 887-900.

Iverson, P., & Krumhansl, C. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the acoustical society of america*, *94*, 2595-2603.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychol Rev*, *99*(3), 561-5.

Jacobsen, T., Schroeger, E., & Sussman, E. (2004). Pre-attentive categorization of vowel formant structure in complex tones. *Brain Res Cogn Brain Res*, *20*(3), 473-9.

James, W. (1884). What is an emotion? *Mind*, *9*, 188-205.

Jaramillo, M., Ilvonen, T., Kujala, T., Alku, P., Tervaniemi, M., & Alho, K. (2001). Are different kinds of acoustic features processed differently for speech and non-speech sounds? *Brain Res Cogn Brain Res*, *12*(3), 459-66.

Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a go/no-go task. *Electroencephalogr Clin Neurophysiol*, *82*(6), 477-82.

Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology*, *23*(4), 367-84.

Johnston, V. S., Miller, D., & Burleson, M. (1986). Multiple P3s to emotional stimuli and their theoretical significance. *Psychophysiology*, *23*(6), 684-694.

Johnston, V. S., & Wang, X. T. (1991). The relationship between menstrual phase and the P3 component of ERPs. *Psychophysiology*, *28*(4), 400-9.

Johnstone, T., Reekum, C. van, & Scherer, K. R. (2001). Vocal expression correlates of appraisal processes. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion* (p. 271-284). Oxford: University Press.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. Haviland-Jones (Eds.), *Handbook of emotions* (p. 220-235). New York: Guilford Press.

Jürgens, U. (1979). Vocalization as an emotional indicator. a neuroethological study in the squirrel monkey. *Behaviour*, *69*(1-2), 88-117.

Jürgens, U. (2003). Zum stimmlichen Ausdruck emotionaler Zustände. eine vergleichend verhaltens- und neurobiologische Untersuchung [on the vocal expression of emotional states. a comparative ethological and neurobiological study]. *Sprache, Stimme, Gehör*, *27*, 71-74.

Jürgens, U., & Cramon, D. Y. von. (1982a). On the role of the anterior cingulate cortex in phonation: a case report. *Brain Lang*, *15*(2), 234-48.

Jürgens, U., Maurus, M., Ploog, D., & Winter, P. (1967). Vocalization in the squirrel monkey (saimiri sciureus) elicited by brain stimulation. *Exp Brain Res*, *4*(2), 114-7.

Jürgens, U., & Ploog, D. (1970). Cerebral representation of vocalization in the squirrel monkey. *Exp Brain Res*, *10*(5), 532-54.

Juslin, P. (1997a). Perceived emotional expression in synthesized performances of a short melody: Capturing the listener's judgment policy. *Musicae Scientiae*, *1*, 225-256.

Juslin, P. (1997b). Emotional communication in music performance: A functionalist perspective and some data. *Music Perception*, *14*(4), 383-418.

Juslin, P. (2001). Communicating emotion in music performance: A review and theoretical framework. In P. Juslin & J. Sloboda (Eds.), *Music and emotion* (p. 309-337). Oxford: Oxford University Press.

Juslin, P., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: different channels, same code? *Psychological Bulletin*, *129*(5), 770-814.

Juslin, P., & Laukka, P. (2004). Expression, perception, and induction of musical emotion: a review and a questionnaire study of everyday listening. *Journal of New Music Research*, *33*(3), 217-238.

Kawasaki, H., Adolphs, R., Oya, H., Kovach, C., Damasio, H., Kaufman, O., et al. (2005). Analysis of single-unit responses to emotional scenes in human ventromedial prefrontal cortex. *J Cogn Neurosci*, *17*(10), 1509-18.

Kayser, J., Bruder, G. E., Tenke, C. E., Stewart, J. E., & Quitkin, F. M. (2000). Event-related potentials (ERPs) to hemifield presentations of emotional stimuli: differences between depressed patients and healthy adults in P3 amplitude and asymmetry. *Int J Psychophysiol*, *36*(3), 211-36.

Keidel, W. D. (1971). D.C.-potentials in the auditory evoked response in man. *Acta Otolaryngol*, *71*(2), 242-8.

Keil, A., Bradley, M. M., Hauk, O., Rockstroh, B., Elbert, T., & Lang, P. J. (2002). Large-scale neural correlates of affective picture processing. *Psychophysiology*, *39*(5), 641-9.

Khalfa, S., Schön, D., Anton, J. L., & Liegeois-Chauvel, C. (2005). Brain regions involved in the recognition of happiness and sadness in music. *Neuroreport*, *16*(18), 1981-4.

Kivy, P. (1989). *Sound sentiment*. Philadelphia: Temple University Press.

Kleinpaul, R. (1888/1972). *Sprache ohne Worte: Idee einer allgemeinen Wissenschaft der Sprache*. The Hague: Mouton.

Klem, G. H., Luders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty

electrode system of the international federation. The International Federation of Clinical Neurophysiology. *Electroencephalogr Clin Neurophysiol Suppl*, *52*, 3-6.

Kodera, K., Hink, R. F., Yamada, O., & Suzuki, J. I. (1979). Effects of rise time on simultaneously recorded auditory-evoked potentials from the early, middle and late ranges. *Audiology*, *18*(5), 395-402.

Koelsch, S., Fritz, T., Cramon, D. von, Müller, K., & Friederici, A. D. (2006). Investigating emotion with music: An fMRI study. *Hum Brain Mapp*, *27*(3), 239-50.

Koelsch, S., Gunter, T. C., Cramon, D. Y. v, Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: a cortical "language-network" serves the processing of music. *Neuroimage*, *17*(2), 956-66.

Koelsch, S., Wittfoth, M., Wolf, A., Müller, J., & Hahne, A. (2004). Music perception in cochlear implant users: an event-related potential study. *Clin Neurophysiol*, *115*(4), 966-72.

Kohlmetz, C., Müller, S. V., Nager, W., Münte, T. F., & Altenmüller, E. (2003). Selective loss of timbre perception for keyboard and percussion instruments following a right temporal lesion. *Neurocase*, *9*(1), 86-93.

Konishi, T., Niimi, S., & Imaizumi, S. (2000). Vibrato and emotion in the singing voice. In G. Woods, G. Luck, R. Brochard, F. Seddon, & J. Sloboda (Eds.), *Proceedings of the sixth international conference for music perception and cognition (icmpc).* Keele, England.

Kotlyar, G., & Morozov, V. (1976). Acoustical correlates of the emotional content of vocalized speech. *Soviet Physics - Acoustics*, *22*(3), 208-211.

Kotz, S. A., Meyer, M., Alter, K., Besson, M., Cramon, D. Y. von, & Friederici, A. D. (2003). On the lateralization of emotional prosody: an event-related functional MR investigation. *Brain Lang*, *86*(3), 366-76.

Kraus, N., & Nicol, T. (2005). Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends Neurosci*, *28*(4), 176-81.

Kriegstein, K. von, Egera, E., Kleinschmidt, A., & Giraud, A. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res Cogn Brain Res*, *17*, 48-55.

Kruskal, J. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*, 1-27.

Kruskal, J. (1964b). Non-metric multidimensional scaling: a numerical method. *Psychometrika*, *29*, 115-129.

Kunej, D., & Turk, I. (2000). New perspectives on the beginnings of music: archeological and musicological analysis of a middle paleolithic bone "flute". In N. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (p. 235-268). Cambridge: The MIT Press.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science*, *197*(4305), 792-5.

Lakshminarayanan, K., Ben Shalom, D., Wassenhove, V. van, Orbelo, D., Houde, J., & Poeppel, D. (2003). The effect of spectral manipulations on the identification of affective and linguistic prosody. *Brain Lang*, *84*(2), 250-63.

Lancker, D. van, & Sidtis, J. J. (1992). The identification of affective-prosodic stimuli

by left- and right-hemisphere-damaged subjects: all errors are not created equal. *J Speech Hear Res*, *35*(5), 963-70.

Lancker, D. R. van, & Canter, G. J. (1982). Impairment of voice and face recognition in patients with hemispheric damage. *Brain Cogn*, *1*(2), 185-95.

Lang, P., Bradley, M., & Cuthbert, B. (1995). *International affective picture system (iaps): Technical manual and affective ratings*. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Lang, P., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation and action. In P. Lang, R. Simons, & M. Balaban (Eds.), *Attention and orienting: Sensory and motivational processes* (p. 97-136). Hillsdale, NJ: Erlbaum.

Lange, C. (1887). *über Gemuethsbewegungen*. Leipzig: Theodor Thomas.

Lattner, S., Maess, B., Wang, Y., Schauer, M., Alter, K., & Friederici, A. D. (2003). Dissociation of human and computer voices in the brain: evidence for a preattentive gestalt-like perception. *Hum Brain Mapp*, *20*(1), 13-21.

Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Hum Brain Mapp*, *24*(1), 11-20.

Laukka, P. (2003). Categorical perception of emotion in vocal expression. *Annals of the New York Acadademy of Sciences*, *1000*, 283-287.

Lazarus, R. (1991). *Emotion and adaptation*. New York: Oxford University Press.

LeDoux, J. E. (1989). Cognitive-emotional interactions in the brain. *Cognition and Emotion*, *3*, 267-289.

LeDoux, J. E. (1991). Emotion and the brain. *Journal of NIH research*, *3*, 49-51.

LeDoux, J. E. (2000). Emotion circuits in the brain. *Annu Rev Neurosci*, *23*, 155-84.

LeDoux, J. E. (2002). The emotional brain revisited. In J. E. LeDoux (Ed.), *Synaptic self: How our brain becomes who we are* (p. 200-234). New York: Viking Penguin.

Liegeois-Chauvel, C., Peretz, I., Babai, M., Laguitton, V., & Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain*, *121 (Pt 10)*, 1853-67.

Loewy, D. H., Campbell, K. B., & Bastien, C. (1996). The mismatch negativity to frequency deviant stimuli during natural sleep. *Electroencephalogr Clin Neurophysiol*, *98*(6), 493-501.

Mah, L., Arnold, M. C., & Grafman, J. (2004). Impairment of social perception associated with lesions of the prefrontal cortex. *Am J Psychiatry*, *161*(7), 1247-55.

Mansuripur, M. (1987). *Introduction to information theory*. Englewood Cliffs, NJ: Prentiss Hall.

Marks, L. E. (2004). Cross-modal interactions in speeded classification. In G. A. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processes* (p. 85-106). Cambridge, MA: MIT Press.

Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory and visual discrimination. *Int J Psychophysiol*, *50*(1-2), 125-45.

Marshall, S., & Cohen, A. (1988). Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Perception*, *6*(1), 95-112.

Massaro, W., & Egan, P. (1996). Perceiving affect from the voice and the face. *Psychodynamic bulletin & review*, *3*(2), 215-221.

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: the cognitive psychology of human audition* (p. 146-198). Oxford: Oxford University Press.

McAdams, S., Winsberg, S., Donnadieu, S., Soete, G. de, & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychol Res*, *58*(3), 177-92.

McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroencephalogr Clin Neurophysiol*, *62*(3), 203-8.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-8.

Meyer, J. (2004). *Akustik und musikalische Aufführungspraxis*. Frankfurt am Main: Verlag Erwin Bochinsky.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annu Rev Neurosci*, *24*, 167-202.

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res*, *14*(1), 115-28.

Molino, J. (2000). Toward an evolutionary theory of music and language. In N. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (p. 165-176). Cambridge: The MIT Press.

Mondloch, C. J., & Maurer, D. (2004). Do small white balls squeak? Pitch-object correspondences in young children. *Cogn Affect Behav Neurosci*, *4*(2), 133-6.

Moon, C. M., & Fifer, W. P. (2000). Evidence of transnatal auditory learning. *J Perinatol*, *20*(8 Pt 2), S37-44.

Moore, B. (2004). *An introduction to the psychology of hearing*. Amsterdam: Elsevier.

Morita, Y., Morita, K., Yamamoto, M., Waseda, Y., & Maeda, H. (2001). Effects of facial affect recognition on the auditory P300 in healthy subjects. *Neurosci Res*, *41*(1), 89-95.

Morris, J., Scott, S., & Dolan, R. (1999). Saying it with feeling: neural responses to emotional vocalizations. *Neuropsychologia*, *37*, 1155-1163.

Münte, T. F., Brack, M., Grootheer, O., Wieringa, B. M., Matzke, M., & Johannes, S. (1998). Brain potentials reveal the timing of face identity and expression judgments. *Neurosci Res*, *30*(1), 25-34.

Münte, T. F., Urbach, T., Düzel, E., & Kutas, M. (2000). Event-related potentials in the study of human cognition and neuropsychology. In F. Boller, J. Grafman, & G. Rizzolatti (Eds.), *Handbook of neuropsychology* (Vol. 1, p. 139-235). Elsevier Science.

Näätänen, R. (1992). *Attention and brain function*. Hillsdale: Erlbaum.

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

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in

mismatch negativity (MMN): a review of the evidence. *Psychophysiology*, *42*(1), 25-32.

Näätänen, R., & Michie, P. T. (1979). Early selective-attention effects on the evoked potential: a critical review and reinterpretation. *Biol Psychol*, *8*(2), 81-136.

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., & Sams, M. (1987). The mismatch negativity to intensity changes in an auditory stimulus sequence. *Electroencephalogr Clin Neurophysiol Suppl*, *40*, 125-31.

Näätänen, R., Paavilainen, P., & Reinikainen, K. (1989). Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? *Neurosci Lett*, *107*(1-3), 347-52.

Näätänen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., & Sokolov, E. N. (1988). Frequency and location specificity of the human vertex N1 wave. *Electroencephalogr Clin Neurophysiol*, *69*(6), 523-31.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). "primitive intelligence" in the auditory cortex. *Trends Neurosci*, *24*(5), 283-8.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus presentation in cognitive neuroscience. *Psychological bulletin*, *6*, 826-859.

Nagel, F., Kopiez, R., Grewe, O., & Altenmüller, E. (in press). 'EMuJoy' software for continuous measurement of perceived emotions in music: Basic aspects of data recording and interface features. *Behavior Research Methods*.

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., et al. (2001). Neural substrates for recognition of familiar voices: a pet study. *Neuropsychologia*, *39*(10), 1047-54.

Naumann, E., Bartussek, D., Diedrich, O., & Laufer, M. (1992). Assessing cognitive and affective information processing functions of the brain by means of the late positive complex of the event-related potential. *Journal of Psychophysiology*, *6*, 285-298.

Neuner, F., & Schweinberger, S. R. (2000). Neuropsychological impairments in the recognition of faces, voices, and personal names. *Brain Cogn*, *44*(3), 342-66.

Niedenthal, P., & Halberstadt, J. (2000). Emotional response as conceptual coherence. In E. Eich, J. F. Kihlstrom, G. Bower, J. Forgas, & P. Niedenthal (Eds.), *Cognition and emotion* (p. 169-203). Oxford: Oxford University Press.

Öhman, A. (1986). Face the beast and fear the face: animal and social fears as prototypes for evolutionary analyses of emotion. *Psychophysiology*, *23*(2), 123-45.

Ortony, A., Clore, G., & Collins, A. (1999). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.

Owings, D., & Morton, E. (1998). *Animal vocal communication: A new approach*. Cambridge: Cambridge University Press.

Paavilainen, P., Karlsson, M. L., Reinikainen, K., & Näätänen, R. (1989). Mismatch negativity to change in spatial location of an auditory stimulus. *Electroencephalogr Clin Neurophysiol*, *73*(2), 129-41.

Palomba, D., Angrilli, A., & A, M. (1997). Visual evoked potentials, heart rate responses

and memory to emotional pictorial stimuli. *International Journal of Psychophysiology*, *27*, 55-67.

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. Oxford: Oxford University Press.

Papoušek, M., Bornstein, M., Nuzzo, C., Papoušek, H., & Symmes, D. (1990). Infant responses to prototypical melodic contours in parental speech. *Infant Behavior and Development*, *13*, 539-545.

Pell, M. D., & Leonard, C. L. (2003). Processing emotional tone from speech in Parkinson's disease: a role for the basal ganglia. *Cogn Affect Behav Neurosci*, *3*(4), 275-88.

Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, *113 (Pt 4)*, 1185-205.

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68*, 111-141.

Pfefferbaum, A., & Ford, J. M. (1988). ERPs to stimuli requiring response production and inhibition: effects of age, probability and visual noise. *Electroencephalogr Clin Neurophysiol*, *71*(1), 55-63.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: an MEG mismatch study. *J Cogn Neurosci*, *12*(6), 1038-55.

Picton, T. W., Alain, C., Otten, L., Ritter, W., & Achim, A. (2000). Mismatch negativity: different water in the same river. *Audiol Neurootol*, *5*(3-4), 111-39.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., et al. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, *37*(2), 127-52.

Picton, T. W., Goodman, W. S., & Bryce, D. P. (1970). Amplitude of evoked responses to tones of high intensity. *Acta Otolaryngol*, *70*(2), 77-82.

Pihan, H., Altenmüller, E., Hertrich, I., & Ackermann, H. (2000). Cortical activation patterns of affective speech processing depend on concurrent demands on the subvocal rehearsal system. a DC-potential study. *Brain*, *123 (Pt 11)*, 2338-49.

Pittam, J., & Scherer, K. R. (1993). Vocal expression and communication of emotion. In M. Lewis & J. Haviland (Eds.), *Handbook of emotions.* New York, London: The Guilford Press.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Theories of emotion* (Vol. 1, p. 3-33). New York: Academic Press.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech communication*, *41*, 245-255.

Pourtois, G., Gelder, B. de, Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport*, *11*(6), 1329-33.

Pratto, F., & John, O. (1991). Automatic vigilance: The attention-grabbing power of negative information. *Journal of Personality and Social Psychology*, *61*, 380-391.

Pritchard, W. S. (1981). Psychophysiology of P300. *Psychol Bull*, *89*(3), 506-40.

Rapoport, E. (1996). Emotional expression code in opera and lied singing. *Journal of New Music Research*, *25*(2), 109-149.

Roach, P., Stibbard, R., Osborne, J., Arnfield, S., & Setter, J. (1998). Transcription of prosodic and paralinguistic features of emotional speech. *Journal of the international phonetic association*, *28*, 83-94.

Roseman, I., Spindel, M., & Jose, P. (1990). Appraisal of emotion-eliciting events: Testing a theory of discrete emotions. *Journal of Personality and Social Psychology*, *59*, 899-915.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci*, *336*(1278), 367-73.

Ross, E. D. (1981). The aprosodias. functional-anatomic organization of the affective components of language in the right hemisphere. *Arch Neurol*, *38*(9), 561-9.

Ross, E. D., Edmondson, J. A., Seibert, G. B., & Homan, R. W. (1988). Acoustic analysis of affective prosody during right-sided wada test: a within-subjects verification of the right hemisphere's role in language. *Brain Lang*, *33*(1), 128-45.

Rothkopf, E. Z. (1957). A measure of stimulus similarity and errors in some paired-associate learning tasks. *J Exp Psychol*, *53*(2), 94-101.

Rugg, M. D., & Coles, M. G. H. (1996). The ERP and cognitive psychology: conceptual issues. In M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind, event-related brain potentials and cognition* (p. 27-39). Oxford: Oxford University Press.

Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161-1178.

Russell, J., & Barrett, L. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*, 805-819.

Sams, M., Kaukoranta, E., Hamalainen, M., & Näätänen, R. (1991). Cortical activity elicited by changes in auditory stimuli: different sources for the magnetic N100m and mismatch responses. *Psychophysiology*, *28*(1), 21-9.

Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalogr Clin Neurophysiol*, *62*(6), 437-48.

Samson, S., Zatorre, R. J., & Ramsay, J. O. (2002). Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. *Brain*, *125*(Pt 3), 511-23.

Sander, K., Brechmann, A., & Scheich, H. (2003). Audition of laughing and crying leads to right amygdala activation in a low-noise fMRI setting. *Brain Res Brain Res Protoc*, *11*(2), 81-91.

Sander, K., & Scheich, H. (2005). Left auditory cortex and amygdala, but right insula dominance for human laughing and crying. *J Cogn Neurosci*, *17*(10), 1519-31.

Savela, J., Kujala, T., Tuomainen, J., Ek, M., Aaltonen, O., & Näätänen, R. (2003). The mismatch negativity and reaction time as indices of the perceptual distance between the corresponding vowels of two related languages. *Brain Res Cogn Brain Res*, *16*(2), 250-6.

Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychol Rev*, *69*, 379-99.

Schapkin, S. A., Gusev, A. N., & Kuhl, J. (2000). Categorization of unilaterally presented emotional words: an ERP analysis. *Acta Neurobiol Exp (Wars)*, *60*(1), 17-28.

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2005). Vocal expression of emotions in normally hearing and hearing-impaired infants. *J Voice*.

Scherer, K. R. (1982a). Methods of research on vocal communication: paradigms and parameters. In K. R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (p. 137-198). Cambridge: Cambridge University Press.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*, 143-165.

Scherer, K. R. (1988). On the symbolic function of vocal affect expression. *Journal of language and social psychology*, *7*(2), 79-100.

Scherer, K. R. (1995). Expression of emotion in voice and music. *J Voice*, *9*(3), 235-48.

Scherer, K. R. (2000). Psychological models of emotion. In J. Borod (Ed.), *The neuropsychology of emotion* (p. 137-166). Oxford/New York: Oxford University Press.

Scherer, K. R. (2001). The nature and study of appraisal: a review of the issue. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (p. 369-391). Oxford: Oxford University Press.

Scherer, K. R., Banse, R., Wallbott, H., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, *15*(2), 123-148.

Scherer, K. R., & Zentner, M. (2001). Emotional effects of music: Production rules. In P. Juslin & J. Sloboda (Eds.), *Music and emotion* (p. 361-392). Oxford: Oxford University Press.

Scherg, M., Vajsar, J., & Picton, T. (1989). A source analysis of human auditory evoked potentials. *Journal of Cognitive Neuroscience*, *1*, 336-355.

Schirmer, A., & Kotz, S. A. (2003). ERP evidence for a sex-specific stroop effect in emotional speech. *J Cogn Neurosci*, *15*(8), 1135-48.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn Sci*, *10*(1), 24-30.

Schirmer, A., Striano, T., & Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport*, *16*(6), 635-9.

Schlosberg, H. (1954). Three dimensions of emotion. *Psychological Review*, *61*, 81-88.

Schönwiesner, M., Rübsamen, R., & Cramon, D. Y. von. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur J Neurosci*, *22*(6), 1521-8.

Schrader, L., & Todt, D. (1993). Contact call parameters covary with social context in common marmosets. *Animal Behaviour*, *46*, 1026-1028.

Schroeger, E. (1997). On the detection of auditory deviations: a pre-attentive activation model. *Psychophysiology*, *34*(3), 245-57.

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Birbaumer, N., & Lang, P. J. (1997).

Probe P3 and blinks: two measures of affective startle modulation. *Psychophysiology*, *34*(1), 1-6.

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., & Lang, P. J. (2000). Affective picture processing: the late positive potential is modulated by motivational relevance. *Psychophysiology*, *37*(2), 257-61.

Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., & Johnson, M. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature*, *385*(6613), 254-7.

Shaw, P., Bramham, J., Lawrence, E. J., Morris, R., Baron-Cohen, S., & David, A. S. (2005). Differential effects of lesions of the amygdala and prefrontal cortex on recognizing facial expressions of complex emotions. *J Cogn Neurosci*, *17*(9), 1410-9.

Silva, F. Lopes da. (1991). Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalogr Clin Neurophysiol*, *79*(2), 81-93.

Sloboda, J. (1990). Empirical studies of the emotional response to music. In M. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication* (p. 33-46). Washington: American Psychological Association.

Sokolowski, K. (2002). Emotion. In J. Müsseler & W. Prinz (Eds.), *Allgemeine Psychologie* (p. 337-384). Heidelberg: Spektrum Akademischer Verlag.

Solomon, R. (2004). *Thinking about feeling : Contemporary philosophers on emotions*. Oxford: Oxford University Press.

Speckmann, E.-J., & Elger, C. (2005). Introduction to the neurophysiological basis of the EEG and DC potentials. In E. Niedermeyer & F. Lopes da Silva (Eds.), *Electroencephalography: basic principles, clinical applications and related fields.* Philadelphia: Lippincott Williams & Wilkins.

Stein, B., & Meredith, M. (1993). *The merging of the senses*. Cambridge, Massachusetts: The MIT Press.

Sugg, M. J., & Polich, J. (1995). P300 from auditory stimuli: intensity and frequency effects. *Biol Psychol*, *41*(3), 255-69.

Sundberg, J. (1987). *The science of the singing voice*. Dekalb, Illinois: Northern Illinois University Press.

Sundberg, J. (1999). The perception of singing. In D. Deutsch (Ed.), *The psychology of music* (p. 171-214). London: Academic Press.

Sussman, E., Gomes, H., Nousak, J. M., Ritter, W., & Vaughan, J., H. G. (1998). Feature conjunctions and auditory sensory memory. *Brain Res*, *793*(1-2), 95-102.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schroeger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Exp Brain Res*, *161*(1), 1-10.

Tervaniemi, M., Maury, S., & Näätänen, R. (1994). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *Neuroreport*, *5*(7), 844-6.

Tervaniemi, M., Winkler, I., & Näätänen, R. (1997). Pre-attentive categorization of sounds by timbre as revealed by event-related potentials. *Neuroreport*, *8*(11), 2571-4.

Thesen, T., Vibell, J., Calvert, G., & Österbauer, R. (2004). Neuroimaging of multisensory processing in vision, audition, touch and olfaction. *Cognitive Processing*, *5*, 84-93.

Tian, B., Reser, D., Durham, A., Kustov, A., & Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science*, *292*(5515), 290-3.

Titova, N., & Näätänen, R. (2001). Preattentive voice discrimination by the human brain as indexed by the mismatch negativity. *Neurosci Lett*, *308*(1), 63-5.

Toiviainen, P., Tervaniemi, M., Louhivuori, J., Saher, M., Huotilainen, R., & Näätänen, R. (1998). Timbre similarity: convergence of neural, behavioral, and computational approaches. *Music Perception*, *16*, 223-241.

Traunmüller, H. (1997). Perception of speaker sex, age, and vocal effort. *Phonum*, *4*, 183-186.

Twist, D., Squires, N., Spielholz, N., & Silverglide, R. (1991). Event-related potentials in disorders of prosodic and semantic linguistic processing. *Neuropsychiatry, Neuropsychology, Behavioral Neurology*, *4*, 281-304.

Vaina, L. M. (1994). Functional segregation of color and motion processing in the human visual cortex: clinical evidence. *Cereb Cortex*, *4*(5), 555-72.

Verleger, R. (1988). Event - related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of p3. *Behavioral and Brain Sciences*, *11*, 343-356.

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proc Natl Acad Sci U S A*, *100*(17), 10038-42.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychol Bull*, *98*(2), 219-35.

Wayman, J. W., Frisina, R. D., Walton, J. P., Hantz, E. C., & Crummer, G. C. (1992). Effects of musical training and absolute pitch ability on event-related activity in response to sine tones. *J Acoust Soc Am*, *91*(6), 3527-31.

Welch, R., & Warren, D. (1986). Intersensory interactions. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (Vol. I: Sensory Processes and Perception, p. 25/1-25/36). New York: Wiley.

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Curr Opin Neurobiol*, *8*(2), 227-33.

Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., et al. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cereb Cortex*, *14*(12), 1384-9.

Wildgruber, D., Pihan, H., Ackermann, H., Erb, M., & Grodd, W. (2002). Dynamic brain activation during processing of emotional intonation: influence of acoustic parameters, emotional valence, and sex. *Neuroimage*, *15*(4), 856-69.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., et al. (2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage*, *24*(4), 1233-41.

Windmann, S., Daum, I., & Güntürkün, O. (2002). Dissociating prelexical and postlexical

processing of affective information in the two hemispheres: effects of the stimulus presentation format. *Brain Lang*, *80*(3), 269-86.

Windmann, S., & Kutas, M. (2001). Electrophysiological correlates of emotion-induced recognition bias. *J Cogn Neurosci*, *13*(5), 577-92.

Winkler, I., Tervaniemi, M., Huotilainen, M., Ilmoniemi, R., Ahonen, A., Salonen, O., et al. (1995). From objective to subjective: pitch representation in the human auditory cortex. *Neuroreport*, *6*(17), 2317-20.

Winkler, I., Tervaniemi, M., & Näätänen, R. (1997). Two separate codes for missing-fundamental pitch in the human auditory cortex. *J Acoust Soc Am*, *102*(2 Pt 1), 1072-82.

Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalogr Clin Neurophysiol*, *79*(3), 170-91.

Wundt, W. (1900). *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte*. Leipzig: Kröner Verlag.

Young, A. W., Newcombe, F., Haan, E. H. de, Small, M., & Hay, D. C. (1993). Face perception after brain injury. selective impairments affecting identity and expression. *Brain*, *116 (Pt 4)*, 941-59.

Zajonc, R. B. (1985). Emotion and facial efference: a theory reclaimed. *Science*, *228*(4695), 15-21.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb Cortex*, *11*(10), 946-53.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends Cogn Sci*, *6*(1), 37-46.

Zatorre, R. J., Bouffard, M., & Belin, P. (2004). Sensitivity to auditory object features in human temporal neocortex. *J Neurosci*, *24*(14), 3637-42.

List of Figures

2.1. Circumplex model 8
2.2. The emotional processing circuit 11
2.3. Intro - example of harmonic partials 15
2.4. The lens model 27
2.5. Model for processing of emotional prosody 30

3.1. Generation of the EEG signal 35
3.2. Recording event-related potentials 36

4.1. Electrode setup for exp. I 48
4.2. MMN-Exp. I - reaction times in active condition 50
4.3. Active condition – Effect of 'deviant type' 51
4.4. Active condition – Effect of emotion 51
4.5. MMN-Exp I – Grand averages 53
4.6. MMN-Exp I – Difference waves 56
4.7. Stimulus waveform 57

5.1. Electrode setup for MMN-Exp II 75
5.2. Grand averages for MMN-Exp. II 78
5.3. MMN-Exp II - Tones arranged in 3-dimensional space 80
5.4. Relative amount of same responses for happy and sad tones 82

6.1. Model of voice perception by Belin et al. (2004) 90
6.2. Electrode setup for Exp. II-01 96
6.3. Exp. II-01, grand averages: main effects 99

6.4. Exp. II-01, grand averages: emotion matching task 100
6.5. Exp. II-01, grand averages: identity matching task 101
6.6. Schema of the experimental design 104
6.7. Exp. II-01, P2 effect in the emotion matching task 105
6.8. Exp. II-01 - Difference waves 105
6.9. Exp. II-01, reaction times 106

7.1. Exp II-02 - effects of unattended picture valnce 122
7.2. Effects of (unattended) voice valence 123
7.3. Exp. II-02 - Effects of (unattended) picture valence 124
7.4. Exp. II-02 - Effects of attended voice valence 125
7.5. Exp. II-02 - task effect 126

.1. Appendix - spectrograms of stimuli in MMN-Exp II 162
.2. Appendix - vibrato of sad stimuli in MMN-II 163
.3. Appendix - vibrato of happy stimuli in MMN-II 163

List of Tables

2.1. Parameters of vocal affect expression 14
2.2. Parameters of musical affect expression 22

4.1. Active condition – reaction-times 50
4.2. MMN-Exp I – F-values 54

5.1. Scaling-Exp – Stimulus material 63
5.2. Scaling-Exp – Discrimination probabilities 68
5.3. Scaling-Exp – Fechnerian Distances 68
5.4. Affect-rating 69
5.5. Arousal-rating 69
5.6. Valence-rating 70
5.7. Affect-rating – post-hoc-comparisons 71
5.8. MMN-Exp II – Design 73
5.9. MMN-Exp II – F-values 76
5.10. Discussion – scaling results 82
5.11. MMN-Exp II - acoustical structure of sad tones 83

6.1. Design of experiment II-01 94
6.2. Exp. II-01 – F-values 103
6.3. Exp. II-01 – Reaction times 106

7.1. Exp II-02 - behavioral data 118
7.2. Exp II-02 - effects of picture valence 120

Appendix

Technique to express certain emotions as reported by the violinists recorded for experiment MMN-II. Abbreviations: Geschw.=Geschwindigkeit, schn.=schnell.

Name	fröhlich	traurig	neutral
G01	schnelles Vibrato schn. Bogengeschw.	langsameres Vibrato leichteres Bogengewicht langs. Bogengeschw.	wenig Vibrato normale Bogengeschw.
G02	kurz ein bisschen akzentuieren schnelles Vibrato	langsameres Vibrato unakzentuiert	kein Vibrato
G03	crescendo, eher forte mit Attacke schnelles Vibrato bis zum Ende vibriert schn. Bogengeschw.	decrecendo, eher piano weich angesetzt Vibrato spät entwickelt langsameres Vibrato langs. Bogengeschwindigkeit	mezzoforte bis forte von Anfang bis Ende mittlere Bogengeschw. mittleres Vibrato, bis Ende durchvibriert
G04	laut kleines schnelles Vibrato harter Bogenansatz Abstrich	leise langsameres Vibrato, bis gar kein Vibrato “fahler Ton” beissender Ton Abstrich	mittel laut kein besonders Vibrato keine besonders charakt. Bogenstelle, Mitte
G05	schnellere Geschw. schnelles Vibrato akzentuierter	langsamer, sachter, zarter weniger o. viel Bogen wenig Druck grösseres Vibrato weichere Übergänge	wenig bis gar kein Vibrato mittelschnell gleichbleibender
G06	viel Vibrato kurze Notenwerte laut schnell	wenig Vibrato weicher Ton leise langsam	etwas Vibrato laut breite Noten
G07	forte näher zum Steg akzentuierter	leise wenig Druck, leicher Bogen näher zum Griffbrett	möglichst wenig Ausdruck mittlere Dynamik weniger Vibrato
G08	Vibrato am Anfang schnell offensive Dynamik	wenig Vibrato langsameres Vibrato Bogen langsam ziehen weniger offensiv, weicher	ohne Vibrato Bogen schnell gezogen

Results of pre-rating of large tone set to find stimulus material for MMN-Exp. II

sound no	first resp. mean (N=9)	SD	sec. resp. mean (N=8)	SD	selected tones
1	2.33	0.87	2.63	0.52	
2	3.00	0.87	2.25	0.89	
3	2.11	0.33	2.50	0.93	tone01
4	2.56	0.73	1.88	0.99	
5	1.89	0.78	1.75	0.46	tone02
6	2.22	0.67	2.25	0.71	
7	3.11	0.93	2.63	0.74	
8	2.78	0.97	2.50	0.93	
9	1.67	0.50	1.50	0.53	
10	1.78	0.67	1.88	0.83	
11	4.00	0.50	3.75	0.89	
12	4.78	0.44	4.88	0.35	tone03
13	4.33	1.00	4.38	0.52	tone04
14	4.22	0.83	3.88	0.99	
15	2.78	0.67	2.63	0.52	tone05
16	2.56	0.88	2.25	0.46	
17	4.22	0.97	4.00	0.76	
18	4.11	0.60	4.38	0.52	tone06
19	4.00	1.00	4.50	0.53	
20	4.44	0.53	4.38	0.74	tone07
21	4.00	1.22	4.50	0.76	tone08
22	3.67	1.00	3.25	0.89	
23	3.89	0.60	3.75	0.71	
24	2.78	1.30	2.75	1.04	
25	2.00	1.32	1.75	1.04	
26	1.89	1.36	1.38	0.52	
27	3.44	0.73	3.50	0.93	
28	3.78	0.67	3.75	0.71	tone09
29	3.56	1.01	3.13	0.64	tone10
30	3.44	0.53	3.38	1.30	
31	2.89	1.05	3.13	0.64	
32	1.89	0.78	2.25	0.46	
33	2.44	0.73	2.88	0.83	
34	2.11	0.33	2.75	0.71	
35	2.67	0.87	2.75	0.89	

Psychoacoustical analysis of the stimuli of MMN-exp. II Fig. .1. depicts the sound spectrograms of the three sad (left column) and the three happy tones (right column). In a sound spectrogram the horizontal dimension corresponds to time (reading from left to right), and the vertical dimension corresponds to frequency (or pitch), with higher sounds shown higher on the display. The relative intensity of the sound at any particular time and frequency is indicated by the darkness of the spectrogram at that point.

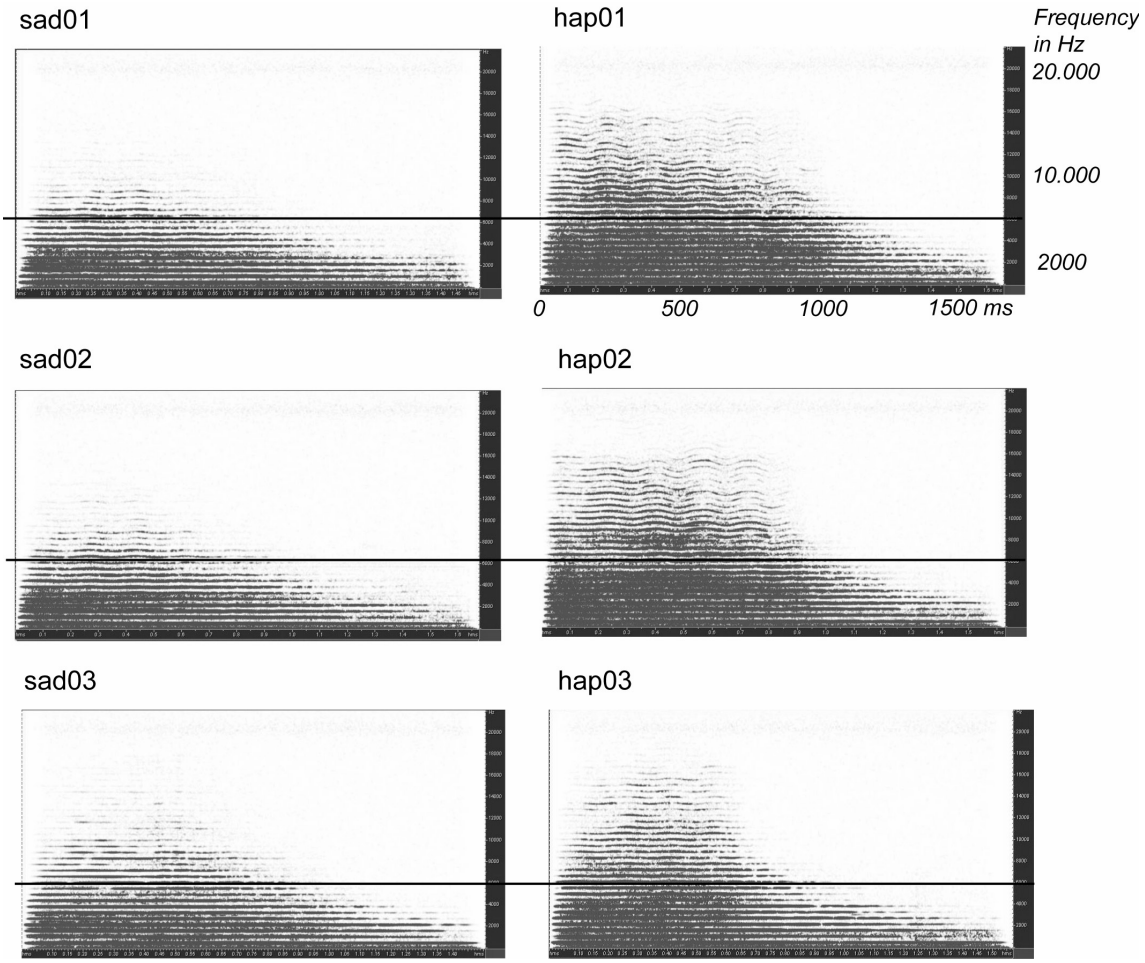

Figure .1.: Spectrograms of the stimuli in MMN-Exp II. The cut-off line for the high-frequency energy is indicated by the gray horizontal line.

Fig. .2. and .3. depict the frequency vibrato of the sad and the happy tones between 0 and 1000 ms. The horizontal dimension corresponds to time (reading from left to right), and the vertical dimension corresponds to pitch.

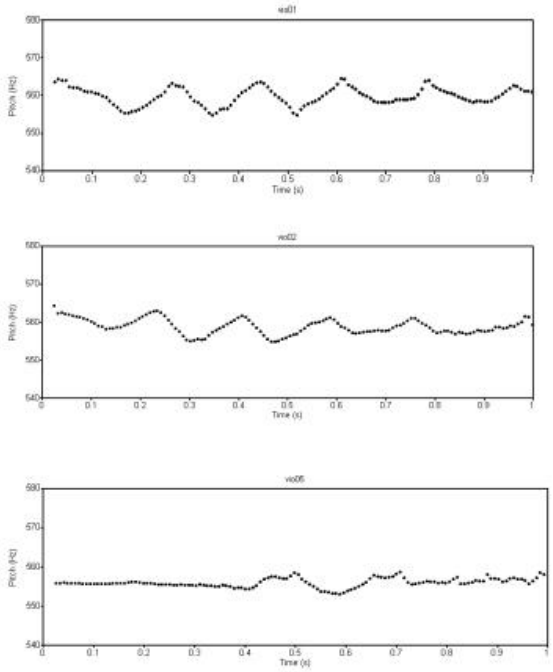

Figure .2.: **Vibrato of sad01, sad02, and sad03.**

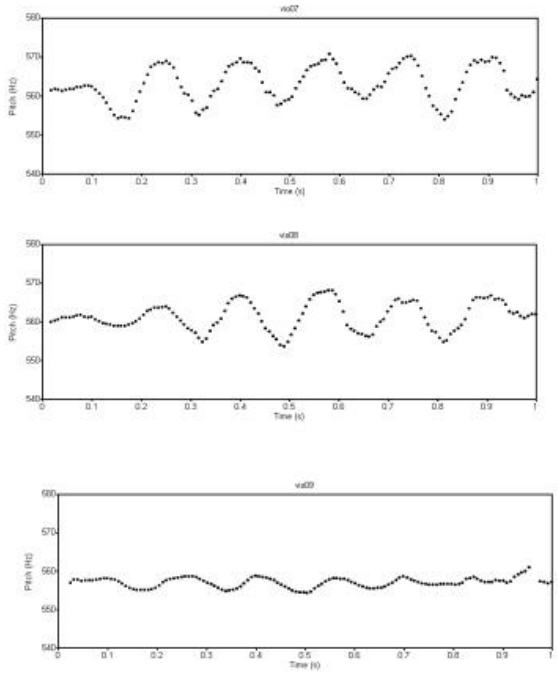

Figure .3.: **Vibrato of hap01, hap02, and hap03**

Erklärung

Hiermit erkläre ich, dass ich die von mir eingereichte Dissertation zu dem Thema

"Neural mechanisms for fast recognition of auditory emotion"

selbständig verfasst, nicht schon als Dissertation verwendet habe und die benutzten Hilfsmittel und Quellen vollständig angegeben wurden.

Weiterhin erkläre ich, dass ich weder diese noch eine andere Arbeit zur Erlangung des akademischen Grades doctor rerum naturalium (Dr. rer. nat.) an anderen Einrichtungen eingereicht habe.

Hannover, den 24.04.2006

K. Spreckelmeyer

Lebenslauf

Katja Spreckelmeyer, geb. Goydke

geboren am 13.04.1976 in Braunschweig
verheiratet,
Staatsangehörigkeit: Deutsch

Ausbildung

2002-2006	Promotionsstudentin and der Otto-von-Guericke-Universität Magdeburg
2003	10-monatiger Forschungsaufenthalt am Department of Cognitive Science an der University of California, San Diego, USA
2002	Diplom in Psychologie
1999-2002	Psychologiestudium an der Heinrich-Heine-Universitt Düsseldorf
1998-1999	Psychologiestudium an der Universität Wien
1998	Vordiplom
1996-1998	Psychologiestudium an der FU Berlin
1995	Abitur am Martino-Katharineum in Braunschweig

Praktische Tätigkeiten

seit 2006	Wissenschaftliche Mitarbeiterin an der Klinik für Psychiatrie und Psychotherapie des Universitätsklinikums Aachen
2002-2006	Wissenschaftliche Mitarbeiterin am Institut für Musikphysiologie und Musikermedizin, Hannover
2000-2002	Studentische Hilfskraft am Institut für physiologische Psychologie der HHU Düsseldorf
2000	Forschungspraktikum am Max-Planck-Institut für Kognitions- und Neurowissenschaften in Leipzig
2000	Tätigkeit als studentische Hilfskraft am Neurologischen Therapiecentrum (NTC) in Düsseldorf
1999	Klinisches Praktikum in der psychiatrischen Abteilung der Charité, Berlin
1995-1996	Freiwilliges Soziales Jahr in der Camphill-Community 'Ochil Tower School' in Auchterarder, Schottland

Auszeichnungen

2003-2006	Promotionsstipendium der Studienstiftung des deutschen Volkes
2003	Auslandsforschungsstipendien (USA) des Deutschen Akademischen Austauschdienst, der GA-Lienert-Stiftung und der G. Daimler- und K. Benz-Stiftung