Author: Witteman, Jurriaan
Title: Towards a cognitive neuroscience of prosody perception and its modulation by alexithymia
Issue Date: 2014-12-18

Towards a Cognitive Neuroscience of Prosody Perception

And its Modulation by Alexithymia

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands

phone: +31 30 253 6111

e-mail: lot@uu.nl
http://www.lotschool.nl

Cover illustration: 'Rootbound', a sculpture by Kate MacDowell. Photo by Bill Bachhuber. Reprinted with permission.

ISBN: 978-94-6093-160-4
NUR 616

Towards a Cognitive Neuroscience of Prosody Perception
And its Modulation by Alexithymia

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C. J. J. M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 18 december 2014
klokke 10.00 uur

door

JURRIAAN WITTEMAN

Promotiecommissie

Promotores:	Prof. dr. Vincent J. J. P. van Heuven
	Prof. dr. Niels O. Schiller
Overige leden:	Dr. Johanneke Caspers
	Prof. Dr. Bernhard Hommel
	Prof. Dr. Marinus H. van IJzendoorn
	Prof. Dr. Marc van Oostendorp
	Prof. Dr. Serge A. R. B. Rombouts

Contents

FIRST EMPIRICAL PART: TOWARDS A COGNITIVE NEUROSCIENCE OF PROSODY PERCEPTION

SECOND EMPIRICAL PART: MODULATION BY ALEXITHYMIA

Acknowledgments

I am indebted to my supervisors, prof. Niels Schiller and prof. Vincent van Heuven for giving me the chance to work on such a fascinating topic. I have very much enjoyed working with you and have learned a lot in the process. Further, I am grateful to prof. Serge Rombouts and colleagues at the Radiology department for their support with the imaging work. Additionally, I want to thank LUCL for the logistic support.

Also, I want to thank my direct colleagues in what used to be the Phonetics lab, Jos Pacilly, Daan van de Velde, Willemijn Heeren, Amanda Post, Junru Wu, Franziska Scholz, Ronya Cui, and Yiya Chen (and many others) for the stimulating and pleasant working environment. I particularly enjoyed the lunch break discussions with Vincent and Jos, during which we have explored many topics ranging from politics to astronomy and which, incidentally, tended to end with Jos teaching us some elementary physics. Also, I would like to thank my indirect colleagues at 'the other side' (the main LUCL building), Margarita, Peppijn, Sita, Jessie, Marijn, Suzanne, Maarten, Annemiek (and many others) for the laughs and sharing 'The PhD-experience' with me.

I have collaborated with various researchers at Leiden University. I am grateful to prof. Marinus van IJzendoorn for collaborating with me and teaching me how to do meta-analysis. Also, I would like to thank dr. Henk van Steenbergen and dr. Ilya Veer for teaching me technical aspects of functional neuroimaging.

My PhD-work has been greatly enhanced by the collaboration with the Neuro Imaging Centre (NIC) Groningen. So, thank you Katharina for both the great times we have had in designing new experiments and writing them up and all the good times we had as we became good friends too. I trust that we stay in touch both as friends and as colleagues in the future. Also, I want to thank dr. Sander Martens and prof. André Aleman for making our collaborative projects possible. Further, I like to thank all the great people I have met at the NIC, so Gemma, Emi, Richard, Johan, Ramona thanks for your company.

Last, I would like to thank my parents, Gonny en Paul, for the constant support they have given me in pursuing my interests.

Chapter 1

General introduction

1.1. Core concepts used in the thesis

Imagine that you are sitting in a train absorbed by a good book and suddenly you hear a mother shouting angrily at her child: "Tim, you are going to annoy the other passengers!" You automatically look into the direction of where the event has just taken place, just as many other passengers probably would.

Apparently, the mother's utterance constitutes a powerful acoustic social signal that is even being processed by your brain when you not intend to. Note that the salience of the utterance is not determined by *what* the mother is saying (i.e., the content of the words she is using) but *how* she is saying it. For instance, the mother could have pronounced exactly same words but with different acoustic characteristics to indicate her surprise or sadness with the fact that Tim is annoying the passengers in the train. This thesis is concerned with the question what network in the human brain processes such acoustic properties of speech, how it does so, and why the network sometimes continues to process these acoustic properties even when we do not pay attention to them.

The succession of the vowels and consonants that make up words (and longer structures) are called segments. We have seen in the example above that it is not the segments of speech that explain how Tim's mother conveyed her anger, because with the same sequence of words she can convey various emotional meanings. The communicative layer of speech that uses the acoustic characteristics that cannot be explained by the mere succession of segments has been termed 'prosody' in linguistics (e.g., Rietveld & Van Heuven, 2009; Hardcastle, Laver, & Gibbon, 2010). The term 'prosody' can roughly be translated from Greek as 'with *(pros)* the song (*oide*)' or 'the musical accompaniment to vowels and consonants that make up the utterance' (Van Heuven & Sluijter, 1996). Indeed, because this 'musical accompaniment' appears to be superimposed on the segmental layer of speech, prosodic aspects of speech have also been called 'suprasegmentals' in linguistics (Lehiste, 1970; Rietveld & Van Heuven, 2009). Note that, fundamentally, 'prosody' therefore is negatively defined (anything that it is not segmental).

Prosody has various communicative functions (Rietveld & Van Heuven, 2009) that seem to be fundamentally different in nature. On the one hand, as we have seen in the example above, prosody can be used to convey *emotional* information. Banse and Scherer (1996) provide a widely accepted definition of 'emotion', i.e. 'an episode of temporary synchronization of all major subsystems of organismic functioning represented by five components (cognition, physiological regulation, motivation, motor expression and monitoring-feeling) in response to internal or external events of major importance to the organism.' Note further that one key characteristic of emotion, and hence of emotional prosody, is that it is dimensional - Tim's mother can express her anger to a certain degree to sound more or less angry. Another key aspect of emotion is that emotions can be qualified as positive or negative (they have 'valence') with respect to the organism experiencing emotion. A last key aspect of emotional processing, which we will return to, is that emotions can signal importance of information with respect to the well-being of the organism. This has the implication that accurate perception of emotional prosody is likely to have had 'fitness value' according to Darwin's theory of evolution – it increases the probability of passing on one's genetic material. For instance, the ability to accurately decipher a dominant group member's angry mental state, and thereby avoiding harm due to physical conflict, is likely to have increased chances of survival and having offspring in our evolutionary past (and, incidentally, it probably still does today). This fitness value of accurate emotional prosody perception renders it likely that the emotional prosody perception network in the human brain has evolved quite early in human evolutionary history and thus is a genetically programmed system that we share with other mammals (Öhman & Mineka, 2001).

On the other hand, prosody can also be used to convey 'linguistic' information. For instance Tim's mother could pronounce the sentence "Tim, you are going to annoy the other passengers?" with a rising tone at the end of the utterance to indicate that an utterance is a question rather than a statement of fact (Rietveld & Van Heuven, 2009). When used linguistically, by definition, prosody conveys information regarding the linguistic structure of the utterance. This means that linguistic prosody conveys information relevant to the abstract rules of language, or in other words, the grammar – phonology, syntax and semantics. Another example of linguistic prosody that will be studied in this thesis is the placement of stress. For instance, by placing stress on the first syllable of the word '**ex**port' (bold indicates a stressed syllable) versus the second syllable 'ex**port'** word meaning changes (from a noun to a verb). Notice again that the segmental layer of speech is identical – it is the suprasegmental characteristic of speech (i.e., linguistic prosody) that is indispensable in deriving word meaning. A key characteristic of linguistic (grammatical) contrasts is that they are *categorical* in nature (Ladd & Morton, 1997; Van Heuven & Kirsner, 2004), i.e., employ categories that obey the grammatical rule system. For instance, if we look at stress, a syllable either has or does not have stress. Further, stress on the second syllable of the word 'neu**ral**' (bold indicates the stressed syllable) is grammatically incorrect. Note that this categorical nature of linguistic prosody is in contrast to the dimensional character of emotional prosody. Further, note that, by definition, linguistic prosodic contrasts are language-specific, whereas at least some aspects of emotional prosody perception have been shown to be culturally (and linguistically) indifferent or 'universal' (Van Bezooijen, Otto, & Heenan, 1983; Scherer, Banse, & Wallbott, 2001), which, incidentally, would also be expected if the underlying emotional prosody perception network is indeed 'hard wired'. Last, notice that while emotional information has valence, linguistic information does not – for instance, placement of stress cannot be called positive or negative, further underscoring the fundamentally different nature of these two prosodic communicative functions.

What acoustic dimensions then, does the prosodic layer of speech use to convey emotional and linguistic meaning? Careful analysis of various acoustic dimensions such as F_0 (pitch), intensity (loudness), duration, spectral distribution (timbre) and the dynamics of these parameters in time, suggest that different vocal emotions have specific 'acoustic profiles', meaning that different emotions use a specific set of values along these dimensions (Banse & Scherer, 1996). Turning to linguistic prosody, the exact acoustic dimensions that are important in conveying meaning depend on the precise linguistic function (e.g., indicating the mode of a sentence, phrase boundary marking or stress). However, when we zoom in on the linguistic prosodic function that will be studied in this thesis, i.e. linguistic stress, it turns out that stressed syllables differ from unstressed syllables in multiple acoustic dimensions – stressed syllables have higher F_0 (pitch), longer duration, greater intensity (loudness) and spectrally more extreme realizations than unstressed syllables (Rietveld & Van Heuven, 2009). In the present thesis, I will examine what neural network analyzes these two communicative functions of prosody, and how (i.e., using what series of neurocognitive operations).

Imagine again that you are sitting in the train being absorbed by a good book and suddenly turning your gaze into the direction of the mother who had just expressed her anger towards her son. This simple observation tells us something very interesting about emotional prosody perception – apparently, the network in your brain analyzes the emotional prosody even when you do not intend to. Further, the network apparently even manages to divert your attention away from your book to devote full attention to the emotional prosodic signal, as is evidenced by the turning of your head. In experimental psychology, such mental operations that occur even when you not intend to, are called 'automatic'. More specifically, automatic processes are psychological processes that occur when there is no intention to process, can continue to operate under low levels of attention, are efficient (rapid) and are hard to control once initiated (Moors & De Houwer, 2006).

Why would processing of anger prosody be automatic? As we have already touched upon, Darwin's theory of evolution could provide an explanation. As pointed out earlier, accurate processing of social signals such as anger likely has had fitness value in our evolutionary history (Silk, 2007) – avoiding the wrath of an angry dominant male by accurately deciphering its angry mental state may have meant the difference between life and death. If true, selective pressure on accurate emotional prosody perception would have caused evolution of a 'hard wired' (i.e., genetically programmed) network dedicated to the perception of emotional prosody (Öhman & Mineka, 2001). Indeed, the perception of at least some emotional categories as communicated through the voice are 'universal', meaning that members of distant cultures (e.g. Koreans and Japanese) can recognize emotional meaning expressed by (for instance) Dutch speakers and vice versa (Scherer et al., 2001; Van Bezooijen, Otto, & Heenan, 1983). Such universality of a capability suggests that it sub-served by a biological substrate that is under genetic control. Thus, the existence of a hard-wired dedicated neural network can explain automaticity of emotional prosody processing, particularly for emotional prosody that would signal danger (such as anger and perhaps fear). In the present thesis, I will examine whether such a specialized neural system exists in the human brain.

If automatic processing of emotional prosody can be explained by an evolved dedicated neural system, what about the processing of a much more recently invented artificial acoustic signal that can powerfully convey emotions? Ask anyone why one listens to music and within a couple of seconds some reference to emotional states will be used. Also, who would not be moved by Johann Sebastian Bach's *Kunst der Fuge*? Clearly then, music, just like 'the music of speech' (prosody), has a powerful capacity to convey emotion. One hypothesis, the 'superexpressive voices hypothesis', even links the two together – music would be such a powerful acoustic medium to convey emotions *because* it mimics and then exaggerates the characteristics of emotional prosody, acting as a supranormal stimulus (Juslin & Västfjäll, 2008) and eliciting strong emotions in the audience by 'emotional contagion' (imitation of the emotional expression by the listener). A clear difference between music and prosody, however, is that music is an artificial signal. More specifically, it has been defined as 'intentionally created non-linguistic acoustical events structured in time and produced in social contexts' (Altenmüller, Kopiez, & Grewe, 2013). Thus, if automaticity of emotional processing can be explained by a dedicated neural system that has (probably) evolved before there was any music, we would not expect automatic processing of emotional music. On the other hand, there is some evidence that recognition of at least some emotional categories in emotional music is universal (Fritz et al., 2009), pointing to a potential fitness value of music perception in evolutionary history and potentially a dedicated neural system for automatic perception of emotional music, too. Alternatively, music may mimic emotional prosody so well that it engages the system dedicated to automatic processing of emotional prosody (Peretz, Aubé, & Armormy, 2013). To shed some light on these issues, I will directly compare perception of emotional prosody and emotional music in this thesis and explore whether automaticity of processing can be demonstrated.

Finally, people do not only show similarities in the way their brains process emotional prosody and emotional music, but they also display differences in affective processing style (Canli, 2004). In this thesis, we will focus on a personality trait that is associated with affective processing style, called 'alexithymia'. Alexithymia can be translated literally as 'having no words (*alexi*) for feelings (*thymia*)'. It is a personality trait that is associated with difficulty in identifying emotions (Sifneos, 1973). People who score in the clinical range on alexithymia have severe difficulty recognizing, identifying and verbalizing emotions. Due to these problems, alexithymic people may be regarded as 'cold' or 'distant' by their social environment, resulting in social problems. Not much is known about the neural processes underlying alexithymia. Further, most previous brain research on alexithymia has focused on how alexithymia influences neural processing of emotion in the visual modality (i.e., through facial expression). In the present thesis I will therefore explore whether, and if so how, alexithymia modulates activity in the emotional prosody perception network. Due to practical limitations however, only modulation by alexithymia in the non-clinical (normal) range will be examined.

Before we turn to neurocognitive models of emotional prosody perception and modulation of emotional prosody perception by alexithymia, I will briefly discuss the methods used in this thesis to answer the questions raised.

1.2. Methods used in the thesis

1.2.1. Behavioral methods

How can behavior inform us about the neural network involved in prosody perception? Two behavioral approaches that have been used widely in the past to study the neural processing of prosody, and that will be used in this thesis, can provide us with such information. These two approaches dominated research into how the brain processes prosody in the early days of the field, when there were no neuroimaging techniques available yet.

The first behavioral approach uses knowledge of the anatomy of the auditory system to inform us about whether one hemisphere is more capable at (or 'specialized in') processing prosody than the other cerebral hemisphere. Auditory information from the ears projects to both the same cerebral hemisphere as the side of acoustic input (the ipsilateral side) as to the hemisphere at the opposite side (the contralateral side). However, it can be shown that if we simultaneously stimulate both ears with highly competing different sources of auditory information (which is called 'dichotic' stimulation), it is primarily projected contralaterally (see Davidson & Hugdahl, 1995). If in such a 'Dichotic Listening (DL) task' a stimulus is presented to the ear contralateral to the specialized hemisphere, the specialized hemisphere can process it immediately. If the stimulus is presented to the ear contralateral to the unspecialized hemisphere, however, it either has to be processed sub-optimally by the unspecialized hemisphere or cross the corpus callosum to reach the specialized hemisphere, causing degradation of the stimulus representation or a time delay, respectively (Grimshaw, Kwasny, Covell, & Johnson, 2003). Therefore, if behavioral performance is systematically superior with stimulation to one of the ears, this is indicative of superiority ('hemispheric specialization') of the contralateral cerebral hemisphere at the perception task, e.g. perceiving prosody.

The second approach is the 'lesion-deficit' approach, which is one of the oldest research methods of neurology. It is grounded in the basic logic that if a certain neural structure is *necessary* for a certain mental capacity, damage to this structure should result in performance deficits in that mental domain as compared to controls without damage to the structure. Thus, in the case of prosody perception, comparing performance of patients with damage to structures hypothesized to be necessary for prosody perception to the performance of undamaged controls on the same prosody perception task, can inform us about what neural structures are indeed crucial for prosody perception.

1.2.2. Neurophysiological methods

An obvious approach to finding out what neural network is involved in prosody perception would be to directly observe brain activity while the brain is engaged in the processing of prosody. Indeed, with the development of neuroimaging techniques this has become possible and these methods have become standard tools of cognitive neuroscience. Two of such techniques will be used in the present thesis.

Brain cells communicate with each other through electrochemical signals. Although these electric potentials are very small (in the order of magnitude of micro-volts), with sufficient amplification the concerted electrical activity of groups of neurons can be measured at the scalp using electrodes (e.g. Luck, 2005). However, the brain is always active, so if we were to use this electroencephalography (EEG) technique while people are listening to prosody, how would we know what activity is related to prosody perception and what activity is not? By repeating a prosodic stimulus many times and averaging the electrical activity immediately after this 'event' over many such trials, the electrical activity not associated with the prosody perception process averages out and we are left with the averaged electrical brain wave that is associated with the perception of the prosodic stimulus – the Event Related Potential (ERP). The shape and distribution over the scalp of these ERPs can inform us about the nature of the prosody perception process.

Brain cells need energy to generate electrochemical signals. Neurons acquire energy by using oxygen to metabolize glucose into Adenosine Tri Phosphate (ATP). Thus, when neurons are active, the blood surrounding these neurons will have relatively low oxygen content, while blood surrounding less active neurons will have relatively high oxygen content. It can be shown that blood low in oxygen has slightly different magnetic properties than oxygen-rich blood. This magnetic consequence of brain activity is called the Blood Oxygenation Level Dependent (BOLD) response to brain activity. Magnetic Resonance Imaging (MRI) takes advantage of the BOLD response to infer brain activity. A participant is placed in a strong magnetic field, for instance while listening to prosody. Areas of the brain that are involved in prosody perception will consume more oxygen to generate activity than areas that are not involved in prosody perception, causing differences in distortion of the magnetic field depending on levels of brain activity. By measuring these small differences in the magnetic field, we can infer which areas were active when the participant was listening to prosody (e.g. see Jezzard, Matthews, & Smith, 2001). Because with this type of magnetic resonance imaging we can make inferences about brain function, it has also been called *functional* MRI (fMRI).

Note that these two neurophysiological methods can be seen as complementary to the lesion-deficit method to study the neural correlates of prosody perception. While the neurophysiological methods tell us what neural structures are *sufficient* for the perception of prosody, the lesion-deficit approach tells us which of these structures are *necessary* for prosody perception. Further, ERPs and fMRI can be seen as complementary too, as the first has high temporal (but low spatial) resolution while the latter has high spatial (but low temporal) resolution. In other words, while ERPs can inform us about *when* brain activity changes, fMRI can tell us *where* activity changes during prosody perception.

1.2.3. Quantitative meta-analysis

When the research for this thesis was initiated, there was already a large body of literature available that had used the lesion-deficit approach to study the neural network involved in prosody perception, but no clear conclusion had been reached. Therefore, it was decided that potentially more insight could be gained by re-analyzing and synthesizing the existing literature, rather than to set up yet another primary lesion study. Indeed, quantitative meta-analysis can provide new insights into existing literature by quantitatively combining the effects found in individual published studies (e.g. Hedges & Oklin, 1985). For instance, in the case of lesion-deficit studies, this would imply for every study computing the effect of damage to a neural structure on performance as compared to performance of subjects without brain damage. Subsequently, all the effect size estimates of individual studies are weighted by measurement precision and then combined into an overall effect size estimate. Through the combined sample size of all primary studies sampled, meta-analysis can test crucial hypotheses in a field with much higher statistical power than individual studies. Further, new insights can be gained by examining which factors - varying across studies - influence the overall effect size (which individual studies, by definition, cannot test).

Similarly, a significant body of fMRI literature into the network involved in emotional prosody perception had already accumulated when the research for this thesis was initiated. Thus, by quantitatively combining the results of these MRI studies, we could, at least in theory, infer with comparably high statistical power which neural network is involved in emotional prosody perception. A problem with fMRI results, however, is that they generally provide a set of locations of activation in the brain and no one obvious effect size that can be used for quantitative research synthesis. Therefore, in recent years a meta-analysis method has been developed that only uses the coordinates of activation of all individual studies to quantitatively synthesize fMRI research findings. At the core of this technique lies an algorithm that, based on all the activation coordinates provided by the primary studies, estimates for every location in the brain whether there is a better-than-chance association of that location with the mental operation (in our case prosody perception) studied. Such 'Activation Likelihood Estimation (ALE-) meta-analysis' (Laird et al., 2005) has become a mainstream tool in quantitatively synthesizing neuroimaging research.

1.2.4. Combined approaches to study automatic perception of emotional prosody

In the first section of this introduction we discussed the possibility that perception of emotional prosody (and perhaps emotional music) might be relatively automatic because of the existence of a hard-wired system dedicated to its perception. But how can we study whether automatic neural processing of emotional prosody perception really exists?

The example that we started with provides a clue as to how we could investigate automatic processing of emotional prosody. If we can show that a neural system continues to be engaged in the processing of emotional prosody processing even when attention is directed at something else (in the example we used, being absorbed by a good book), this is evidence that the processing of emotional prosody by that neural structure is automatic (Anderson, Christoff, Panitz, De Rosa, & Gabrieli, 2003). In this thesis we will similarly divert attention away from emotional prosody while we record brain activity using fMRI to see what structures in the brain might process emotional prosody automatically.

Recall that another aspect of 'automaticity' is that automatic processes are very fast (Moors & De Houwer, 2001). The so-called 'affective priming paradigm', developed by Fazio (2001) takes advantage of this attribute of automaticity to study whether an affective stimulus is being processed automatically. With this paradigm first positive (e.g. 'happy') or negative (e.g. 'sad') prosody is presented. Almost immediately - 200 milliseconds (ms) - later, a target word is presented that is either positive (e.g. HEALTH) or sad (e.g. PAIN) and participants are asked to decide as quickly as possible whether the word is positive or negative. If the emotional prosody is indeed processed rapidly (within 200 ms), it should be able to affect the speed of the response to the target words. If so, we expect responses to be slower for target words that are affectively incongruent with the primes (e.g., happy prosody – PAIN) than for congruent primes (e.g., sad prosody – PAIN). In this thesis, the affective priming paradigm will be used to study whether such rapid processing of emotional prosody (and perhaps of emotional music) can be demonstrated.

When the work for this thesis was initiated, there were two competing explanations for the nature of such rapid affective priming effects. On the one hand, it had been proposed that affective priming effects (APEs) occur because of rapid spreading of activation through a conceptual associative network from the affective concept activated by the affective prime to concepts with the same affective valence (Bargh, Chaiken, Raymond, & Hymes, 1996). Responses to affective targets would then be faster in the affectively congruent condition than incongruent condition because the target concept would already have been pre-activated by the affective prime in the congruent condition. On the other hand, it had been proposed that affective primes pre-activate a response tendency to press 'positive' or 'negative' – responses to targets in congruent conditions would then be faster than in incongruent conditions because the correct response would already have been pre-activated by the prime in the congruent condition, while in affectively incongruent conditions there would be interference by the pre-activated incorrect response in selecting the correct response, slowing down reaction time (De Houwer, Hermans, Rothermund, & Wentura, 2002). In Chapter 2.5 of this thesis it will be tested whether fast automatic affective priming can be demonstrated for emotional music and emotional prosody perception, and if so, which of these two mechanisms, 'the spreading of activation account' vs. 'the response competition account' can best explain APEs.

1.3. The cognitive neuroscience of prosody perception and its modulation by alexithymia

1.3.1. Interhemispheric neurocognitive models of prosody perception

In Figure 1, a schematic visualization can be found of the prosody perception pathway as hypothesized by two major models of (emotional) prosody perception (Schirmer & Kotz, 2006; Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009) that existed when the research work for this thesis was initiated. As will be explained in more detail in the next section, these models suggest that prosody perception is a multi-stage process with elementary acoustic processing first taking place in the primary auditory cortex (Heschl's gyrus, HG), after which there are further stages of acoustic processing with increasing complexity in the middle superior temporal gyrus (mid-STG) and subsequently in either the anterior STG (a-STG) or posterior STG (p-STG). Finally, abstract evaluation of (emotional) prosody would be sub-served by the inferior frontal gyrus (IFG).

Early work on the cognitive neuroscience of prosody perception focused on the question of whether the right hemisphere would be more adept to, or 'specialized' in, the processing of emotional prosody than the left hemisphere. It was hypothesized that the right hemisphere might be specialized in the processing of emotional prosody similarly to left hemispheric specialization for propositional language and that the classical Wernicke-Broca model of propositional language perception by the left hemisphere could be applied to right hemisphere processing of emotional prosody (Ross, 1981). In Figure 1, right-hemispheric specialization for emotional prosody perception is symbolized by the bold circles in the right hemisphere. Note that the hypothesis that the right hemisphere is specialized in the processing of emotional prosody is strictly 'interhemispheric' – it leaves unspecified at what stage of processing within the right hemisphere the purported right-hemispheric advantage arises.

The 'right hemisphere hypothesis of emotional prosody perception' (e.g., Ross, 1981) stimulated many lesion studies to investigate whether patients with right-hemispheric damage would show a more pronounced deficit in emotional prosody perception task performance than patients with equivalent damage to the left hemisphere. These studies were soon followed up by hypotheses about *why* the right hemisphere might be better at the processing of emotional prosody than the left. Broadly, two hypotheses were postulated. First, according to the '*functional* lateralization hypothesis', the cerebral hemispheres are specialized in the processing of *functional categories,* with the right hemisphere being specialized in processing of emotional information and the left in the processing of linguistic information (Van Lancker, 1980). In contrast, *acoustic* lateralization hypotheses proposed that specialization for prosody processing depends on the *acoustic cues* that are critical for the extraction of meaning. One prominent acoustic lateralization hypothesis stated that the left hemisphere is better at processing of temporal information while the right hemisphere is superior at spectral processing (e.g. Van Lancker & Sidtis, 1992). For instance, as variation in pitch is an important acoustic cue to the meaning of emotional prosody, right-hemispheric superiority for emotional prosody processing could then be explained on the basis of right hemispheric specialization for pitch processing. Note that, if cue-dependent lateralization hypotheses were correct, we would expect hemispheric asymmetries early on in the prosody perception pathway (Figure 1) in areas devoted to acoustic processing. If, on the other hand, hemispheric specialization is driven by specialization for more abstract (functional) categories we would expect hemispheric asymmetries to emerge later in the prosody perception pathway, when a more abstract level of processing has been reached.

In the first three empirical chapters of the first empirical section of this thesis, lesion-deficit data (Chapter 2.1), dichotic listening data combined with ERPs (Chapter 2.2) and neuroimaging data (Chapter 2.3) will be examined to test whether hemispheric specialization for prosody perception can be demonstrated and if so, which of the above hypotheses can best explain hemispheric specialization for prosody perception.

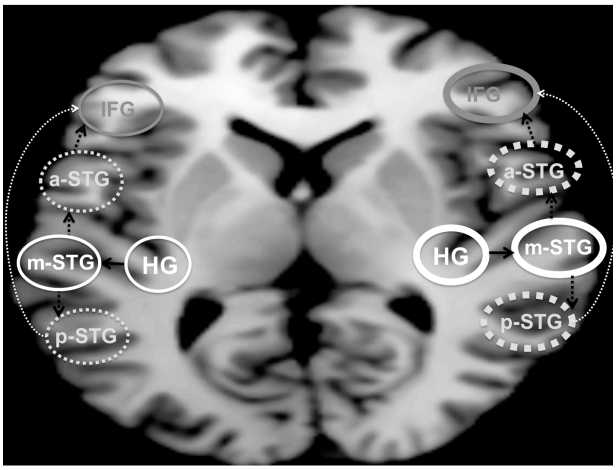

Figure 1. Schematic visualization of the prosody perception pathway as hypothesized by two dominant models of (emotional) prosody perception, superimposed on an axial slice of an MRI scan. Solid circles and arrows indicate regions and connections that are hypothesized to be important in (emotional) prosody perception by both models. Dashed circles and arrows indicate areas and connections hypothesized to be involved in prosody perception by only one of the two models. White circles indicate areas hypothesized to be involved in the first stage, light grey circles in the second stage and dark grey circles in the final stage of (emotional) prosody perception. Bold circles in the right hemisphere as compared to the left indicate hypothesized right-hemispheric superiority for emotional prosody perception. Abbreviations: HG = Heschl's gyrus, m-STG = middle superior temporal gyrus, p-STG = posterior superior temporal gyrus, a-STG = anterior superior temporal gyrus, IFG = inferior frontal gyrus.

1.3.2. Intrahemispheric neurocognitive models of prosody perception

With the development of neuroimaging methods it became possible to further probe the human brain to reveal which areas within the cerebral hemispheres are involved in prosody perception. These neuroimaging studies inspired two similar three-stage (emotional) prosody perception models, developed by two leading German research groups (Schirmer & Kotz, 2006; Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009). The neural network of emotional prosody perception that these two models proposed, and the disagreement between the two models is illustrated in Figure 1.

When auditory information reaches the cerebral cortex, first elementary acoustic processing takes place in the primary auditory cortex (PAC), which is located in the transverse temporal gyrus or Heschl's gyrus (HG). According to both the model of the Kotz group (Schirmer & Kotz, 2006) and the Wildgruber group (Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009) in an initial stage (white circles in Figure 1) basic acoustic features important for (emotional) prosody perception are extracted in HG and the middle superior temporal gyrus (mid-STG). Both models propose a subsequent second stage (light-grey circles in Figure 1) of (emotional) prosody perception where the acoustic features extracted in stage one are integrated into a 'gestalt' (more abstract units of representation), but the models disagree on the brain area which is responsible for this stage of processing. While the Wildgruber model proposes that stage-two emotional prosody perception takes place in the posterior STG (p-STG), the Kotz model places this stage further anterior along the auditory 'what pathway' (auditory object recognition stream) in the anterior STG (a-STG). Last, in a final third stage (dark-grey circles in Figure 1) both models propose that abstract evaluation of prosodic information and integration of prosodic information with other layers of the speech signal (such as semantics) takes place in the inferior frontal gyrus (IFG). In Chapter 2.3 of this thesis, the neuroimaging literature of emotional prosody perception that was available when the work for this thesis was initiated will be meta-analyzed, to investigate which of these two models is best supported by the neuroimaging evidence base.

In addition to the cortical areas mentioned above, subcortical structures have been implicated in emotional prosody perception. According to the Kotz model, the amygdala plays a role in the automatic perception of emotional prosody. Indeed, the amygdala has been implied in fast (and hence possibly 'automatic') emotional processing in the classic dual route model by LeDoux (2001) and Öhman and Mineka (2001) have subsequently proposed that it might be the amygdala that has evolved as the neural substrate dedicated to the (automatic) processing of potentially harmful social signals (such as anger prosody). The two approaches to studying automaticity of processing described in section 1.2.4 will therefore be applied in this thesis to test whether two key features of 'automaticity' can be demonstrated for the perception of emotional prosody (and perhaps emotional music): (1) that the perception process occurs even when participants do not intend to analyze the emotional signal and (2) that the perception process is fast. First, in Chapter 2.4 it will be tested using fMRI whether the emotional prosody perception process continues to operate at the neural level even when subject do not intend to process emotional prosody. Furthermore, it will be examined whether the amygdala might indeed be the neural substrate of unintentional emotional prosody perception. Second, in Chapter 2.5 it will be tested using the affective priming paradigm and EEG whether rapid perception of emotional prosody and emotional music can be demonstrated. Additionally, it will be tested which of the two proposed fast mechanisms described in section 1.2.4, spreading of activation in a conceptual network of affective concepts vs. response competition, can best explain such affective priming effects.

1.3.3. Modulation by alexithymia

As was explained in section 1.1 of this introduction, *alexithymia* is a personality trait that is associated with difficulty in verbalizing and identifying emotions. Therefore, it seems plausible to assume that at some level in the emotional prosody perception pathway, individual differences in this personality trait should be reflected. When the work for this thesis was initiated, the few neuroimaging studies that had been performed pointed to a decreased response of both subcortical (Kugel et al., 2008) and cortical areas (Kano et al., 2003, but see Berthoz et al., 2002 for mixed evidence) to emotional stimulation for people scoring high on this personality trait. Therefore, there is discussion about whether alexithymia is primarily reflected by subcortical (and possibly 'automatic' perception of emotion) or cortical mechanisms (perhaps reflecting more abstract and deliberate interpretation of emotional information). Functional connectivity studies had proposed an alternative explanation: some evidence had been found for altered coupling of cortical and limbic areas during emotional processing in alexithymics, possibly explaining the reduced ability of alexithymics to reflectively process emotional information by disturbed flow of information between subcortical and cortical areas (Mériau et al., 2006). However, all these studies were based on emotional processing through the visual system, rendering it uncertain whether the above hypotheses would also hold for emotional perception through the auditory system. Thus, in Chapter 3.1 it will be tested whether the purported automatic (rapid) perception of emotional prosody and emotional music as measured with the affective priming paradigm is modulated by (non-clinical) alexithymia. Last, in Chapter 3.2 it will be examined whether (non-clinical) alexithymia modulates automatic (unintentional) perception of emotional prosody and whether such modulation is mediated by relatively early (subcortical) or late (cortical) mechanisms.

The following references correspond to chapters in this thesis:

Chapter 2.1. Witteman, J., Van IJzendoorn, M. H., Van de Velde, D., Van Heuven, V. J., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. *Neuropsychologia, 49,* 3722-3738.

Chapter 2.2. Witteman, J., Goerlich, K. S., Martens, S., Aleman, A., Van Heuven, V. J., & Schiller, N. O. (2014). The nature of hemispheric specialization for prosody perception. *Cognitive Affective & Behavioral Neuroscience, 14,* 1104-1114.

Chapter 2.3. Witteman, J., Van Heuven, V. J., & Schiller, N. O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia, 50,* 2752-2763.

Chapter 2.4. Witteman, J., Veer, I. M., Van Steenbergen, H., Rombouts, S. A., Van Heuven, V. J., & Schiller, N. O. (Submitted). Selective automatic activation of the superior temporal gyrus for the perception of angry prosody.

Chapter 2.5. Goerlich, K. S., Witteman, J., Schiller, N. O., Van Heuven, V. J., Aleman, A., & Martens, S. (2012). The nature of affective priming in music and speech. *Journal of Cognitive Neuroscience, 24,* 1725-1741.

Chapter 3.1. Goerlich, K. S., Witteman, J., Aleman, A., & Martens, S. (2011). Hearing feelings: Affective categorization of music and speech in alexithymia, an ERP study. *PloS One, 6,* e19501.

Chapter 3.2. Goerlich, K. S., Witteman, J., Schiller, N. O., Van Heuven, V. J., Aleman, A., & Martens, S. Blunted feelings: Alexithymia is associated with a diminished neural response to speech prosody. *Social Affective and Cognitive Neuroscience, 9,* 1108-1117.

2. References

Altenmüller, E., Kopiez, R., & Grewe, O. (2013). A contribution to the evolutionary basis of music: Lessons from the chill response. In A. Altenmüller, E., Schmidt, S., & Zimmermann, E. (Eds.), *Evolution of Emotional Communication: From sounds in Nonhuman Mammals to Speech and Music in Man* (pp. 313-335). Oxford: Oxford University Press.

Anderson, A. K., Christoff, K., Panitz, D., De Rosa, E., & Gabrieli, J. D. E. (2003). Neural correlates of the automatic processing of threat facial signals. *The Journal of Neuroscience, 23*, 5627-5633.

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

Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditionally automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology, 32,* 104-128.

Berthoz, S., Artiges, E., Van de Moortele, P. F., Poline, J. B., Rouquette, S., Consoli, S. M., & Martinot, J. L. (2002). Effect of impaired recognition and expression of emotions on frontocingulate cortices: An fMRI study of alexithymia. *American Journal of Psychiatry, 159,* 961-967.

Canli, T. (2004). Functional brain mapping of extraversion and neuroticism: Learning from individual differences in emotion processing. *Journal of Personality, 72*, 1105-1132.

Davidson, R. J., & Hugdahl, E. K. (Eds.). (1995). *Brain Asymmetry.* Cambridge, MA: MIT Press.

De Houwer, J., Hermans, D., Rothermund, K., & Wentura, D. (2002). Affective priming of semantic categorisation responses. *Cognition and Emotion, 16,* 643-666.

Fazio, R. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, *15*, 115-141.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I. Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal recognition of three basic emotions in music. *Current Biology, 19,* 573-576.

Grimshaw, G. M., Kwasny, K. M., Covell, E., & Johnson, R. A. (2003). The dynamic nature of language lateralization: Effects of lexical and prosodic factors. *Neuropsychologia, 41,* 1008-1019.

Hardcastle, W. J., Laver, J., & Gibbon, F. B. (Eds.). (2010). *The handbook of phonetic sciences* (2nd ed.) Oxford: Wiley-Blackwell.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

Jezzard, P., Matthews, P. M., & Smith, S. M. (Eds). (2001). *Functional MRI: An introduction to the methods.* Oxford: Oxford University Press.

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31,* 575-621.

Kano, M., Fukudo, S., Gyoba, J., Kamachi, M., Tagawa, M., Mochizuki, H., Itoh, M., Hongo, M., & Yanai, K. (2003). Specific brain processing of facial expressions in people with alexithymia: An $H_2{}^{15}O$-PET study. *Brain, 126,* 1474-1484.

Kugel, H., Eichman, M., Dannlowski, U., Ohrman, P., Bauer, J., Arolt, V., Heindel, W., & Suslow, T. (2008). Alexithymic features and automatic amygdala reactivity to facial emotion. *Neuroscience Letters, 435,* 40-44.

Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics, 25,* 313-342.

Laird, A. R., Fox, P. M., Price, C. J., Glahn, D. C., Uecker, A. M., Lancaster, J. L., . . . Fox, P. T. (2005). ALE meta-analysis: Controlling the false discovery rate and performing statistical contrasts. *Human Brain Mapping, 25,* 155-164.

LeDoux, J. E. (2000). Emotion Circuits in the Brain. *Annual Review of Neuroscience, 23,* 155-184.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge, MA: MIT press.

Luck, S. J. (Ed.). (2005). *An introduction to the event related potential technique.* Cambridge, MA: MIT press.

Mériau, K., Wartenburger, I., Kazzer, P., Prehn, K., Lammers, C. H., Van der Meer, E., Villringer, A., & Heekeren, H. R. (2006). A neural network reflecting individual differences in cognitive processing of emotions during perceptual decision making. *NeuroImage, 33,* 1016-1027.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132,* 297-326.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*, 483-522.

Peretz, I., Aubé, W., & Armorny, J. L. (2013). Towards a neurobiology of musical emotions. In A. Altenmüller, S. Scmidt, & E. Zimmermann (Eds.), *Evolution of Emotional Communication: From sounds in Nonhuman Mammals to Speech and Music in Man* (pp. 313-335). Oxford: Oxford University Press.

Rietveld, A. C. M., & Van Heuven, V. J. (2009). *Algemene fonetiek* [General Phonetics]. Bussum: Coutinho.

Ross, E. D. (1981). The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology, 38,* 561-569.

Sifneos, P. E. (1973). The prevalence of 'alexithymic' characteristics in psychosomatic patients. *Psychotherapy & Psychosomatics, 22*, 255–262.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32,* 76-92.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10,* 24-30.

Silk, J. B. (2007). Social components of fitness in primate groups. *Science, 317,* 1347-1351.

Van Bezooijen R., Otto, S. A., & Heenan, T. A. (1983). Recognition of vocal expressions of emotion. *Journal of Cross-Cultural Psychology, 14*, 387-406.

Van Heuven, V. J., & Kirsner, R. (2004). Phonetic or phonological contrasts in Dutch boundary tones? *Linguistics in the Netherlands, 21,* 102-113.

Van Heuven, V. J., & Sluijter, A. (1996). Notes on the phonetics of word prosody. In: Goedemans, R., Van der Hulst, E., & Visch, E. (Eds.). *Stress patterns in the world, part 1: Background* (233-269). The Hague: Holland Institute of Generative Linguistics, Leiden/Holland Academic Graphics.

Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Papers in Linguistics: International Journal of Human Communication, 13,* 201-277.

Van Lancker, D., & Sidtis, J. J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research, 35,* 963-970.

Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology, 11,* 277-281.

FIRST EMPIRICAL PART: TOWARDS A COGNITIVE NEUROSCIENCE OF PROSODY PERCEPTION

Chapter 2.1

Hemispheric specialization for prosody perception: A meta-analysis of the lesion literature

Witteman, J., Van IJzendoorn, M. H., Van de Velde, D., Van Heuven, V. J., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. *Neuropsychologia, 49,* 3722-3738.

Abstract

It is unclear whether there is hemispheric specialization for prosodic perception and, if so, what the nature of this hemispheric asymmetry is. Using the lesion-approach, many studies have attempted to test whether there is hemispheric specialization for emotional and linguistic prosodic perception by examining the impact of left- versus right hemispheric damage on prosodic perception task performance. However, so far no consensus has been reached. In an attempt to find a consistent pattern of lateralization for prosodic perception, a meta-analysis was performed on 38 lesion studies (including 450 left hemisphere damaged patients, 534 right hemisphere damaged patients and 491 controls) of prosodic perception. It was found that both left- and right hemispheric damage compromise emotional and linguistic prosodic perception task performance. Furthermore, right hemispheric damage degraded emotional prosodic perception more than left hemispheric damage (trimmed g = −0.37, 95% CI [−0.66; −0.09], N = 620 patients). It is concluded that prosodic perception is under bihemispheric control with relative specialization of the right hemisphere for emotional prosodic perception.

1. Introduction

How we say something can be as important as *what* we say when conveying a message to our audience. This prosodic (supra-segmental) layer of speech uses a variety of acoustic cues such as speaking rate, pitch and intensity to convey different communicative functions. On the one hand, prosody can be used to convey information regarding the linguistic structure of an utterance (for a review of linguistic prosodic functions see Cutler, Dahan, & Van Donselaar, 1997). This 'linguistic prosody' can be used to stress syllables, group words into intonational phrases, emphasize importance of constituents in a sentence and to signal whether an utterance is meant as a question or a statement. On the other hand, prosody can be used to convey paralinguistic information such as the emotional state of the speaker (for a review see Scherer, 1986), which henceforth will be referred to as 'emotional prosody'.

Over the last four decades a considerable body of literature has accumulated on the question how the brain processes prosody (for recent reviews see Wong, 2002; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006; Kotz, Meyer, & Paulmann, 2006; Schirmer & Kotz, 2006; Alves, Fukusima, & Aznar-Casanova, 2008; Kotz & Paulmann, 2011). Understanding how prosody is processed in the brain is not only interesting from a fundamental cognitive neuroscience point of view but could also be clinically relevant as, for instance, impairment of prosodic processing has recently been found to be a core deficit in schizophrenia (Hoekert, Kahn, Pijnenborg, & Aleman, 2007). A central question that has remained unresolved so far is whether there is hemispheric specialization for prosodic perception, and if so, which mechanism drives this hemispheric asymmetry.

Concerning lateralization of prosodic perception on the cortical level, four hypotheses have emerged:

(1) The right cerebral hemisphere is specialized in the processing of *all* prosodic information (Klouda, Robin, Graff-Radford, & Cooper, 1988);

(2) The *Right hemisphere hypothesis* posits that the right hemisphere is specialized in emotional prosodic processing (Ross, 1981; Blonder, Bowers, & Heilman, 1991; Borod et al., 1998);

(3) The *Functional lateralization hypothesis* (Van Lancker, 1980) proposes that hemispheric specialization is dependent on the communicative function of prosodic material: emotional prosodic information is processed in the right hemisphere while linguistic prosody is processed in the left;

(4) The *Cue dependent lateralization hypothesis* proposes that lateralization of prosodic processing depends on the acoustic cues that are critical for the extraction of meaning: the left hemisphere would be better adapted to processing of durational information while the right hemisphere is superior in spectral processing (Van Lancker & Sidtis, 1992). As variation in pitch is an important acoustic cue to the meaning of emotional prosody (but not the only cue; see Scherer, 2003), right hemispheric superiority for emotional prosodic processing could then be explained on the basis of rightward lateralization for pitch processing.

On the one hand, the first three hypotheses all assume a specialized (and lateralized) module for structuring of incoming acoustic information into *prosodic categories*. Note that these three "categorical" hypotheses need not necessarily be mutually exclusive; for instance, the functional lateralization hypothesis can be seen as a refinement of the right hemisphere hypothesis. On the other hand, cue-dependent hypotheses posit that lateralization of prosodic processing is determined by non-prosody specific acoustic processes. The cue-dependent and the "categorical" lateralization hypotheses are not mutually exclusive either: they could represent different stages of prosodic processing which might be differentially lateralized (see Schirmer & Kotz, 2006).

One way to test these hypotheses is through the lesion approach. Typically, lesion studies compare a group of patients with acquired left or right hemispheric brain damage to a group of healthy controls on a prosodic perception task. If there is hemispheric specialization for prosodic processing then damage to the specialized hemisphere should (1) compromise performance on prosodic tasks as compared to controls (and equivalent damage to the non-specialized hemisphere should degrade performance relative to controls less) and (2) deteriorate performance as compared to equivalent damage to the non-specialized hemisphere. This approach provides information about which hemisphere of the brain is *necessary* for prosodic perception.

In order to differentiate between the "categorical" lateralization hypotheses it is necessary for lesion studies to have (1) a right hemispheric damage (RHD) group, a left hemispheric damage (LHD) group and a normal control (NC) group as well as (2) an emotional and a linguistic prosodic perception task. Unfortunately most lesion studies that have been published to date do not fulfill these criteria. The studies that did fulfill these criteria give an inconsistent picture with some presenting evidence favoring global right-hemisphere superiority for prosodic processing (Blonder et al., 1991; Borod et al., 1998) the right hemisphere hypothesis (Heilman, Bowers, Speedie, & Coslett, 1984) and the functional lateralization hypothesis (Walker, Daigle, & Buzzard, 2002) while others do not support any of the hypotheses (Pell & Baum, 1997; Pell, 1998; Breitenstein, Daum, & Ackermann, 1998; Geigenberger & Ziegler, 2001; Zgaljardic, Borod, & Sliwinski, 2002; Kho et al., 2007). When these studies and studies that included all relevant groups but imposed only one prosodic task find detrimental effects of hemispheric damage, many find that damage to *each* of the two hemispheres compromises emotional (Heilman et al., 1984; Van Lancker et al., 1992; Lalande, Braun, Charlevois, & Whitaker, 1992; Peper & Irle, 1997; Pell, 1998; Breitenstein, 1998; Zaidel, Kasher, Soroker, & Batori, 2002; Kucharska-Pietura, Phillips, Gernand, & David, 2003; Shamay-Tsoory, Tomer, Goldsher, Berger, & Aharon-Peretz, 2004; Pell, 2006; Kho et al., 2007, but for evidence of hemisphere specific degradation see Tompkins & Flowers, 1985; Bowers, Coslet, Bauer, Speedie, & Heilman, 1987; Blonder et al., 1991; Geigenberger et al., 2001; Walker et al., 2002; Charbonneau, Scherzer, Aspirot, & Cohen, 2003) and linguistic (Heilman et al., 1984; Pell et al., 1997; Baum, 1998; Geigenberger et al., 2001; Aasland & Baum, 2003; Seddoh, 2006b, but for hemisphere specific degradation see Bryan, 1989; Blonder et al., 1991; Perkins, Baran, & Gandour, 1996; Pell, 1998; Borod et al., 1998; Walker, Fongemie, & Daigle, 2001; Walker et al., 2002; Abada & Baum, 2006) prosodic perception performance, suggesting that both hemispheres provide necessary contributions to both prosodic functions.

To disentangle the contribution of "categorical" versus cue-dependent hemispheric specialization in prosodic perception it is necessary to vary the function of the prosodic material while keeping acoustics constant or vice versa and observe whether there is differential impact of left- versus right-hemispheric damage on prosodic perception performance as compared to performance by NC. One approach has been to selectively remove durational or fundamental frequency (F_0) variation in linguistic or emotional prosodic stimuli and to observe whether LHD or RHD differentially degrades perception performance as compared NC. Unfortunately these studies (Pell, 1998; Baum, 1998; Aasland & Baum, 2003) have not consistently found differential degradation of performance after removal of F_0 variation for LHD and after removal of durational information for RHD, as would have been expected based on the Cue-dependent lateralization hypothesis (Van Lancker et al., 1992). Adopting a different approach, Van Lancker et al. (1992) used discriminant analysis to analyze which acoustic properties of emotional prosody could predict the pattern of errors made by LHD and RHD patients on an emotional categorization task. It was shown that the errors of the RHD patients could be predicted by misuse of F_0 variability. The authors concluded that the right hemisphere might contribute to emotional prosodic perception through a specialization in pitch processing. However, this conclusion must be considered with caution as Baum and Pell (1997) failed to replicate the result.

Several factors have been suggested in the literature that can moderate the impact of lateralized brain damage on prosodic perception performance. Ross, Thompson, and Yenkosky (1997) propose that apparent emotional prosodic processing deficits after LHD are not caused by emotional prosodic processing deficits per se, but that these patients have problems linking emotional meaning from the prosodic layer to the propositional layer of the speech signal. These authors predict that when the 'verbal-articulatory demands' (whether lexical meaning and syllables are present) of an (affective) prosodic perception task are increased LHD performance should degrade while RHD performance should remain unaffected. Secondly, as was already evident in our discussion of the non-mutual exclusivity of the cue-dependent versus "categorical" hypotheses of prosodic perception, prosodic processing can be conceptualized as a *process* consisting of several stages. For instance, in a recent review Schirmer and Kotz (2006) propose that there are at least three stages in prosodic perception (see also Kotz et al., 2006). In an initial stage, complex acoustical analysis of the speech signal is performed; in the second stage, emotional or linguistic information is identified; and in a final stage, this information becomes available to higher-order cognitive processes for further evaluation or integration with other layers of speech (such as the propositional content). This proposal implies that performance for prosodic perception tasks such as those used in the lesion literature reflects a combination (i.e. summation or even interaction) of these stages, each of which might be differentially lateralized (Gandour, 2004). Lastly, as Hoekert et al. (2007) have pointed out in a meta-analytic review of emotional prosodic impairment in schizophrenia, the quality of the prosodic perception task used might influence the findings. Tasks with high psychometric quality can be expected to give a better picture of prosodic performance degradation due to lateralized brain damage than tasks of low psychometric quality.

In sum, although a considerable body of lesion literature has accumulated, no consensus has been reached on the degree and nature of hemispheric lateralization for prosodic perception. Most studies that had the appropriate design for differentiating between the "categorical" lateralization hypotheses of prosodic perception find that both LHD and RHD can affect prosodic perception, suggesting that hemispheric specialization for prosodic perception is a matter of *degree* rather than type. The small number of studies that have manipulated acoustic cues in prosodic perception tasks have not consistently supported the cue-dependent lateralization hypothesis. The most important problems in testing the "categorical" lateralization hypotheses mentioned is that many of the lesion studies published to date do not have the appropriate design to disentangle the various lateralization hypotheses (Wong, 2002). Furthermore, studies typically had fewer than 15 subjects per experimental group, limiting statistical power to detect effects.

In the present study, a meta-analysis was employed to review the lesion literature on hemispheric specialization for prosodic perception. By (1) including RHD, LHD and NC groups as well as (2) both emotional and linguistic perception tasks in the meta-analysis, it was possible to overcome the main weaknesses of previous individual studies and meta-analytically differentiate between the "categorical" lateralization hypotheses of prosodic perception (the number of studies manipulating acoustics while keeping prosodic function constant or vice versa was too low to contrast the "categorical" hypotheses with the cue-dependent lateralization hypothesis). For instance, although individual studies that have only included a NC group and a LHD or RHD group and one prosodic function (e.g. Baum, Daniloff, Daniloff, & Lewis, 1982; Wertz, Henschel, Auther, Ashford, & Kirshner, 1998; Seddoh, 2006; Weintraub, Mesulam, & Kramer, 1981; Harciarek, Heilman, & Jodzio, 2006) cannot give any definitive information on lateralization of prosodic processing (i.e. whether one of the hemispheres is *specifically* necessary for emotional or linguistic prosodic perception or both), such studies are valuable for discriminating between the "categorical" lateralization hypotheses when compared meta-analytically. Furthermore, by including a large number of subjects for each experimental group in the meta-analysis, the lateralization hypotheses of prosodic perception could be tested with high statistical power, making it possible to demonstrate subtle effects that individual studies cannot detect. By summarizing studies quantitatively a more precise and objective insight in the effect of lateralized brain damage on prosodic perception task performance can be gained as compared to the traditional qualitative reviews. Additionally, we tested the influence of factors that have been suggested to moderate the relationship between lateralized brain damage and prosodic perception performance by moderator-analysis. Moderating variables (such as whether patients and controls are matched on demographic variables) typically vary only between studies. Therefore, while most individual studies cannot test for such effects, meta-analytic moderator analysis can provide novel insights by assessing the impact of moderators. Finally, since the last qualitative review of the lesion literature on prosodic processing was published in 2002 (Wong, 2002), the current review additionally covers almost a decade of relevant research.

2. Methods

2.1. Literature search

The PubMed and PsycLit databases were searched for relevant articles published until January 2011 using the search string '*prosod** AND *brain* NOT *EEG* NOT *ERP* NOT *MRI*' (where the asterisk denotes a wildcard) in the title or abstract. Additionally, the reference lists of all articles included in the meta-analysis and available reviews (Wong, 2002; Kotz et al., 2006, 2011) were manually checked for previously published potentially relevant articles. This search yielded 80 publications that were considered for inclusion in the meta-analysis.

2.2. Study selection

Studies needed to fulfil the following set of criteria to be eligible for inclusion in the meta-analyses. First, (i) at least an adult group with acquired left hemispheric damage (LHD) and a group with right hemispheric damage (RHD) or (ii) at least one of these brain damaged groups and a normal control (NC) group had to be present. Second, objective (CT/MR imaging or surgical) evidence for the lateralized nature of the damage had to be presented for the majority of the patients. Third, the brain damage had to be primarily cortical and focal: lesions had to be clearly localizable and situated primarily in the cerebral cortex which implied that most lesions were caused by cerebrovascular damage or surgical intervention (such as tumor extirpation or resection for intractable epilepsy). We excluded non-focal (diffuse) etiology such as diffuse traumatic brain injury or Parkinson and Huntington pathology. As part of our general strategy to include as many lesion studies as possible in order to maximize the scope of our results and test for potentially moderating effects factors such as etiology rather than exclude studies, no further restrictions were imposed on etiology of the brain damage. However, as can be seen in Tables 1 and 2, the vast majority of included studies presented patients with brain damage of vascular origin. Fourth, at least one explicit linguistic or emotional prosodic perception task had to be reported and performance needed to be qualitatively interpretable (i.e. we excluded rating tasks that do not allow for the interpretation of test scores in terms of good/bad). Emotional prosodic perception tasks typically present a set of prerecorded utterances that are intoned in a variety of emotional categories by an actor; patients typically have to identify (or rate the intensity of) the intended emotion or discriminate between the emotional intonation of utterance pairs. Linguistic prosodic perception tasks also typically demand subjects to identify or discriminate between prosodic categories, but in this case prosody imparts linguistically relevant (and hence categorical) information such as sentence modality, sentential stress, syntactic structure through phrase marking or metrical stress. One of these studies (Grosjean, 1996) used an atypical task: the ability of subjects to predict from sentence prosody whether a sentence continues or stops at a certain target word. We considered this task to tap linguistic prosodic processing because it is a measure of the ability to perceive phrase structure on the basis of prosodic features (which can be considered a linguistic prosodic function, e.g. for a review see Cutler, Dahan, & Van Donselaar, 1997). Fifth, the study had to report original material (e.g., we excluded Ross & Monnot (2008), who reanalyzed data from their research database) in order to prevent data from entering a meta-analysis twice (which violates the assumption of independence) and had to be published in an English language international peer-reviewed journal. Implicit prosodic perception tasks (e.g., Wunderlich, Ziegler, & Geigenberger, 2003) were not included as they might engage different neural systems (for a review of this issue see Wildgruber et al., 2006) and by definition tap other processes in addition to prosodic processing (i.e. the patient is actively engaged in a different task than evaluating the prosodic information). Furthermore, there were not enough studies that reported implicit tasks to look at the moderating effect of the explicit versus implicit nature of the task. Sixth, we excluded studies on tonal languages as the literature suggests that these might have a different lateralization pattern than non-tonal languages (e.g., see Gandour et al., 2003). Finally, the study had to report sufficient information to be able to compute or accurately estimate the standardized difference in means. This required the study to report (i) sample size and means and standard deviations or (ii) sample size and test statistics such as *Z*- or *T*- or *F*-values with means or (iii) sample size and exact or categorical *p*-values.

2.3. Data analysis strategy

We tested the effect of RHD versus LHD and of lateralized damage versus NC performance for both emotional and linguistic prosodic tasks. For meta-analysis to be statistically valid, it is necessary that each study and each subject only contributes to the analysis once. This assumption of independence required six separate meta-analyses: three comparisons (NC vs. LHD; NC vs. RHD; LHD vs. RHD) for each of the two prosodic functions (emotional vs. linguistic).

Additionally, we wanted to assess differences between comparisons in the mean weighted effect size (ES). For instance, we asked whether damage to the specialized hemisphere compromises performance (as compared to NC) *more* than damage to the non-specialized hemisphere, or whether damage to a hemisphere *differentially* disturbs performance for linguistic or emotional prosodic processing. Therefore, the 85% confidence interval (CI) around the mean weighted ES under the random effects model was computed for each meta-analysis. Non-overlapping 85% CIs of two mean weighted effect sizes indicate a significant difference at the traditional 5% type I error threshold (Goldstein & Healy, 1995). This approach allowed us to assess whether there were differences in the mean weighted ES between meta-analyses and to present a graphical overview of the analyses.

Many studies reported multiple (and hence non-independent) measures of prosodic processing. To preserve independence we pooled multiple effect sizes to obtain one measure of ES per study for each of the six comparisons. This strategy had the following implications. First, we pooled across multiple measures of a prosodic function. However, if a study reported unimodal (only prosodic) and multimodal prosodic tasks (for instance, tasks in which emotional semantics can be congruent or incongruent with the emotional meaning of the prosody), only the unimodal task was used (as the former is a purer measure of *prosodic* processing). Second, if data were presented for subgroups with damage to different locations within a hemisphere we pooled to a mean ES for the whole hemisphere (since *inter*hemispheric differences are the focus of the meta-analysis). Third, if pre- and post-surgical data (tumor extirpation, resection of epileptic foci) were reported, we only analyzed the post-surgery data (preserving the overall logic of analyzing the effect of acquired brain damage on prosodic processing between groups).

2.4. Computation of effect size and meta-analytic procedures

Hedges's measure of effect size *g* (the standardized difference in means) was calculated for each comparison. Hedges's *g* is very similar to Cohen's *d* (the difference between means divided by the pooled standard deviation) but is less biased if the sample size is small (Hedges & Olkin, 1985). When comparing patient groups to NC we gave the ES a negative sign if patients performed worse than NC to indicate deterioration of performance due to brain damage. When comparing the LHD to RHD group we gave the ES a minus sign if RHD performed worse than LHD since most hypotheses of hemispheric specialization for prosodic processing predict performance degradation for the RHD group relative to LHD.

The included studies varied widely in the quality of the reported statistics. When possible, ES was calculated from the sample size and the (pooled) means and standard deviations. If the standard deviation was zero we entered the value $1 \cdot 10^{-10}$ because the Comprehensive Meta Analysis package (Borenstein, hedges, Higgins, & Rothstein, 2005) does not accept standard deviations of zero. If the study reported sample size and either a t-test or sample size and a *p*-value, we used Comprehensive Meta Analysis (CMA) to convert to Hedges's *g*. When (one way) *F*-tests were reported, we used the "Effect size determination program" developed by Wilson (http://mason.gmu.edu/~dwilsonb/downloads/ES_Calculator.xls) to convert to *g*. When categorical *p*-values were reported, we entered the upper border in CMA (i.e. $p = 0.05$ when $p < 0.05$ was reported), and when 'no effect' was reported, we entered an ES of 0, adopting a conservative approach.

Two additional sources of information were used to estimate *g*. When the main effect *F*-value of a multiple way ANOVA was reported, we used the 'Effect size determination program' to convert it to *g*. This procedure potentially overestimates the true ES as the error term of the *F*-test will be smaller due to the inclusion of an additional factor (apart from the factor of interest: lateralized brain damage versus control). We also entered *Z*-values (performance scores that have been normalized using the mean and standard deviation of the control group) as *g* in the meta-analysis for comparisons of patient groups versus controls. These measures of ES may also overestimate ES since it is not the pooled standard deviation but the (likely smaller) standard deviation of the control group that is used in the denominator. We later tested in a moderator analysis whether these last two sources of information gave a significantly larger ES than the other measures of ES but this test failed to reach significance. Therefore, we decided to include these measures of ES in the meta-analyses in order to increase the representativeness of the results.

For each meta-analysis, the distribution of effect sizes was checked for outliers by first converting all *g*'s to Fisher's *Z* and subsequently converting to standard normal scores. Outliers were defined as values outside the −3.29 to 3.29 range (corresponding to a probability that is lower than 0.001 in the normal distribution). As none of the study outcomes fulfilled this criterion all the included outcomes remained in the meta-analyses.

All further meta-analytic procedures were performed with Comprehensive Meta-Analysis (Borenstein et al., 2005). The mean weighted ES *g* and 95% CI were computed using the random effects model. The random effects model was chosen because it is reasonable to assume that the true ES varied among included studies (see Borenstein, Hedges, Higgins, & Rothstein, 2009) and it produced more conservative results. For each meta-analysis the mean weighted effect size and its confidence interval was computed using the inverse variance weight of each study under the random effects model, ensuring that the measurement precision of contributing studies was taken into account when computing the summary statistics. Subsequently, the *Q*-statistic was computed to test whether there was significant heterogeneity in the ES distribution (Lipsey & Wilson, 2001). Follow-up moderator analyses were performed to explore whether theoretically or methodologically plausible factors (see below) could explain variance in the ES distribution.

Seven studies fulfilled all inclusion criteria but failed to report sufficient information to calculate *g*. In order to give a complete overview of the issue at hand these studies were analyzed through narrative review.

2.5. Publication bias

A threat to the validity of meta-analysis is the so-called 'file drawer problem' (Rosenthal, 1979). This refers to the phenomenon that studies that find statistically significant differences are published more easily than studies failing to reject the null-hypothesis (they remain in the file drawer). These statistically non-significant studies have a lower probability of being included in the meta-analysis than studies that do find statistically significant differences resulting in the risk of overestimation of the true ES by meta-analysis. This issue can be examined by plotting the ES of each study against the precision (defined as the inverse of the standard error) in a so-called 'funnel-plot' (Lipsey & Wilson, 2001). If there is no publication bias, this plot should look like a funnel where the less precise studies are scattered more widely around the point estimate than the more precise studies. If there is publication bias, studies with lower precision and a small ES (that did not get published) should be missing causing an asymmetry at the base of the funnel. For each of the six meta-analyses we checked the funnel plot for publication bias. Subsequently, we formally tested whether there was funnel plot asymmetry using Egger's regression test (Egger, Smith, Schneider, & Minder, 1997). This test aims to quantify funnel plot asymmetry by regressing the standard normalized ES against its precision (the inverse of the standard error). If there is funnel plot asymmetry the intercept of the regression line should be significantly different from zero (where $p = 0.10$ is chosen as the statistical threshold to compensate for the limited power of the test). Furthermore, we computed Orwin's fail-safe number of studies (Orwin, 1983) which is the number of unpublished and statistically non-significant studies that is needed to reduce the observed ES to a negligible effect (which we defined as a g of -0.20). A large fail-safe number of studies gives credence to the robustness of the observed ES. Lastly, we used the trim-and-fill method (Duval & Tweedie, 2000) to correct the observed ES for publication bias. This method iteratively trims small studies on the positive side of the funnel plot until it is symmetric, 'fills' the funnel plot with the trimmed studies and their mirror images (in order not to underestimate the variance), and recalculates an adjusted pooled ES. It is assumed that the adjusted ES is a more precise estimate of the true ES since it also incorporates unpublished studies.

2.6. Moderating variables

As discussed in the Introduction, the lesion literature suggests several variables that can moderate the relationship between lateralized brain damage and performance on prosodic perception tasks. These can be broadly categorized as either sample or task characteristics.

2.6.1. Sample characteristics

First, it is possible that hemispheric specialization patterns for prosodic processing differ between languages, especially for linguistic prosody (e.g., see Gandour et al., 2003). Therefore, we tested whether the mean weighted ES differed between studies using (American) English subjects and studies that used non-English speaking populations. There was not enough variation in languages between studies to further differentiate between specific languages. Second, studies varied in the proportion of the patient sample that had objective (radiologically or surgically confirmed) evidence for the lateralized nature of the brain damage. Studies that provide objective evidence for the lateralized nature of the brain damage for all patients will provide a more accurate picture of the effect of LHD and RHD on performance than studies that do not (as in the latter case the possibility cannot be excluded that damage was in fact contralateral or bihemispheric for some patients). We therefore tested whether there was a difference in the mean weighted ES between studies that presented objective evidence for the lateralized nature of the damage for the total sample versus studies that did so for the majority of the sample (but not all patients). Third, when experimental groups significantly differ in demographic (such as sex or age) or neuropsychological variables (such as attention deficit) that could affect prosodic processing (see Fecteau, Armony, Joanette, & Belin, 2004; Schirmer, Striano, & Friederici, 2005), it is possible to erroneously conclude that differences between groups in prosodic processing are due to differences in lateralized brain damage while in reality the aforementioned confounders are (partially) responsible. Therefore, we tested whether there were differences in the mean weighted ES between studies that had matched groups on at least one demographic or neuropsychological variable versus studies that did not match experimental groups. Finally, it is possible that the nature of the brain damage influences the occurrence of prosodic perception deficits. Therefore we tested for a moderating influence of etiology of the brain damage by comparing studies that tested patients with only vascular damage, only other than vascular damage or vascular and non-vascular damage.

2.6.2. Task characteristics

First, similarly to Hoekert et al. (2007) we tested the following task quality parameters: (1) whether or not the article reported psychometric (reliability, validity) information about the task; (2) whether the actors that produced the prosodic material were professional actors or phoneticians or not; (3) whether at least six items per prosodic category were used for the task or less (since fewer than six items compromises the reliability of the task) and (4) whether six or more or less than six prosodic categories were used for the task (as more than six prosodic categories might tax working memory of the patients too much and confound degradation of performance due to *prosodic* processing deficits).

The influence of 'verbal-articulatory load' (Ross, Thompson, & Yenkosky 1997) was tested in multiple ways. First, it was tested whether there was a difference in the mean weighted ES for tasks that used speech material that contained lexical semantics versus tasks that used material without lexical semantics (such as pseudo-language or low-pass filtered speech). Secondly, we tested whether there was an effect of the extent to which the response procedure taxed verbal abilities: we contrasted tasks that demanded a verbal response (such as pointing to a verbal label) versus tasks that required a nonverbal response (such as pointing to a facial expression) versus tasks that allowed for both modes of responding. This moderator is also interesting from a different perspective than the variation in verbal load: one could argue that tasks that allow for multiple ways of responding are a more valid measure of *prosodic* processing than tasks that only allow one kind of response as performance is less affected by response-specific (non prosodic) factors. Lastly, we compared identification with discrimination performance. Identification tasks tax verbal capacities to a greater extent than discrimination tasks (Pell, 2006) as in the former case prosodic information needs to be associated with verbal categories while in the latter case no such verbal categorization is required. Because many studies that used discrimination tasks also used identification tasks (forcing us to pool those tasks to preserve independence), we decided to contrast studies that used both tasks to studies that only used identification tasks instead of comparing discrimination to identification orthogonally.

Contrasting the effect of lateralized brain damage on discrimination versus identification performance could also be interesting with regard to the proposed subdivision of the prosodic perception process in at least three stages (Schirmer & Kotz, 2006). One could argue that discrimination tasks primarily tap early processes (acoustic analysis, determination of emotional or linguistic significance) while identification tasks additionally tap the later evaluative processes. Therefore, contrasting these two tasks allows one to investigate whether lateralized brain damage differentially disturbs earlier or later prosodic processing stages.

Lastly, a number of variables were checked for moderating effects but these variables did not vary enough between studies to perform a sufficiently powerful moderator analysis (i.e. there were fewer than five studies per level of the moderator). These moderators included the size of the (prosody carrying) unit (e.g., Gandour et al., 2003) as most studies used sentences, the length of the lesion-onset-testing interval (nearly all studies tested patients in the 'chronic stage' i.e., used intervals of longer than 3 months) and whether lesion localization was only cortical or cortical and also subcortical (nearly all studies involved patients with cortical and subcortical damage). Similarly, we also aimed to test the hypothesis that perception of positive versus negative (or approach versus withdrawal) emotions is lateralized differently but the number of studies presenting data for these emotional categories separately was too small for a meaningful statistical analysis.

2.7. Explorative intra-hemispheric analysis

Although the current meta-analysis is focused on *inter*hemispheric differences in prosodic perception performance, several authors have proposed specific roles for different *intra*hemispheric loci. For instance, Schirmer and Kotz (2006) have proposed that the (bilateral) superior temporal sulcus and gyrus (STS/STG) are involved in the determination of emotional significance of prosodic cues while at a later stage prosodic information is evaluated and integrated with other cognitive processes in the frontal cortex. Ross and Monnot (2008) have also proposed a different role for temporal lobe and frontal lobe areas in emotional prosodic perception. These authors posit that the temporal operculum of the right hemisphere is crucial for adequate emotional prosodic perception and suggest that while damage to the right hemispheric temporal operculum differentially degrades emotional prosodic perception performance as compared to frontal cortical damage, for the left hemisphere the intrahemispheric cortical locus of the lesion does not predict prosodic perception performance.

Given these intrahemispheric models of emotional prosodic perception it would be interesting to meta-analytically test whether the intrahemispheric locus of the brain damage influences prosodic perception performance. Unfortunately, there were not enough studies reporting performance measures separately for subgroups of patients with damage restricted to certain intrahemispheric cortical loci, preventing us from directly meta-analytically investigating this issue. However, a more explorative analysis was undertaken. For each study, the number of patients with temporal and frontal lobe damage (as reported by the authors) was counted and the percentage of patients with temporal and frontal damage was derived. Subsequently, for each patient group as compared to controls and for emotional and linguistic prosody, we performed a meta-regression analysis (e.g., see Lipsey & Wilson, 2001) to investigate the influence of percentage temporal and frontal lobe damage on effect size for prosodic perception performance. A weighted least squares regression analysis was performed with percentage temporal and percentage frontal damage as the predictors, the inverse variance under the random effects model as the regression weight and the effect size as the criterion. Similarly to Alink et al. (2008), we converted the effect size to Fisher's Z and used this metric as the dependent for the meta-regression analysis as it has superior distribution characteristics (Lipsey & Wilson, 2001). To test statistical significance of predictor beta-weights, a corrected standard error was used as suggested by Lipsey and Wilson (2001).

We also counted the number of patients with subcortical damage and white matter lesions (as reported in the article) for each study. Unfortunately, the percentage of patients with white matter or subcortical damage did not vary enough between studies to allow for a meta-regression analysis. As white matter or subcortial lesions have been implicated in prosodic perception performance (e.g. see Ross & Monnot, 2008) we were concerned that cortical lateralization effects might be confounded with these lesions. Therefore, we used an independent t-test to check whether there was a systematic difference between the right and the left hemisphere in the percentage of patients with subcortical or white matter lesions.

3. Results

A total of 38 studies fulfilled the inclusion criteria: 28 studies provided data for the meta-analyses on emotional prosody (Table 1), and 20 studies for the meta-analyses on linguistic prosody (Table 2). An independent rater (DV) coded the part of the coding form involving the study, sample and task characteristics for 10 studies (26% of all included studies). The mean inter-rater agreement was 97.9% suggesting a high level of reliability of the coding process.

3.1. Emotional prosody

3.1.1. LHD vs. NC

A total of 21 studies (with a total of 287 LHD-patients and 399 NC) provided sufficient data for this comparison (see studies with an NC and LHD group in Table 1). The ES distribution with 95% CIs (the forest plot) can be found in Figure 1. The mean weighted ES *g* (95% CI) under the random effects model was −1.06 (−1.40; −0.71) suggesting a large effect of left hemispheric damage on emotional prosodic processing. Inspection of the ES to measurement precision scatterplot (the funnel plot: Figure 2) suggested publication bias. In Table 3 it can be observed that Egger's regression test confirmed the existence of publication bias. The number of statistically non-significant unpublished studies (Orwin's fail-safe N) to reduce the observed ES to a negligible effect was sufficiently large to give credence to the robustness of the observed effect. Trim and Fill did not result in an adjustment of the mean weighted ES.

The heterogeneity statistic was significant ($Q(20) = 81.23$, $p < 0.0001$) suggesting significant unexplained variance in the ES distribution. None of the moderators explained a statistically significant amount of variance in the ES distribution.

Table 1. Overview of studies on *emotional* prosody.

Study	**Language**	**Etiology**	**CT/MR**	**N_L**	**N_R**	**N_C**
Heilmanet al. (1984)	English	CVA	Most	9	8	15
Tomkins et al. (1985)	English	CVA	Most	11	11	11
Bowers et al. (1987)	English	?	?	10	10	12
Ehlers et al. (1987)	Danish	Mixed	All	5	11	-
Blonderet al. (1991)	English	CVA	All	10	10	10
Brådviket al. (1991)	Swedish	CVA	Most	-	20	18
Lalande et al. (1992)	French	CVA	Most	10	12	16
Van Lancker (1992)	English	CVA	Most	24	13	37
Hornak et al. (1996)	English	Mixed	Most	4	11	16
Pellet al. (1997)	English	CVA	?	10	9	10
Peper et al.(1997)	German	Resection	All	21	19	12
Ross et al. (1997)	English	CVA	All	10	12	16
Schmitt et al. (1997)	German	CVA	?	25	27	26
Breitenstein et al. (1998)	German	Mixed	All	16	16	10
Pell (1998)	English	CVA	All	11	9	10
Wertz et al. (1998)	English	CVA	Most	-	20	18
Karow et al. (2001)	English	Mixed	All	10	10	5
Walker et al. (2002)	English	CVA	All	8	8	8
Adolphs et al. (2002)	English	Mixed	All	25	26	?
Zgaljardic et al. (2002)	English	CVA	L: Most R: All	7	9	7
Charbonneau et al. (2003)	French	CVA	Most	17	15	16
Hornak et al. (2003)	English	Mixed	Most	9	16	48
Kucharska et al. (2003)	Polish	CVA	All	30	30	50
Shamay-Tsoory et al.(2004)	Hebrew	Mixed	All	18	16	19
Harciarek et al. (2006)	Polish	CVA	All	-	30	31
Pell (2006)	English	CVA	All	11	9	12
Rymarczyk et al. (2007)	Polish	CVA	All	-	37	26
Kho et al. (2008)	Dutch	Resection	All	16	15	47

Language = native language of the subjects; CT/MR = objective (CT/MR imaging, or surgical-) verification of lateralized damage for all or most patients; N_L = number of patients with left hemispheric damage; N_R = number of patients with right hemispheric damage; N_C = number of healthy controls.

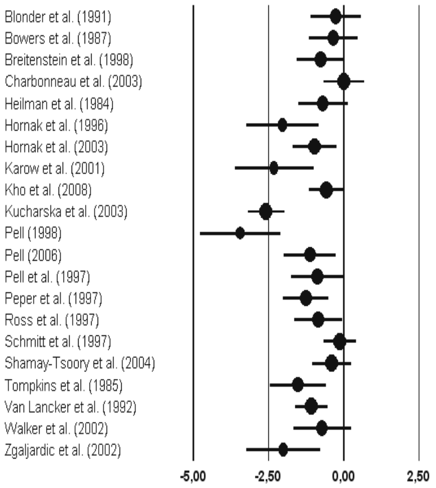

Figure 1. Forest plot of the mean effect size (*g*) and 95% confidence interval for each of the studies included in the comparison of LHD to NC for emotional prosodic perception. Larger symbols indicate a larger number of subjects included in the respective study.

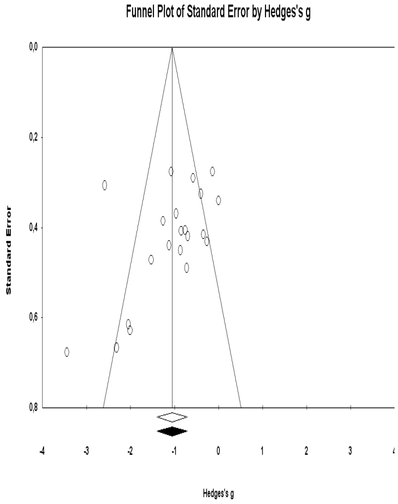

Figure 2. Funnelplot of the studies included in the comparison LHD to NC for emotional prosodic perception. The effect size *g* of each study (x-axis) is plotted against its standard error (y-axis). The vertical line represents the mean weighted effect size and the diagonal lines the 95% confidence interval.

3.1.2. RHD vs. NC

A total of 26 studies (with a total of 402 RHD patients and 508 NC) provided sufficient data for this comparison (see studies with an NC and RHD group in Table 1). The forest plot can be found in Figure 3. The mean weighted ES (95% CI) was −1.41 (−1.76; −1.05) suggesting a large effect of right hemispheric damage on emotional prosodic processing. Inspection of the funnel plot (see Figure 4) suggested publication bias. However, as can be observed in Table 3 Egger's regression test did not reach significance and Orwin's fail-safe gave credence to the robustness of the observed effect. Trim and Fill did not result in an adjustment of the mean weighted ES.

The heterogeneity statistic was significant ($Q(25) = 134.201$, $p < 0.0001$). The mean weighted ES was significantly larger ($Q_B(1) = 4.15$, $p < 0.05$) for studies that presented objective evidence for the lateralized nature of the brain damage for all patients ($g = -1.72$; k = 14) than for studies that provided objective evidence for the majority of the patients or for studies for which this was unknown ($g = -0.04$; k = 12).

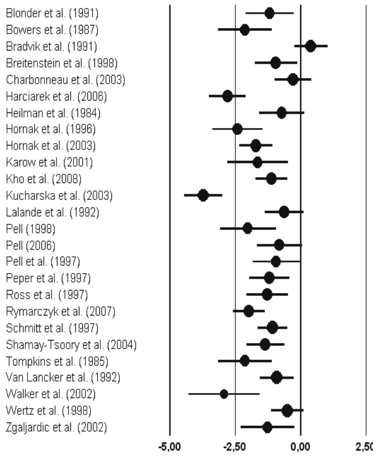

Figure 3. Forest plot of the mean effect size (*g*) and 95% confidence interval for each of the studies included in the comparison of RHD to NC for emotional prosodic perception. Larger symbols indicate a larger number of subjects included in the respective study.

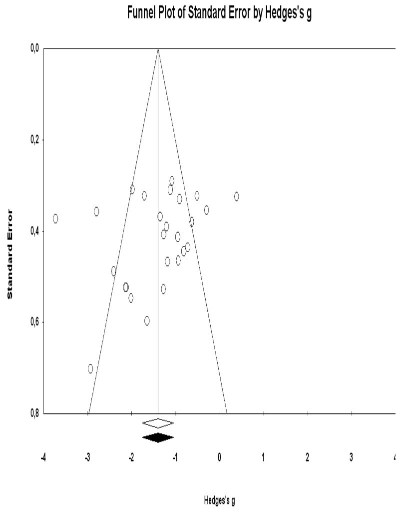

Figure 4. Funnelplot of the studies included in the comparison RHD to NC for emotional prosodic perception. The effect size *g* of each study (x-axis) is plotted against its standard error (y-axis). The vertical line represents the mean weighted effect size and the diagonal lines the 95% confidence interval.

3.1.3. LHD vs. RHD

A total of 22 studies (with a total of 314 LHD and 306 RHD-patients) provided sufficient data for this comparison (see studies with an LHD and RHD group in Table 1). The forest plot can be found in Figure 5. The mean weighted ES (95% CI) was −0.47 (−0.74; −0.20) suggesting a significantly larger effect of right hemispheric damage than left hemispheric damage on emotional prosodic processing. Inspection of the funnel plot (Figure 6) suggested publication bias but as can be observed in Table 2 Egger's regression test did not reach significance and Orwin's fail safe N supported the robustness of the observed effect. Trim and Fill identified two missing studies and led to a slight downward adjustment of the mean weighted ES (95% CI) to −0.37 (−0.66; −0.09).

The heterogeneity statistic was significant (*Q*(21) = 58.74, *p* < 0.0001). None of the moderators explained a statistically significant amount of variance in the ES distribution.

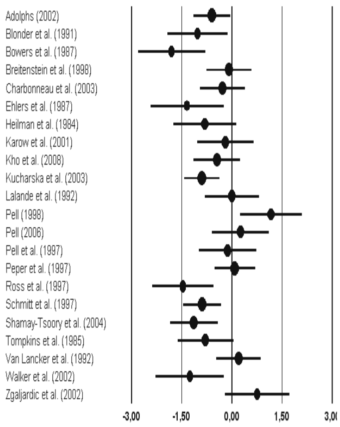

Figure 5. Forest plot of the mean effect size (*g*) and 95% confidence interval for each of the studies included in the comparison of LHD to RHD for emotional prosodic perception. Larger symbols indicate a larger number of subjects included in the respective study.

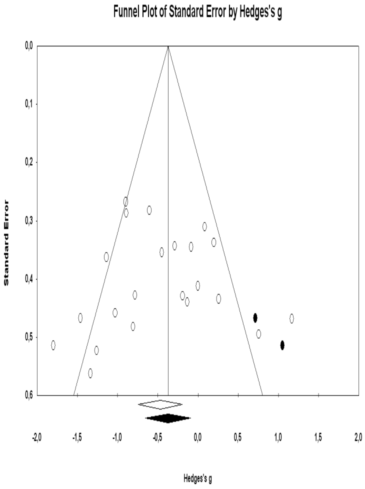

Figure. 6. Funnelplot of the studies included in the comparison LHD to RHD for emotional prosodic perception. The effect size *g* of each study (x-axis) is plotted against its standard error (y-axis). The vertical line represents the mean weighted effect size and the diagonal lines the 95% confidence interval.

Table 2. Publication bias analyses for emotional prosody.

Comparison	**Observed *g* (95% CI)**	**Fail-safe N**	**Egger's intercept**	**Trim and Fill (studies added)**	**Adjusted *g* (95% CI)**
LHD - NC	−1.06 (−1.40; −0.71)	78	−3.22*	0	-
RHD - NC	−1.41 (−1.76; −1.05)	146	−2.94	0	-
LHD - RHD	−0.47 (−0.74; −0.20)	30	−0.13	2	−0.37 (−0.66; −0.09)

LHD - NC = left sided damage vs. controls, RHD - NC = right sided damage vs. controls, LHD - RHD = left sided damage vs. right sided damage. *p < 0.10

3.2. Linguistic prosody

An overview of the studies included in the meta-analysis of linguistic prosody perception can be found in Table 3.

3.2.1. LHD vs. NC

A total of 17 studies (with a total of 211 LHD patients and 249 NC) provided sufficient data for this comparison (see studies with a LHD and NC-group in Table 2). The forest plot can be found in Figure 7. The mean weighted ES (95% CI) was −1.05 (−1.39; −0.71) suggesting a large effect of left hemispheric damage on linguistic prosodic processing. Inspection of the funnel plot (Figure 8) suggested publication bias. As can be observed in Table 4 publication bias was confirmed by Egger's regression test but Orwin's fail-safe N gave credence to the robustness of the observed effect. Trim and Fill identified three missing studies and led to a downward adjustment of the mean weighted ES (95% CI) to −0.81 (−1.20; −0.43).

The heterogeneity statistic was significant ($Q(16) = 44.89$, $p < 0.0001$). None of the moderators explained a statistically significant amount of variance in the ES distribution.

Table 3. Overview of studies on linguistic prosody.

Study	Language	Etiology	CT/MR	N_L	N_R	N_C
Weintraubet al. (1981)	English	Mixed	?	-	9	10
Baumet al. (1982)	English	CVA	All	8	-	8
Heilmanet al. (1984)	English	CVA	Most	9	8	15
Emmoreyet al. (1987)	English	CVA	Most	15	7	15
Bryan(1989)	English	CVA	Most	30	30	30
Blonderet al. (1991)	English	CVA	All	10	10	10
Brådviket al. (1991)	Swedish	CVA	Most	-	20	18
Grosjean(1996)	French	Mixed	Most	10	10	20
Perkins et al.(1996)	English	CVA	All	8	8	8
Baumet al. (1997)	English	CVA	?	10	10	10
Baum et al. (1998)	English	CVA	All	12	10	10
Pell et al. (1997)	English	CVA	?	10	9	10
Borod et al.(1998)	English	CVA	All	10	11	15
Breitenstein et al. (1998)	German	Mixed	All	16	10	10
Pell (1998)	English	CVA	All	11	9	10
Walkeret al. (2002)	English	CVA	All	8	8	8
Zgaljardic et al. (2002)	English	CVA	L: Most R: All	7	9	7
Seddoh (2006)	English	CVA	All	21	-	16
Rymarczyk et al. (2007)	Polish	CVA	All	-	37	26
Kho et al.(2008)	Dutch	Resection	All	16	15	47

Language = native language of the subjects; CT/MR = objective (CT/MR imaging, or surgical-) verification of lateralized damage for all or most patients; N_L = number of patients with left hemispheric damage; N_R = number of patients with right hemispheric damage; N_C = number of healthy controls.

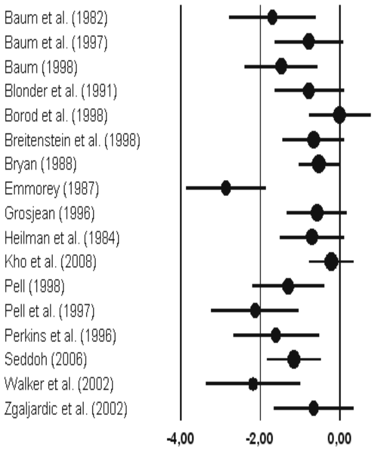

Figure 7. Forest plot of the mean effect size (*g*) and 95% confidence interval for each of the studies included in the comparison of LHD to NC for linguistic prosodic perception. Larger symbols indicate a larger number of subjects included in the respective study.

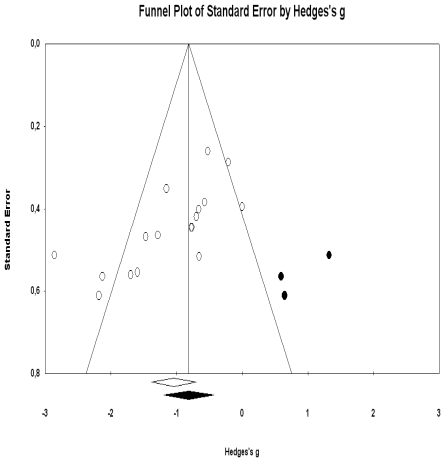

Fig. 8. Funnelplot of the studies included in the comparison LHD to NC for linguistic prosodic perception. The effect size *g* of each study (x-axis) is plotted against its standard error (y-axis). The vertical line represents the mean weighted effect size and the diagonal lines the 95% confidence interval.

3.2.2. RHD vs. NC

A total of 18 studies (with a total of 236 RHD patients and 271 NC) provided sufficient data for this comparison (see studies with an RHD and NC group in Table 2). The forest plot can be found in Figure 9. The mean weighted ES (95% CI) was −0.88 (−1.11; −0.64) suggesting a large effect of right hemispheric damage on linguistic prosodic processing. Inspection of the funnel plot (Figure 10) suggested publication bias. In Table 4 it can be observed that Eggers regression test confirmed publication bias but Orwin's fail-safe N supported the robustness of the observed effect. Trim and Fill identified four missing studies and led to a downward adjustment of the mean weighted ES (95% CI) to −0.73 (−0.99; −0.47).

The heterogeneity statistic failed to reach significance ($Q(17) = 25.88, p > 0.05$) suggesting a homogeneous ES distribution. None of the moderators explained a significant amount of variation in the ES distribution.

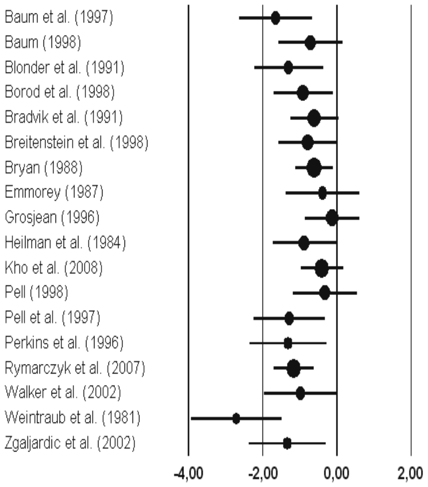

Figure 9. Forest plot of the mean effect size (*g*) and 95% confidence interval for each of the studies included in the comparison of RHD to NC for linguistic prosodic perception. Larger symbols indicate a larger number of subjects included in the respective study.

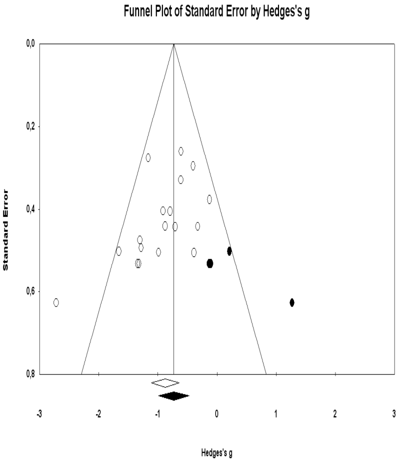

Fig. 10. Funnelplot of the studies included in the comparison RHD to NC for linguistic prosodic perception. The effect size *g* of each study (x-axis) is plotted against its standard error (y-axis). The vertical line represents the mean weighted effect size and the diagonal lines the 95% confidence interval.

3.2.3. LHD vs. RHD

A total of 14 studies (with a total of 172 LHD patients and 160 RHD patients) provided sufficient data for this comparison (see studies with an LHD- and RHD-group in Table 2). The forest plot can be found in the Figure 11. The mean weighted ES (95% CI) was 0.12 (−0.29; 0.52) indicating a negligible (and nonsignificant-) difference between the two cerebral hemispheres in linguistic prosodic processing capability. The funnelplot (Figure 12) was symmetric. However as can be seen in Table 4 Egger's test did suggest publication bias but Trim and Fill did not identify any missing studies.

The heterogeneity statistic was significant ($Q(13) = 43.30, p < 0.0001$). None of the moderator variables explained a significant amount of variance in the ES distribution.

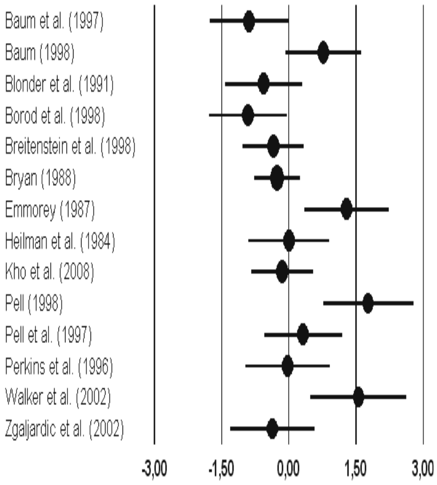

Figure 11. Forest plot of the mean effect size (*g*) and 95% confidence interval for each of the studies included in the comparison of LHD to RHD for linguistic prosodic perception. Larger symbols indicate a larger number of subjects included in the respective study.

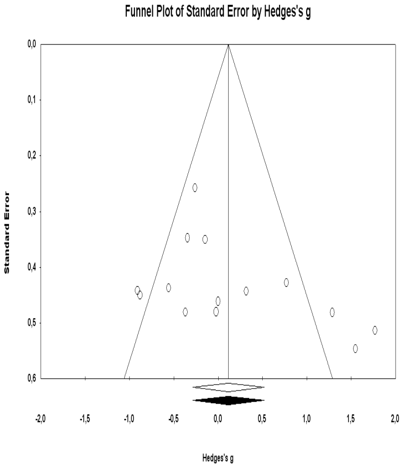

Figure 12. Funnelplot of the studies included in the comparison LHD to RHD for linguistic prosodic perception. The effect size *g* of each study (x-axis) is plotted against its standard error (y-axis). The vertical line represents the mean weighted effect size and the diagonal lines the 95% confidence interval.

Table 4. Publication bias analyses for meta-analyses for linguistic prosody.

Comparison	**Observed *g* (95% CI)**	**Fail-safe N**	**Egger's intercept**	**Trim and Fill (studies added)**	**Adjusted *g* (95% CI)**
LHD - NC	−1.05 (−1.39; −0.71)	59	−4.8*	3	−0.81 (−1.20; −0.43)
RHD - NC	−0.88 (−1.11; −0.64)	58	−2.4*	4	−0.73 (−0.99; −0.47)
LHD - RHD	0.12 (−0.29; 0.52)	-	3.7*	0	-

3.3. Overview and comparison of the meta-analyses

The mean weighted ES and 85% CI under the random effects model for all six meta-analyses are presented in Figure 13. It can be observed that damage to each hemisphere compromises performance on both emotional and linguistic prosodic functions as compared to controls but that there is a non-significant trend for emotional prosody to be more disturbed following RHD than LHD while the reverse holds for linguistic prosody. Furthermore RHD compromises emotional prosodic processing more than linguistic prosodic processing. Finally, when comparing RHD to LHD directly there is evidence for right-hemispheric specialization for emotional prosody but no clear hemispheric specialization for linguistic prosodic processing.

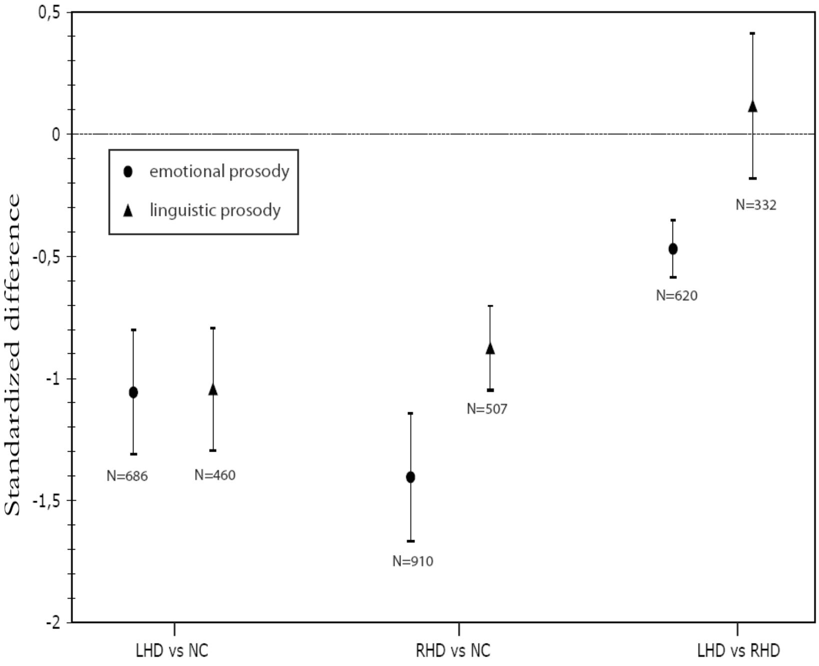

Figure 13. The mean weighted effect size *g* (y-axis) with 85% CI is shown for each of the six meta-analyses (x-axis). It can be observed that both left hemispheric damage (LHD) and right hemispheric damage (RHD) compromises both emotional (circles) and linguistic (triangles) prosodic perception as compared to normal controls (NC). Furthermore, RHD degrades emotional processing more than linguistic processing. Lastly, RHD compromises emotional prosodic perception as compared to LHD while there is no significant difference between these groups for linguistic prosodic perception.

3.4. Qualitative mini-review of studies reporting insufficient data

A total of seven studies did not report enough statistics to compute Hedges's *g* but did fulfill the other inclusion criteria. To give a complete review of the literature on hemispheric specialization for prosodic perception, these studies will now be briefly reviewed.

3.4.1. Emotional prosody

Four studies used emotional prosodic perception tasks. Cancelliere and Kertesz (1990) focused on the influence of the intrahemispheric location of brain damage on emotional processing. Semantically neutral sentences that had been pronounced in happy, sad, angry, or neutral prosody were presented to 28 RHD, 17 LHD patients, and 20 NC. Participants were requested to identify the prosodic category. Because the focus was on *intra*hemispheric damage, the two lateralized damaged groups were not compared directly. Using the lesion-overlap method, the authors concluded that the basal ganglia are important for prosodic perception (irrespective of the side of the lesion).

Starkstein, Federoff, Price, Leiguarda and Robinson (1994) examined the effects of acute lateralized brain damage on emotional prosodic perception. A total number of 59 patients (numbers of LHD or RHD patients unclear) and 17 NC were presented with two prosodic perception tasks. For the first task, semantically neutral sentences were intoned in happy, sad, or angry prosody and participants were asked to identify the prosodic category. In a second task, the instructions were the same but the sentences contained emotional semantics that was either congruent or incongruent with the prosodic layer of the utterance. For the first task, LHD and RHD groups were not directly compared but for the second task the authors reported worse performance for RHD as compared to LHD.

Weddell (1994) investigated the effect of cortical and subcortical damage on emotional processing. Ten patients with damage to the wall or floor of the third ventricle (hypothalamus), 27 RHD and 24 LHD patients (which included cortical and subcortical lateralized damage) and 15 controls with spinal lesions had to identify the prosodic category (happy, sad, angry, surprised or neutral) of semantically neutral or congruent utterances. The authors reported poorer performance for the RHD group as compared to the LHD group but only when emotion was expressed solely through prosody. Furthermore, basal ganglia damage (irrespective of the side of the lesion) impaired emotional prosodic processing. The authors concluded that the bilateral basal ganglia are important for emotional perception.

Zaidel et al. (2002) administered an affective prosodic perception test to 23 RHD, 12 LHD patients and 21 NC among a set of tasks that tapped pragmatics. Participants were requested to identify the emotional prosodic category (happy, sad, angry, neutral) of semantically neutral sentences. Both brain damaged groups performed worse than NC but the authors did not report a significant difference between the brain damaged groups. The authors concluded that there is no clear hemispheric specialization for pragmatic aspects of language.

3.4.2. Linguistic prosody

Three studies used linguistic prosodic perception tasks. Aasland and Baum (2003) investigated the sensitivity of LHD and RHD patients to durational cues in determining phrasal boundaries. The duration of pauses and of pre-boundary words in the utterance

"Pink and black and green" was systematically varied to obtain different phrase structures. In two experiments, 10 LHD, 9 RHD patients, and 10 NC were asked to identify the phrase structure by pointing to a picture that depicted the correct grouping of constituents (colors). This design allowed the authors to parametrically investigate the interaction of durational cues and lateralized hemispheric damage on phrasal boundary processing. Overall, both patient groups performed worse than NC in identifying phrasal boundaries, but the LHD group showed improved performance when temporal cues were exaggerated while the RHD group did not. The authors concluded that while a heightened temporal cue threshold might be responsible for the deficiency in phrasal boundary processing for the LHD group, the exact mechanism for the deficit in RHD group is unclear.

Abada et al. (2006) investigated whether the left and right hemisphere are differentially sensitive to metrical stress. Monosyllabic words (such as 'mint') were embedded in disyllabic nonwords where the second syllable was either weak (such as in "mintef") or strong (such as in "mintayf"). Previous research had shown that detection of the proper word is faster when the nonsense syllable is weak as compared to a strong nonsense syllable as there is a tendency to attempt lexical access at strong syllables. Groups of 10 LHD, 10 RHD patients and 10 NC were required to press a button as they detected a proper word. All three groups showed the strong syllable effect but overall the LHD group performed worse than NC while RHD patients did not differ from the other two groups which led the authors to conclude that LHD induces difficulties utilizing stress patterns.

Seddoh (2006b) presented meaningful and nonsense sentences that were intonated as statements or questions to 13 LHD, 8 RHD patients and 12 NC. Participants were asked to identify whether the utterance was a question or a statement. Both patient groups performed worse than the control group but only for utterances with question intonation. Seddoh suggested that the patients might have had difficulties in processing specific components of the pitch contour.

3.4.3. Conclusion of the mini-review

Taken together the studies on emotional prosodic perception indicate bilateral processing of emotional prosody with subcortical involvement and a possible relative right hemispheric advantage. This is in line with the quantitative analysis of emotional prosodic perception studies. The small number of studies on linguistic prosody generally points to bilateral processing, which is also in accordance with the quantitative analysis.

3.5. Explorative intrahemispheric analysis

Table 5 presents the percentage of patients with damage to intrahemispheric locations for each study. To explore a possible differential impact of temporal versus frontal damage on prosodic processing performance, a weighted least squares random effects meta-regression analysis was performed with the percentage of patients with temporal lobe damage and the percentage of patients with frontal lobe damage as predictors and the Fisher's Z transformed ES as the dependent variable. For none of the meta-analyses percentage temporal or percentage frontal damage explained a significant amount of variance in the effect size distribution (for all: $-0.3 < \beta < 0.7$, $p > 0.05$). A series of independent t-tests revealed no difference between the LHD and RHD groups in the percentage of patients with subcortical damage or white matter lesions for any of the meta-analyses (for all: $-0.71 < t < 0.78, p > 0.05$).

Table 5. Percentage of patients with damage to intrahemispheric loci for each study.

Study	**Used in comparison**	**Patient group**	**White matter**	**Sub-cortical**	**Temporal**	**Frontal**
Weintraubet al. (1981)	Linguistic	Right	0	11.11	66.67	22.22
Baum et al. (1982)	Linguistic	Left	?	?	?	?
Heilman et al. (1984)	Linguistic+ Emotional	Right	0	0	0	50.00
		Left	0	0	22.22	55.56
Tomkins et al. (1985)	Emotional	Right	?	?	?	?
		Left	?	?	?	?
Bowers et al. (1987)	Emotional	Right	?	?	?	?
		Left	?	?	?	?
Emmorey et al. (1987)	Linguistic	Right	0	28.57	14.28	28.57
		Left	0	0	6.67	60.00
Bryan(1989)	Linguistic	Right	0	0	36.67	26.67
		Left	0	0	23.33	13.33
Blonder et al. (1991)	Linguistic+ Emotional	Right	0	0	60.00	60.00
		Left	0	0	80.00	50.00
Brådvik et al. (1991)	Emotional	Right	?	?	?	?
Lalande et al. (1992)	Linguistic+ Emotional	Right	?	?	?	?
Van Lancker (1992)	Emotional	Right	?	?	?	?
		Left	?	?	?	?
Grosjean (1996)	Linguistic	Right	20.00	30.00	60.00	60.00
		Left	0	70.00	50.00	70.00
Hornak et al. (1996)	Emotional	Right	9.09	27.27	27.27	72.72
		Left	50.00	0	0	50.00
Perkins et al.(1996	Linguistic	Right	12.50	25.00	62.50	62.50
		Left	25.00	37.50	50.00	50.00
Pell et al. (1997)	Linguistic + Emotional	Right	?	?	?	?
		Left	?	?	?	?
Baum et al. (1997)	Linguistic	Right	20.00	0	20.00	0
		Left	0	0	10.00	40.00
Peper et al.(1997)	Emotional	Right	?	?	?	?
		Left	?	?	?	?

Ross et al. (1997)	Emotional	Right	?	?	?	?
		Left	90.00	30.00	30.00	50.00
Schmitt et al. (1997)	Emotional	Right	?	63.00	?	?
		Left	?	56.00	?	?
Baum et al. (1998)	Linguistic	Right	20.00	0	30.00	0
		Left	0	8.33	8.30	33.33
Borod et al. (1998)	Linguistic	Right	27.27	36.36	0	45.45
		Left	40.00	20.00	30.00	30.00
Breitenstein et al. (1998)	Linguistic+ Emotional	Right	0	12.50	37.50	50.00
		Left	0	6.25	31.25	50.00
Pell (1998)	Linguistic+ Emotional	Right	9.09	0	27.27	9.09
		Left	0	9.09	9.09	36.36
Wertz et al. (1998)	Emotional	Right	?	25.00	30.00	25.00
Karow et al. (2001)	Emotional	Right	20.00	50.00	40.00	50.00
		Left	10.00	50.00	20.00	40.00
Walker et al. (2002)	Linguistic+ Emotional	Right	0	0.25	37.50	37.50
		Left	0	0	37.50	87.50
Zgaljardic et al. (2002)	Linguistic+ Emotional	Right	22.22	0	44.44	22.22
		Left	14.29	14.29	57.14	28.57
Charbonneau et al. (2003)	Emotional	Right	33.33	26.67	6.67	40.00
		Left	23.53	5.88	17.64	29.41
Hornak et al. (2003)	Emotional	Right	0	0	0	100.00
		Left	0	0	0	100.00
Kucharska et al. 2003)	Emotional	Right	?	6.67	33.33	30.00
		Left	?	1.00	30.00	30.00
Shamay-Tsoory et al. (2004)	Emotional	Right	0	0	0	56.25
		Left	0	0	0	55.56
Pell (2006)	Emotional	Right	11.11	22.22	11.11	0
		Left	0	18.18	9.09	36.36
Seddoh (2006)	Linguistic	Right				
		Left				
Rymarczyk et al. (2007)	Linguistic+ Emotional	Right	?	28.85	?	28.88
Kho et al. (2008)	Linguistic+ Emotional	Right	0	100.00	100.00	0
		Left	0	100.00	100.00	0

4. Discussion

The current meta-analyses suggest that both cerebral hemispheres are necessary for adequate explicit emotional and linguistic prosodic perception. Within Cohen's (1988) framework for qualifying effect size, the detrimental effect of both RHD and LHD (as compared to NC) on both linguistic and emotional prosodic perception performance can be considered large (i.e., g was larger than 0.80 for all comparisons). However, when comparing the detrimental effect of LHD and RHD (as compared to NC), it was shown that while LHD degrades emotional and linguistic prosodic perception to a similar degree (−1.06 and −1.05 respectively) RHD degraded emotional prosodic perception performance more (g = −1.41) than linguistic prosodic perception performance (g = −0.88). Finally, when comparing both patient groups directly, RHD degraded emotional prosodic processing more than LHD (g = −0.47) while there was no differential degradation of lateralized brain damage for linguistic prosodic perception performance. These results are therefore compatible with the notion of bihemispheric control over the perception of linguistic and emotional prosody with a *relatively* greater contribution of the right hemisphere to emotional prosodic perception.

Strong versions (absolute lateralization) of the "categorical" lateralization hypotheses mentioned in the introduction can therefore not be supported. The results of the present study clearly do not support even a weak version of global control of the right hemisphere over *all* prosodic perception (Klouda et al., 1988) since LHD compromised linguistic prosodic perception performance to a comparable degree as RHD did. A relative version of the Functional lateralization hypothesis (Van Lancker, 1980) can not be supported either. Although RHD compromised emotional perception more than LHD no significantly larger performance degradation for LHD than RHD was found for linguistic prosodic processing, as would have been predicted by this hypothesis. However, our results are compatible with a weak version of the right hemisphere hypothesis (Ross, 1981; Blonder et al., 1991; Borod et al., 1998) where there is bilateral processing for emotional prosodic perception but with larger right than left hemispheric contribution. As discussed in the introduction, the current design does not permit us to differentiate between the "categorical" and cue-dependent lateralization hypotheses of prosodic perception. Hence, it is unclear whether the right hemispheric superiority for emotional prosodic processing originates in superior processing of acoustics necessary for adequate emotional prosodic perception (Van Lancker et al., 1992) or superiority in processing of emotional prosodic categories (in partial support of the Functional lateralization hypothesis) or both (for an informative discussion of this issue see Pell, 1998).

Another important line of evidence regarding hemispheric specialization for prosodic perception is provided by the neuroimaging literature. Activation maps of fMRI/PET studies are highly dependent on the exact experimental and control condition (i.e. contrast) used. For instance, contrasting emotional versus linguistic decisions for the same prosodic material in a categorization task is mainly sensitive to later stages of the prosodic perception process (identification of emotional or linguistic information), whereas contrasting emotional prosody to neutral prosody additionally taps into earlier stages (acoustic analysis). Hence, comparing neuroimaging studies is not without difficulty. With this caveat in mind, imaging studies to date have found bilateral temporofrontal (STG/STS, IFG) activations for emotional (Buchanan et al., 2000; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Kotz et al., 2003;

Grandjean et al., 2005; Ethofer et al., 2006a; Ethofer et al., 2006b; Mitchell & Ross, 2008; Ethofer et al., 2009; Ethofer, Van de Ville, Scherer, & Vuilleumier, 2009) and linguistic (Gandour et al., 2003; Wildgruber et al., 2004; Doherty, West, Dilley, Shattuck-Hufnagel, & Caplan, 2004; Humphries, Love, Swinney, & Hickok, 2005; Aleman et al., 2005) prosodic perception tasks. Some of these imaging studies have additionally found more (extended) right than left hemispheric activation for emotional prosodic processing (Buchanan et al., 2000; Mitchell et al., 2003; Ethofer et al., 2006a; Ethofer et al., 2009) while for the limited number of studies that used linguistic prosodic perception tasks, the results are mixed with some finding more (extended) activation in the left (Wildgruber et al., 2004; Aleman et al., 2005) and others in the right (Gandour et al., 2003; Doherty et al., 2004) hemisphere. These studies therefore suggest that there is a bilateral temporofrontal network of areas involved in the perception of linguistic and emotional prosody with possibly a relative right hemispheric superiority for emotional prosodic perception. The results of our meta-analysis are clearly in keeping with the neuroimaging literature and additionally suggest that within this bilateral network of sufficient areas both hemispheres are also *necessary* for the perception of prosody. Moreover, our results suggest that the stronger activation of the right hemisphere found by some neuroimaging studies might reflect relative superiority of the right hemisphere in the perception of emotional prosody. Our explorative meta-regression analysis of intrahemispheric contributions to prosodic perception did not reveal a differential impact of temporal versus frontal damage on prosodic perception performance. However, due to the indirect nature of this analysis this result should not be taken as conclusive.

Although we tested in multiple ways whether the 'verbal-articulatory demands' of prosodic perception tasks (Ross et al., 1997) moderated the effect of lateralized brain damage on prosodic perception performance, we failed to find such effects. Increased verbal demands on prosodic tasks (such as the presence of lexical semantics or the need to respond verbally) did not increase the effect of left hemispheric damage on emotional perception performance. However, one moderating variable yielded a statistically robust effect. Studies that provided objective evidence that brain damage was indeed right lateralized for all patients, found a significantly larger effect of RHD on emotional prosodic perception performance than studies that did not provide such evidence for all patients. It is therefore possible that in the latter case patients with left or bilateral brain damage had been included, which might have decreased the ES (and hence the reported mean weighted ES for the effect of RHD on emotional prosodic perception reported here might represent an underestimation). A clear recommendation for future studies, then, is that presumed lateralized brain damage for all patients should be radiologically confirmed in order to get an accurate measure of the effect of lateralized brain damage.

There were some limitations to the present investigation. First, the study design does not permit us to conclude that the observed division of labor between the hemispheres is *specific* to prosodic processing as we have not included non-prosodic control tasks. Hence, although we cannot exclude the possibility that the results might generalize to linguistic and emotional processing in other modalities, we can conclude that the results are *at least* valid for prosodic processing. A second and related issue that has already been pointed out is that the current design does not permit us to differentiate cue-dependent versus "categorical" lateralization (which might be modality-independent) explanations of the right hemispheric advantage for emotional prosodic processing. Thirdly, to maximize sensitivity of our meta-analysis to hemispheric specialization we were forced to pool effect sizes over the levels of moderators. This may have decreased the power of our moderator analyses so that the (absence of) effects in our moderator analysis should not be taken as conclusive. Fourth, an often mentioned critique on meta-analysis is that it quantitatively combines studies that are incomparable and therefore should not be combined. In this light, it might be possible that different linguistic prosodic functions follow a separate pattern of hemispheric lateralization and that the net bilateral control that was found in the current meta-analysis reflects this heterogeneity in hemispheric specialization. Unfortunately, the number of studies per prosodic linguistic phenomenon was too small for a statistically robust test of this issue leaving this a matter for future investigation. Finally, although we focused on cortical lateralization of prosodic perception, many of the studies that were analyzed in the current meta-analysis included patients with cortical *and* subcortical damage. Therefore, we cannot rule out a contribution of subcortical structures to the observed pattern of hemispheric involvement in prosodic perception (indeed many studies have suggested involvement of subcortical structures in prosodic processing such as the basal ganglia, see e.g. Cancelliere & Kertesz, 1990; Blonder, Gur, & Gur, 1989). Hence, when assessing the effect size of lateralized brain damage on prosodic perception performance as compared to controls in isolation it should be kept in mind that the effect likely reflects both cortical, subcortical, and white matter damage. However, as it seems likely that subcortical involvement was present in the right and left hemisphere to a similar degree (which was confirmed by the analysis of the percentage of patients with subcortical damage as reported in each paper) this does not complicate our comparisons of left with right hemispheric damage.

In sum, we conclude that the network of necessary areas dedicated to the perception of linguistic and emotional prosody is *bilateral* with only *relative* right hemispheric specialization for emotional prosodic perception at best.

5. References

References that were included in the meta-analysis are marked with an asterisk (*)

Aasland, W. A., & Baum, S. R. (2003). Temporal parameters as cues to phrasal boundaries: A comparison of processing by left- and right-hemisphere brain-damaged individuals. *Brain and Language, 87*, 385-399.

Abada, S. H., & Baum, S. R. (2006). Sensitivity to metrical stress patterns in word recognition in individuals with left and right hemisphere brain damage. *Brain and Language, 99,* 39-40.

* Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology, 12,* 169-177.

Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., De Haan, E. H. F., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex, 15,* 221-228.

Alink, L. R. A., Van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Mesman, J., Juffer, F., & Koot, H. M. (2008). Cortisol and externalizing behavior in children and adolescents: Mixed meta-analytic evidence for the inverse relation of basal cortisol and cortisol reactivity with externalizing behavior. *Developmental Psychobiology, 50,* 427-450.

Alves, N. T., Fukusima, S., & Aznar-Casanova, A. (2008). Models of brain asymmetry in emotional processing. *Psychology & Neuroscience, 1,* 63-66.

* Baum, S. R. (1998). The role of fundamental frequency and duration in the perception of linguistic stress by individuals with brain damage. *Journal of Speech, Language & Hearing Research, 41,* 31.

* Baum, S. R., Daniloff, J. K., Daniloff, R., & Lewis, J. (1982). Sentence comprehension by broca's aphasics: Effects of some suprasegmental variables. *Brain and Language, 17,* 261-271.

* Baum, S. R., Pell, M. D., Leonard, C. L., & Gordon, J. K. (1997). The ability of right- and left-hemisphere-damaged individuals to produce and interpret prosodic Cues marking phrasal boundaries. *Language & Speech, 40,* 313-330.

* Blonder, L. X., Bowers, D., & Heilman, K. M. (1991). The role of the right hemisphere in emotional communication. *Brain, 114,* 1115-1127.

Blonder, L. X., Gur, R. E., & Gur, R. C. (1989). The effects of right and left hemiparkinsonism on prosody. *Brain and Language, 36,* 193-207.

* Borod, J. C., Cicero, B. A., Obler, L. K., Welkowitz, J., Erhan, H. M., Santschi, C., . . . Whalen, J. R. (1998). Right hemisphere emotional perception: Evidence across multiple channels. *Neuropsychology, 12,* 446-458.

* Bowers, D., Coslet, H. B., Bauer, R. M., Speedie, L. J., & Heilman, K. M (1987). Comprehension of emotional prosody following unilateral hemispheric lesions: Processing defect versus distraction defect. *Neuropsychologia, 25,* 317-328.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis.* Englewood NJ, Biostat. Ref Type: Computer Program.

Borenstein, M., Hedges, L., Higgins J., & Rothstein, H. (2009). *Introduction to Meta-Analysis.* Chichester: Wiley.

* Brådvik, B., Dravins, C., Holtås, S., Rosén, I., Ryding, E., & Ingvar, D. (1991). Disturbances of speech prosody following right hemisphere infarcts. *Acta Neurologica Scandinavica, 84,* 114-126.

* Breitenstein, C., Daum, I., & Ackermann, H. (1998). Emotional processing following cortical and subcortical brain damage: Contribution of the fronto-striatal circuitry. *Behavioural Neurology, 11,* 29-42.

* Bryan, K. L. (1989). Language prosody and the right hemisphere. *Aphasiology, 3,* 285-299.

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., . . .Jäncke, L. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research, 9*, 227-238.

Cancelliere, A. E., & Kertesz, A. (1990). Lesion localization in acquired deficits of emotional expression and comprehension. *Brain and Cognition, 13,* 133-147.

* Charbonneau, S., Scherzer, B. P., Aspirot, D., & Cohen, H. (2003). Perception and production of facial and prosodic emotions by chronic CVA patients. *Neuropsychologia, 41,* 605-613.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*, 141-201.

Doherty, C. P., West, W. C., Dilley, L. C., Shattuck-Hufnagel, S., & Caplan, D. (2004). Question/statement judgments: An fMRI study of intonation processing. *Human Brain Mapping, 23,* 85-98.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455-463.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629-634.

* Ehlers, L., & Dalby, M. (1987). Appreciation of emotional expressions in the visual and auditory modality in normal and brain-damaged patients. *Acta Neurologica Scandinavica, 76,* 251-256.

* Emmorey, K. D. (1987). The neurological substrates for prosodic aspects of speech. *Brain and Language, 30,* 305-320.

Ethofer, T., Anders, S., Wiethoff, S., Erb, M., Herbert, C., Saur, R., . . . Wildgruber, D. (2006a). Effects of prosodic emotional intensity on activation of associative auditory cortex. *Neuroreport, 17,* 249-253.

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., . . Wildgruber, D. (2006b). Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. *NeuroImage, 30,* 580-587.

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience, 21*, 1-14.

Ethofer, T., Van de Ville, D., Scherer, K., & Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology, 19,* 1028-1033.

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 and Cognition, 55, 300*-302.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., . . . Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language, 84,* 318-336.

Gandour, J., Tong, Y., Wong, D., Tavalage, T., Dzemidzic, Xu, Y., . . . Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *NeuroImage, 23,* 344-357.

Goldstein H., & Healy, M. J. R. (1995). The graphical presentation of a collection of means. *Journal of the Royal Statistical Society A, 158,* 175–177.

Geigenberger, A., & Ziegler, W. (2001). Receptive prosodic processing in aphasia. *Aphasiology, 15,* 1169-1187.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience, 8,* 145-146.

* Grosjean, F. (1996). Using prosody to predict the end of sentences in English and French: Normal and brain-damaged subjects. *Language and Cognitive Processes, 11,* 107-134.

* Harciarek, M., Heilman, K. M., & Jodzio, K. (2006). Defective comprehension of emotional faces and prosody as a result of right hemisphere stroke: Modality versus emotion-type specificity. *Journal of the International Neuropsychological Society, 12,* 774-781.

* Heilman, K. M., Bowers, D., Speedie, L., & Coslett, H. B. (1984). Comprehension of affective and nonaffective prosody. *Neurology, 34,* 917.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego, CA: Academic Press.

Hoekert, M., Kahn, R. S., Pijnenborg, M., & Aleman, A. (2007). Impaired recognition and expression of emotional prosody in schizophrenia: Review and meta-analysis. *Schizophrenia Research, 96,* 135-145.

* Hornak, J., Rolls, E. T., & Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia, 34,* 247- 261.

* Hornak, J., Bramham, J., Rolls, E. T., Morris, R. G., O'Doherty, J., Bullock, P. R., & Polkey, C. E. (2003). Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain, 126,* 1691-1712.

Humphries, C., Love, T., Swinney, D., & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Human Brain Mapping, 26,* 128-138.

* Karow, C. M., Marquardt, T. P., & Marshall, R. C. (2001). Affective processing in left and right hemisphere brain-damaged subjects with and without subcortical involvement. *Aphasiology, 15,* 715-729.

* Kho, K. H., Indefrey, P., Hagoort, P., Van Veelen, C. W. M., Van Rijen, P. C., & Ramsey, N. F. (2007). Unimpaired sentence comprehension after anterior temporal cortex resection. *Neuropsychologia, 46,* 1170-1178.

Klouda, G. V., Robin, D. A., Graff-Radford, N. R., & Cooper, W. E. (1988). The role of callosal connections in speech prosody. *Brain and Language, 35,* 154-171.

Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language, 86,* 366-376.

Kotz, S. A., Meyer, M., & Paulmann, S. (2006). Lateralization of emotional prosody in the brain: an overview and synopsis on the impact of study design. *Progress in Brain Research, 156,* 285-294.

Kotz, S. A., & Paulmann, S. (2011). Emotion, Language, and the Brain. *Language and Linguistics Compass, 5,* 108-125.

* Kucharska-Pietura, K., Phillips, M. L., Gernand, W., & David, A. S. (2003). Perception of emotions from faces and voices following unilateral brain damage. *Neuropsychologia, 41,* 1082-1090.

* Lalande, S., Braun, C.M., Charlevois, N., & Whitaker, H. A. (1992). Effects of right and left hemisphere cerebrovascular lesions on discrimination of prosodic and semantic aspects of affect in sentences. *Brain and Language, 42,* 165-186.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. R. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia, 41,* 1410-1421.

Mitchell, R. L. C., & Ross, E. D. (2008). fMRI evidence for the effect of verbal complexity on lateralisation of the neural response associated with decoding prosodic emotion. *Neuropsychologia, 47,* 2880-2887.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational and Behavioral Statistics, 8,* 157-159.

* Pell, M. D. (1998). Recognition of prosody following unilateral brain lesion: Influence of functional and structural attributes of prosodic contours. *Neuropsychologia, 36,* 701-715.

* Pell, M. D. (2006). Cerebral mechanisms for understanding emotional prosody in speech. *Brain and Language, 96,* 221-234.

* Pell, M. D., & Baum, S. R. (1997). The ability to perceive and comprehend intonation in linguistic and affective contexts by brain-damaged adults. *Brain and Language, 57,* 80-99.

* Peper, M., & Irle, E. (1997). Categorical and dimensional decoding of emotional intonations in patients with focal brain lesions. *Brain and Language, 58,* 233-264.

* Perkins, J. M., Baran, J. A., & Gandour, J. (1996). Hemispheric specialization in processing intonation contours. *Aphasiology, 10,* 343-362.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86,* 638-641.

Ross, E. D. (1981). The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology, 38,* 561-569.

Ross, E. D., & Monnot, M. (2008). Neurology of affective prosody and its functional-anatomic organization in right hemisphere. *Brain and Language, 104,* 51-74.

* Ross, E. D., Thompson, R. D., & Yenkosky, J. (1997). Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. *Brain and Language, 56,* 27-54.

* Rymarczyk, K., & Grabowska, A. (2007). Sex differences in brain control of prosody. *Neuropsychologia, 45,* 921-930.

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

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40,* 227-256.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10,* 24-30.

Schirmer, A., Striano, T., & Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport, 16,* 635-639.

* Schmitt, J. J., Hartje, W., & Willmes, K. (1997). Hemispheric asymmetry in the recognition of emotional attitude conveyed by facial expression, prosody and propositional speech. *Cortex, 33,* 65-81.

* Seddoh, S. A. (2006a). Basis of intonation disturbance in aphasia: Perception. *Journal of Neurolinguistics, 19,* 270-290.

Seddoh, S. A. (2006b). A perceptual investigation of intonation categories in patients with unilateral left and right hemisphere damage. *Brain and Language, 99,* 43-44.

* Shamay-Tsoory, S. G., Tomer, R., Goldsher, D., Berger, B. D., & Aharon-Peretz, J. (2004). Impairment in cognitive and affective empathy in patients with brain lesions: Anatomical and cognitive correlates. *Journal of Clinical and Experimental Neuropsychology, 26,* 1113-1127.

Starkstein, S. E., Federoff, J. P., Price, T. R., Leiguarda, R. C., & Robinson, R. G. (1994). Neuropsychological and neuroradiologic correlates of emotional prosody comprehension. *Neurology, 44,* 515.

* Tompkins, C. A., & Flowers, C. R. (1985). Perception of emotional intonation by brain-damaged adults: The influence of task processing levels. *Journal of Speech and Hearing Research, 28,* 527-538.

Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Papers in linguistics: International Journal of Human Communication, 13,* 201-277.

* Van Lancker, D., & Sidtis, J. J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research, 35,* 963-970.

Walker, J. P., Fongemie, K., & Daigle, T. (2001). Prosodic facilitation in the resolution of syntactic ambiguities in subjects with left and right hemisphere damage. *Brain and Language, 78,* 169-196.

* Walker, J. P., Daigle, T., & Buzzard, M. (2002). Hemispheric specialisation in processing prosodic structures: Revisited. *Aphasiology, 16,* 1155-1172.

Weddell, R. A. (1994). Effects of subcortical lesion site on human emotional behavior. *Brain and Cognition, 25,* 161-193.

* Weintraub, S., Mesulam, M. M., & Kramer, L. (1981). Disturbances in prosody. A right-hemisphere contribution to language. *Archives of Neurology, 38,* 742-744.

* Wertz, R. T., Henschel, C. R., Auther, L. L., Ashford, J. R., & Kirshner, H. S. (1998). Affective prosodic disturbance subsequent to right hemisphere stroke: A clinical application. *Journal of Neurolinguistics, 11,* 89-102.

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research, 156,* 249-268.

Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., & Ackermann, H. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex, 14,* 1384-1389.

Wilson, D. *Effect Size Determination Program.* [mason.gmu.edu/~dwilsonb/downloads/ES_Calculator.xls]. Accessed on May 16, 2011.

Wong, P. C. M. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin, 59,* 83-95.

Wunderlich, A., Ziegler, W., & Geigenberger, A. (2003). Implicit processing of prosodic information in patients with left and right hemisphere stroke. *Aphasiology, 17,* 861.

Zaidel, E., Kasher, A., Soroker, N., & Batori, G. (2002). Effects of right and left hemisphere damage on performance of the "right hemisphere communication battery". *Brain and Language, 80,* 510-535.

* Zgaljardic, D. J., Borod, J. C., & Sliwinski, M. (2002). Emotional perception in unilateral stroke patients: recovery, test stability, and interchannel relationships. *Applied Neuropsychology, 9,* 159-172.

Chapter 2.2

The nature of hemispheric specialization for prosody perception: ERP evidence

Witteman, J., Goerlich, K. S., Martens, S., Aleman, A., Van Heuven, V. J., & Schiller, N. O. (2014). The nature of hemispheric specialization for prosody perception. *Cognitive Affective & Behavioral Neuroscience, 14,* 1104-1114.

Abstract

Recent evidence suggests that there is relative right hemispheric specialization for emotional prosody perception while linguistic prosody perception is under bilateral control. It is still unknown, however, how hemispheric specialization for prosody perception might arise. Two main hypotheses have been put forward. Cue-dependent hypotheses on the one hand, propose that hemispheric specialization is driven by specialization for non-prosody specific processing of acoustic cues. The functional lateralization hypothesis on the other hand, proposes that hemispheric specialization is dependent on the communicative function of prosody with emotional and linguistic prosody processing lateralized to the right and left hemisphere, respectively. In the present study, the functional lateralization hypothesis of prosody perception was systematically tested by instructing one group of participants to evaluate the emotional prosody and another group to the linguistic prosody dimension of bi-dimensional prosodic stimuli in a dichotic listening paradigm while event related-potentials (ERPs) were recorded. The results showed that the right ear advantage was associated with a decreased latency of an early negativity in the contralateral hemisphere. No evidence was found for functional lateralization. These findings suggest that functional lateralization effects for prosody perception are small and support the structural model of dichotic listening.

1. Introduction

Speech prosody is an important means to convey emotions (e.g., indicating whether the speaker is angry or sad) and linguistic structure (e.g., signifying a statement or a question). A longstanding question in the neuropsychological literature has been whether there is hemispheric specialization for the perception of prosody. Recent meta-analytic evidence suggests that there is a *relative* right hemispheric specialization for emotional prosody perception, while the processing of linguistic prosody seems to be controlled bilaterally (Witteman, Van IJzendoorn, Van de Velde, Van Heuven, & Schiller, 2011).

However, the nature of this relative right hemispheric specialization for emotional prosody is currently unknown. Two mechanisms have been proposed as to *how* hemispheric specialization for prosody perception might arise. On the one hand, *cue-dependent* lateralization hypotheses propose that right-hemisphere specialization for emotional prosody perception can be explained by a (non-prosody specific) advantage of the right hemisphere for early acoustic processing, such as spectral processing (Van Lancker & Sidtis, 1992), as spectral parameters appear to be particularly important for decoding emotional prosody (Scherer, 2003). Indeed, a recent meta-analysis of the neuroimaging literature of emotional prosody perception found preliminary evidence for relative right hemispheric specialization of the primary and secondary auditory cortex (Witteman, Van Heuven, & Schiller, 2012), which could be interpreted as (indirect) support for the cue-dependent lateralization hypothesis.

On the other hand, the *functional* lateralization hypothesis posits that the right hemisphere is specialized in the processing of the emotional communicative function of emotional prosody (Van Lancker, 1980), whereas the left hemisphere might be specialized in the processing of linguistic prosodic function. The dynamic dual pathway model by Friederici and Alter (2004) further suggests that when linguistic prosody is more bound to segments (such as in the case of metrical stress), it is left lateralized, while linguistic prosody at the sentence level (such as boundary marking) is right lateralized. The goal of the present investigation was to systematically test whether there is functional lateralization for prosody perception.

To comprehensively test the functional lateralization hypothesis it is necessary to vary the communicative function of prosodic material while keeping acoustics constant, and to observe whether there is a change in the difference of activity between the hemispheres. Note that the acoustic and functional lateralization hypotheses are non-mutually exclusive and could represent different stages of the prosody perception process. Indeed, recent neural models of prosody perception have suggested that there might be acoustically driven lateralization in an initial processing stage, but more semantically (functionally) driven lateralization in subsequent stages (Bruck, Kreifelts, & Wildgruber, 2011; Kotz & Paulmann, 2011). To shed light on the issue of when in time functional lateralization arises the present ERP study systematically manipulated the function of prosody by instructing one group of subjects to evaluate the emotional prosody dimension and a different group of participants to evaluate the linguistic prosody dimension of identical bi-dimensional stimuli.

Interestingly, Paulmann, Jessen and Kotz (2012) recently reported such a direct comparison of emotional to linguistic prosody perception using the cross-splicing paradigm. A so called 'prosodic expectancy positivity (PEP)' was found that was more pronounced for emotional than linguistic prosody expectancy violations, suggesting prioritized processing of emotional prosodic cues. However, the authors did not find task-driven effects and suggested that the absence of such effects at the electrophysiological level might have been caused by a lack of statistical power to detect (presumably small) task effects, requiring investigation of this issue using larger study samples. Furthermore, a within-subjects manipulation of task demands was employed, which required subjects to switch between an emotional and a linguistic task set. As the authors pointed out, this procedure might have reduced task-driven differences in prosody processing, warranting a more extended investigation of this issue using a between subjects manipulation of task demands.

The dichotic listening (DL) paradigm is particularly suited to study hemispheric specialization in the auditory modality (Greenwald & Jerger, 2003). In the divided-attention version of this paradigm, two different stimuli are presented to each ear. The participant has to divide attention over the auditory channels and react only to the target stimulus presented to one of the ears. It is generally agreed that in the DL-situation the ipsilateral projection of information from the ears to the cerebral hemispheres is inhibited, rendering auditory information from the ears to be primarily available to the contralateral hemisphere (e.g. see Davidson & Hugdahl, 1995). Therefore, if there were hemispheric specialization for the processing of an auditory stimulus, this would be observable as a performance advantage of the ear contralateral to the specialized hemisphere as it has direct access to the specialized processing module (for a discussion of the exact mechanisms behind DL, see Grimshaw, Kwasny, Covell, & Johnson, 2003).

Only one previous study has combined the DL-paradigm with ERPs to study lateralization of emotional prosody perception. Erhan, Borod, Tenke and Bruder (1998) presented participants with dichotic pairs of nonsense syllables, each of which was spoken in one of seven emotional prosodic categories. Participants had to detect a pre-specified emotional prosodic category as fast as possible while the EEG was recorded. Auditory target detection studies like these generally find an initial negativity followed by a positivity for targets as compared with non-targets (e.g., see Fitzgerald & Picton, 1983). In the case of the positivity, it has been demonstrated that its amplitude increases when target probability decreases, while for the earlier negativity this effect is weak at best (Polich & Bondurant, 1996). The positivity can be subdivided in the P3b component and the slow wave, which together have been termed Late Positive Potential or LPP (e.g. see Briggs & Martin, 2009). Erhan et al. (1998) indeed found an N1 and a sustained negativity, followed by a late positivity and a slow wave. Further, at the behavioral level a left ear advantage (LEA) was found for accuracy, in line with a right-hemispheric specialization for emotional prosody. The sustained negativity (300-879 ms post stimulus onset) was identified as a potential electrophysiological marker of the behavioral ear advantage, hypothesized to reflect the emotional categorization process. However, although the relatively late latency of the component might be interpreted to reflect fairly late (and therefore possibly functional) processing, strictly speaking it is unclear whether this lateralized component reflected early acoustic lateralization, more abstract functional hemispheric specialization or both as the function of prosody was not manipulated independently of the acoustics (or vice versa).

In the present ERP study, linguistic and emotional prosodic task demands were manipulated between-subjects with comparably high statistical power while keeping acoustics constant in a divided-attention dichotic auditory target detection. We predicted that, if there is functional lateralization of prosodic perception, there should be a right-ear advantage for the linguistic prosody perception task and a left-ear advantage for the emotional task. At the electrophysiological level, we hypothesized that the distribution of the N2 or LPP over the hemispheres might change correspondingly, demonstrating functional lateralization. Lastly, we considered the possibility that an earlier component (reflecting acoustic processing) might also be sensitive to variation in task demands, reflecting top-down task dependent (and possibly lateralized) modulation of earlier acoustic processing, as has been demonstrated previously (Sussman, Winkler, Huotilainen, Ritter, & Näätänen, 2002).

2. Methods

2.1. Participants

A total of 82 subjects participated in the experiment. Five subjects were excluded because of lower than chance level performance and three due to noisy EEG-data, resulting in a total of 74 participants; 41 (21 male; mean age 23, range 18-37) for the emotional prosodic task and 33 different participants (16 male; mean age 23, range 19-36) for the linguistic prosodic task. The two groups did not differ in male-female ratio or age (for all: one-way ANOVA, $F < 1.01$, $p > 0.5$) All participants were right-handed as assessed with the Edinburgh Handedness Inventory (Oldfield, 1971), were native speakers of Dutch, had normal or corrected-to-normal vision and no (neuro)psychiatric problems in present or past. Participants showing a mean inter-ear hearing threshold difference greater than 12 dB on .5 kHz, 1 kHz, 2 kHz, 4kHz sinusoid tones were excluded from the study. Participants received 20 Euros for their participation in the two hour EEG session. Informed consent was obtained from all participants prior to the experiment. The study was approved by the local ethics committee and conducted in accordance with the Declaration of Helsinki.

2.2. Materials

Pseudowords (see Appendix A) with a bisyllabic structure were generated by randomly combining monosyllables that were composed of random combinations of an initial consonant, a vowel and a final consonant. All pseudowords obeyed Dutch phonotactics and were checked to verify the absence of semantic content. All pseudowords were uttered with angry and sad prosody and with stress on the first and second syllable by a professional actress and recorded at 16 bit resolution and a 44.1 kHz bit sampling rate in a sound-proofed booth. Items were intensity normalized and had a mean duration of 750 msec. In line with previous literature (Cutler, 2005; Scherer, 2003) unstressed syllables differed from stressed syllables primarily in duration while sad and angry prosody additionally differed in F_0, F_0 variability, and variation in intensity (see Table 1). Note that angry and sad items did not show a large intensity difference as the stimuli had been intensity-normalized. To verify the validity of the intended prosodic contrasts, a panel of five healthy volunteers classified each sad and angry prosodic stimulus (in addition to neutral, happy and surprised prosodic stimuli that were not used in the present study) in a forced choice task and rated each item on a five-point typicality scale (1 = very atypical, 5 = very typical). Only pseudowords were selected for which the emotional prosodic contrasts (angry and sad intonation) were classified correctly by at least 4 out of 5 panel members and which had a typicality rating of at least 3.5.

There is evidence that increasing spectral overlap between the target and the competing stimulus increases the suppression of ipsilateral afferent routes from the ears to the cerebral hemispheres, which enhances ear advantages (Della Penna et al., 2007).

Therefore, to maximize spectral overlap, a competing babble stimulus was created by selecting random small segments of speech of the actress and superimposing them onto each other. Dichotic stimuli were created by selecting a random sample of the babble stimulus with the exact same duration of the target stimulus and combining the two (with the target presented in one channel and the babble stimulus in the other channel). This procedure ensured constant and high competition between the target stimulus and the distracter.

Two dichotic target detection tasks were created that presented identical prosodic dichotic targets. From the pool of validated dichotic stimuli, six that had sad prosody and stress on the first syllable served as targets. For the emotional task, subjects were instructed to press as fast as possible when they heard a sad target. For the linguistic task, subjects responded as fast as possible when targets had word initial stress. Hence, identical stimuli served as targets for both tasks while only the task demands (emotional versus linguistic decision) varied. Ten items with angry prosody and stress on the second syllable served as non-targets in both target detection tasks. Additionally, four task-specific (non-target) filler items were presented ten times to each ear but were not analyzed. For the emotional task, angry items with stress on the first syllable were added to prevent subjects from using a linguistic strategy to detect emotional prosody. For the linguistic task, sad items with stress on the second syllable were added to prevent subjects from using an emotional strategy to detect stress position. Each item was presented ten times to each ear, resulting in a total of 120 target and 280 non-target trials per ear. Hence, a target occurred in 30% of the trials for both the emotional and the linguistic DL-task, and target probability was 50% between the two ears. Task-irrelevant prosodic categories (e.g. word-initial versus word-final stress for the emotional task) had a 50% probability of occurring.

Table 1. Acoustic properties of the linguistic and emotional prosodic contrasts (SD in parentheses).

	Stressed syllable	**Unstressed syllable**	**Angry word**	**Sad word**
Mean intensity (dB)	74.22 (2.56)	73.28 (2.41)	72.51 (1.18)	73.94 (1.57)
Mean F_0 (Hz)	239.62 (54.71)	219.95 (36.59)	255.30 (13.81)	202.37 (35.35)
Total duration (s)	0.37 (0.058)	0.30 (0.05)	0.89 (0.057)	0.71 (0.018)
Mean variation (SD) F_0	57.53 (34.29)	31.98 (15.30)	65.88 (9.88)	42.66 (16.47)
Mean variation (SD) intensity	8.70 (2.74)	6.07 (1.65)	13.68 (1.78)	6.67 (1.40)

2.3. Procedure

Each participant completed one of the two dichotic target-detection tasks in a sound-proof booth. Subjects were instructed that they would hear a prosodic stimulus in one ear and people babbling in the other ear. They were told to ignore the babble stimulus and to decide as fast and accurately as possible when they heard the target prosody by pressing the spacebar with the index finger. Response hand was counterbalanced across participants. Participants were instructed that they could respond while the stimulus was still playing (i.e., RTs were recorded from the onset of the stimulus). All instructions were exactly the same for both tasks except for the description of the task-specific prosodic stimulus categories.

Each dichotic listening task started with twelve dichotic practice trials. Participants kept practising these items until a performance level of at least 75% correct was reached. Subsequently, the experimental trials started which encompassed a total of 400 dichotic trials (120 targets, 200 non-targets and 80 fillers). Stimulus presentation order was (pseudo-) random with the restriction that no more than two consecutive presentations of a target were allowed.

An experimental trial started with a black fixation cross that was presented for 1500 ms. Subsequently, a red fixation cross and a binaural warning tone of 500 Hz were presented for 500 ms, after which the dichotic stimulus was presented while the fixation cross remained red. The trial ended 2000 ms after stimulus onset or when participants made a response. Participants were instructed to fix their gaze on the fixation cross throughout the experiment in order to reduce eye movements and not to blink while the fixation cross was red (i.e., when the stimuli were presented).

Stimulus presentation was controlled using E-prime 1.2 and stimulus material was presented at 16 bit resolution and a 44.1 kHz sampling frequency through headphones at a comfortable intensity level.

2.4. EEG Recordings

Electroencephalogram (EEG) was recorded from 64 tin electrodes mounted in an elastic electro-cap organized according to the international 10/20 system. EEG data were recorded with a linked mastoid physical reference and were re-referenced using an average reference. Vertical and horizontal electrooculogram (VEOG and HEOG) were recorded with two pairs of electrodes, one pair placed above and below the left eye, the other pair placed beside the two eyes. The ground electrode was applied to the sternum. Impedance of all electrodes was kept below 5 kΩ for each participant. EEG was continuously recorded with a sampling rate of 500 Hz, amplified, and off-line digitally low-pass filtered with a cut-off frequency of 30 Hz. Participants were seated in front of a monitor at a distance of approximately 50 cm in a dimly lit, electrically shielded and sound-attenuated booth.

2.5. Behavioral analysis

A RM-MANOVA was performed with Ear (left, right) as a within-subject factor, and Task (emotional, linguistic) as between-subject factor. Dependent measures were proportion correct responses and mean reaction time for correct responses.

2.6. ERP-analysis

The EEG data were analyzed with Brain Vision Analyzer (version 1.05). Prior to averaging, trials with eye-movement and blink artefacts were excluded from analysis. Criteria for artefact rejection within an epoch were a maximal voltage step of 50 μV, a maximal difference between two values in a segment of 100 μV, and a minimal and maximal amplitude of -100 μV and 100 μV, respectively.

ERP epochs were computed in a 1000 ms time-window and aligned to a 100 ms pre-stimulus baseline. Individual averages included at least 38 correctly detected target trials per ear. Based on the previous literature on attentive auditory target detection (e.g., see Fitzgerald & Picton, 1983; Nager et al., 2001; Becker & Reinvang, 2007), we expected to find negativities in the N2 range, followed by a late (> 300 ms) positive potential (LPP) including a P3b and a slow wave, when comparing targets to non-targets. Visual inspection of the ERP waves confirmed our expectations and additionally suggested an early negativity in the N1 latency range for targets as compared to non-targets (see Figure 2). Although both peak and mean amplitude analyses were considered, we decided to exclusively report mean amplitude analyses as only the early negativity exhibited a clear peak (as can be observed in Figure 2). For each participant, mean amplitudes for correct target detection were computed for 100-140 ms (early negativity), 180-320 ms (N2) and 350-900 ms (LPP). Because the early negativity exhibited a clear peak, peak latency could be analyzed additionally.

Due to the relatively large sample size of the present investigation, an analysis of the electrophysiological correlate of the ear advantage was possible. Participants were divided into a left ear advantage (LEA) group and a right ear advantage (REA) group, defined as the negative or positive difference between reaction times for correct responses of the left and right ear, respectively. This group analysis served to identify differences in ERP components depending on the laterality of the ear advantage (LEA versus REA).

For all analyses, a set of 36 electrodes was used (see Figure 1). For each hemisphere, the 18 electrodes were divided into six regions of interest (ROIs) comprising three electrodes each. Topographic effects were then analysed using two factors: Laterality (left hemisphere, proximal electrodes: F3, FC3, FC1, C3, C1, FC3, P3, PO3, P1; lateral electrodes: F7, F5, FC5, T7, C5, CP5, P7, P5, PO7; right hemisphere, proximal electrodes: F4, FC4, FC2, C4, CP4, C2, P4, PO4, P2; lateral electrodes: F8, F6, FC6, T8, C6, CP6, P8, P6, PO8), and a factor Longitudinality (left hemisphere, frontal electrodes: F7, F5, FC5, F3, FC3, FC1; central electrodes: T7, C5, CP5, C3, C1, CP3; posterior electrodes: P7, P5, PO7, P3, P1, PO3; right hemisphere, frontal electrodes: F8, F6, FC6, F4, FC4, FC2; central electrodes: T8, C6, CP8, C4, C2, CP4; posterior electrodes: P8, P8, PO8, P4, P2, PO4).

For the main analyses, a multivariate analysis of variance (MANOVA) was conducted using PASW Statistics with Ear (left, right), Hemisphere (left, right), Laterality (proximal, lateral) and Longitudinality (frontal, central, posterior) as within-subject factors and Task (linguistic, emotional) as between-subject factor and mean amplitude as dependent variable (for the N1 component latency was additionally analyzed as dependent variable). An additional MANOVA was performed using exactly the same factors but with the additional factor EA Group to investigate the electrophysiological correlates of the behavioral ear advantage. Greenhouse-Geisser corrected *p*-values are reported.

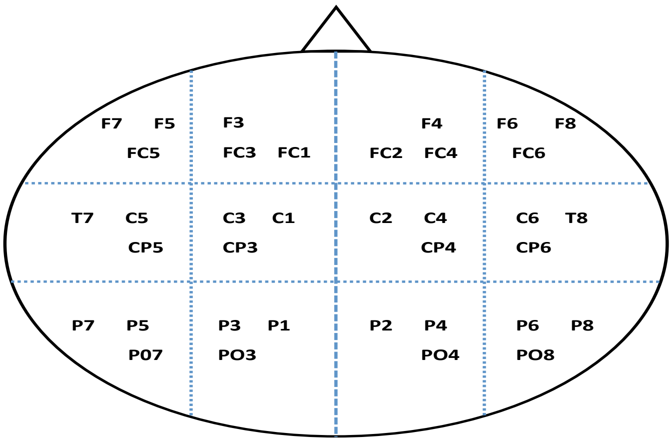

Figure 1. Electrode array used for the experiment. Dashed lines indicate the combination of the levels of the factor Laterality (central, lateral) and Longitudinality (anterior, central, posterior) for each Hemisphere (left, right) .

3. Results

3.1 Behavioral results

The behavioral results can be found in Table 2. Repeated Measures (RM)-MANOVA suggested that the emotional task was easier than the linguistic task as indicated by a main effect of Task ($F(2,71) = 91.54$, $p < 0.001$, $\eta_p^2 = .72$). Follow-up univariate tests confirmed that both reaction time was faster ($F(1,71) = 180.91$, $p < 0.001$, $\eta_p^2 = .72$) and accuracy higher ($F(1,71) = 57.84$, $p < 0.001$, $\eta_p^2 = .45$) for the emotional as compared to the linguistic task.

In sum, the behavioral results suggest that the emotional prosody task was easier than the linguistic prosody task, but no statistically reliable ear advantage was observed.

Table 2. Behavioral results for the experimental conditions (SD in parentheses).

	Linguistic task		**Emotional task**	
	Left ear	Right ear	Left ear	Right ear
Accuracy (% correct)	88 (0.10)	86 (0.11)	99 (0.02)	99 (0.01)
Reaction time (ms)	1016 (178)	1025 (191)	591 (83)	598 (76)

3.2. Electrophysiological results –task effects

The ERP waves of the early negativity, N2 and LPP are visualized in Figure 2. As the present paper is focused on task effects, in the results section below, only task-related effects and the most complete interactions are discussed. For the qualified main effects and simpler interactions and non-task related effects, we refer the reader to the Supplementary information.

3.2.1. Early negativity

The RM-MANOVA for mean amplitude revealed a Hemisphere × Ear × Longitudinality × Task interaction ($F(2,142) = 3.34$, $p < 0.05$, $\eta_p^2 = .05$). Following up with separate ANOVAs for each level of Longitudinality, revealed that at frontal sites, the left ear elicited a larger negativity than the right for the emotional task in both hemispheres while for the linguistic task each ear elicited a larger negativity in the contralateral hemisphere ($F(1,72) = 6.25$, $p < 0.05$, $\eta_p^2 = .08$) while at central ($F(1,72) = 0.2$, $p > 0.05$, $\eta_p^2 = .003$) and posterior sites ($F(1,72) = 0.1$, $p > 0.05$, $\eta_p^2 = .001$) there was no three-way interaction.

An identical MANOVA with latency as dependent variable revealed a Hemisphere × Ear × Task interaction ($F(1,71) = 5.93$ $p < 0.05$, $\eta_p^2 = .08$) with shorter latency in the contralateral hemisphere for each ear for the emotional task, while for the linguistic task latency was shorter in the hemisphere ipsilateral to each ear.

In sum, at frontal sites the left ear elicited a stronger negativity than the right ear in both hemispheres for the emotional task, while for the linguistic task each ear elicited a larger negativity in the contralateral hemisphere. Last, latency was shorter in the hemisphere contralateral to the stimulated ear for the emotional task, while for the linguistic task latency was shorter in the ipsilateral hemisphere.

3.2.2. N2

The MANOVA for mean amplitude revealed a Longitudinality × Task interaction ($F(2,142) = 19.74$, $p < 0.001$, $\eta_p^2 = .22$). Following up with separate ANOVAs for each level of Longitudinality revealed a larger N2 amplitude for the emotional task compared to the linguistic task at frontal sites ($F(1,72) = 13.24$, $p = 0.001$, $\eta_p^2 = .16$), a trend for the emotional task to elicit a greater negativity than the linguistic task at central sites ($F(1,72) = 3.76$, $p = 0.057$, $\eta_p^2 = .05$), and a greater negativity for the linguistic task than the emotional task at posterior sites ($F(1,72) = 6.04$, $p < 0.05$, $\eta_p^2 = .08$)

In sum, the emotional task elicited a larger negativity than the linguistic task at frontocentral sites, while the linguistic task elicited a larger negativity than the emotional task at posterior sites.

3.2.3. LPP

The MANOVA for mean amplitude revealed a Hemisphere × Longitudinality × Task interaction ($F(2,144) = 6.23$, $p < 0.01$, $\eta_p^2 = .08$). Following up the interaction with separate ANOVAs for each level of Longitudinality revealed a significant Hemisphere × Task interaction at posterior sites ($F(1,72) = 8.28$, $p < 0.01$, $\eta_p^2 = .10$) with a clear positivity for the emotional task that was larger for the left than the right hemisphere while for the linguistic task there was a much smaller positivity that was larger for the right than the left hemisphere. For frontal and central sites there was no significant Task × Hemisphere interaction ($p > 0.05$).

Further, there was a Hemisphere × Laterality × Task interaction ($F(2,144) = 7.65$, $p < 0.01$, $\eta_p^2 = .10$) Follow-up ANOVAs for each task revealed that for the emotional task there was a significant Hemisphere × Laterality interaction ($F(1,40) = 13.03$, $p < 0.01$), $\eta_p^2 = .25$) with a positivity for proximal sites that was larger for the right than left hemisphere and a negativity for lateral sites that was larger for the right than the left hemisphere, while for the linguistic task there was no significant interaction ($p > 0.05$).

To summarize, for the emotional task, there was a late positivity that was larger in the left than the right hemisphere at posterior sites and larger in the right than left hemisphere for proximal sites irrespective of intra-hemispheric location, while for the linguistic task there was a highly reduced positivity that was larger in the right than left hemisphere (see Figure 2).

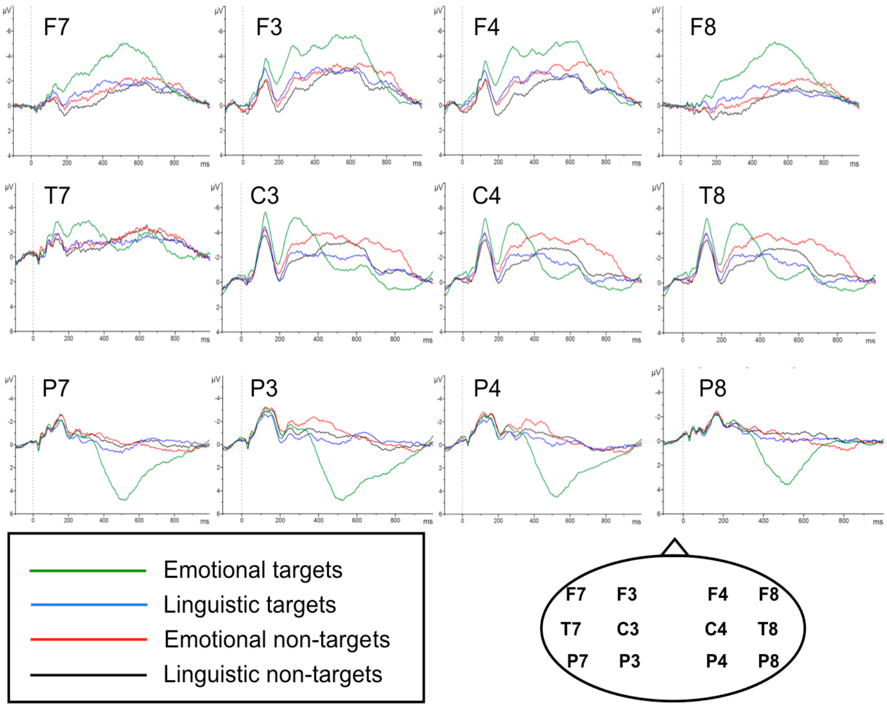

Figure 2. Grand average for the perception of emotional (targets: green, non-targets: red) and linguistic (targets: blue, non-targets: black) prosody for a representative set of electrodes used in the analyses. Averages are shown for a 1000 ms time window post target onset.

3.3. Electrophysiological results - ERP correlates of the ear advantage

The LEA and REA group consisted of 46 and 28 participants respectively. The two groups did not differ in male-female ratio, age or task performance (for all: $p > 0.05$). The groups showed a very large difference in the mean ear advantage $T(1,71) = -10.17$, $p < 0.001$, Cohen's $d = 2.46$) with a very large mean LEA and REA of −34 ms (SD = 26; Cohen's $d = -1.30$) and 33 ms (SD = 30; Cohen's $d = 1.14$), respectively. For both groups, the ear advantage was very robust (paired t-test for both groups; $p < 0.001$). In the section below, only significant interactions with EA-group are discussed (for a complete overview of effects, please see the Supplementary information).

3.3.1. *Early negativity*

No significant main effects or interactions were found for the MANOVA with mean amplitude as dependent variable.

The MANOVA with peak latency as dependent variable revealed a Hemisphere × Laterality × EA-group interaction ($F(1,69) = 4.96$, $p < 0.05$, $\eta_p^2 = .10$). Follow-up MANOVAs for each group showed that for the LEA-group the two-way interaction was non-significant ($p > 0.05$) but for the REA group there was a significant interaction ($F(1,26) = 11.75$, $p < 0.01$, $\eta_p^2 = .31$) with shorter latency for the left hemisphere than the right, but only at lateral sites (see Figure 3).

3.3.2. N2

No main effects or interactions were found.

3.3.3. LPP

No main effects or interactions were found.

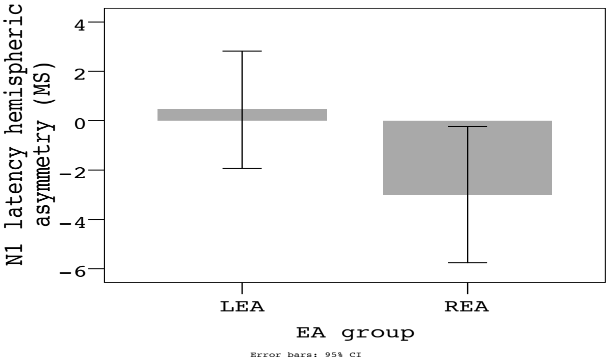

Figure 3. Hemispheric asymmetry of N1 peak latency (left hemisphere peak latency – right hemisphere peak latency) for lateral electrodes for the REA and LEA group.

4. Discussion

The goal of the present investigation was to test whether functional hemispheric specialization for prosody perception could be demonstrated. This was achieved by varying the emotional versus linguistic prosodic processing mode between participants using identical prosodic stimuli in a dichotic listening ERP paradigm. No functional lateralization effects were observed at the behavioral or electrophysiological level. Overall, the emotional task generated a larger response than the linguistic task. Last, the behavioral ear advantage correlated with hemispheric asymmetry of early negativity latency for the REA group. The absence of functional lateralization effects, task effects and the neural correlate of the right ear advantage will be discussed below.

4.1. Functional hemispheric specialization for prosody perception

As outlined in the introduction, recent meta-analytic evidence suggests that there is *relative* right hemispheric specialization for emotional prosody perception while linguistic prosody perception is under bilateral control (Witteman, Van IJzendoorn, Van de Velde, Van Heuven, & Schiller, 2011). Hemispheric specialization on the one hand has been proposed to result from a non-prosody specific right hemispheric advantage in the processing of relevant acoustic cues (i.e., the *cue-dependent* lateralization hypotheses) while the *functional* lateralization hypothesis on the other hand proposes that the left and right hemisphere are specialized in the processing of the *functional categories* of emotion versus linguistic prosody, respectively. The dynamic dual pathway model by Friederici and Alter (2004) further specifies that in the case of linguistic prosody perception, lateralization may depend on the degree to which prosody is bound to segmental structure, with linguistic prosody that is relatively closely bound to segments (such as metical stress, as used in the present investigation) being left-lateralized and linguistic prosody that is not closely bound to the segmental structure (such as boundary marking) being right-lateralized. Note that the functional hypothesis requires that an abstract categorical level of processing has been reached (in order to deduce the functional category of prosodic information) before functional hemispheric specialization can take place.

As outlined in the introduction, recent models propose a three-stage process of (emotional) prosody perception (Kotz & Paulmann, 2011; Brück, Kreifelts, & Wildgruber, 2011), including (1) extraction of acoustic properties in the primary auditory cortex, (2) integration of acoustic properties into a meaningful suprasegmental representation in the associative auditory cortex, and (3) explicit evaluation in frontal cortical areas. It has been suggested that it takes at least 100 ms to reach the first more abstract stage (stage-two) level of processing (Schirmer & Kotz, 2006). Therefore, assuming a strictly serial model, functional lateralization effects are expected to be present in the electrophysiological signal 100 ms after the presentation of a prosodic stimulus. However, it seems likely that the prosody perception process is dynamic, allowing for the possibility of task demands modulating (hemispheric specialization for) earlier stages of prosodic perception (e.g., Brechmann & Scheich, 2005) through top-down modulation.

As both acoustic and functional properties have been hypothesized to influence hemispheric specialization for prosody perception (Pell, 1998) it is necessary to vary the functional task demands of a prosody perception task while keeping acoustics constant and to observe whether there is a shift in hemispheric asymmetry of neural activity to test the functional hemispheric lateralization hypothesis. Therefore, in the present study we instructed one group of participants to categorize the emotional and another group to categorize the linguistic prosodic dimension of the same bi-dimensional prosodic stimuli. Furthermore, to maximize power to detect hemispheric specialization effects, we presented stimuli in a dichotic listening paradigm with a relatively large sample of participants as compared to previous research. Indeed, using G-power (Faul, Erdfelder, Lang, & Buchner, 2007) we calculated that even a small-to-moderate effect of the critical Ear × Task interaction could be detected with sufficient (80%) power in the present study. However, no functional hemispheric specialization effects were found at the behavioral level (as indexed by the ear advantage) or at the electrophysiological level. Two explanations seem plausible for the absence of this effect: (1) functional lateralization effects are of considerable magnitude but the present paradigm was not able to detect the effect of interest; (2) functional lateralization effects are too small to be picked up even by the present (relatively high powered) study. The first explanation seems unlikely as a standard dichotic target detection paradigm was employed in the present investigation that has frequently been used in previous studies. Further, an ERP-waveform was detected that is typically observed in target detection paradigms (initial negativity followed by a positivity) demonstrating the validity of the paradigm and suggesting that the current paradigm should in principle have been able to detect functional lateralization effects.

Concerning the second explanation, an examination of the previous evidence in favor of functional lateralization is required. Only three previous studies had the necessary design to test the functional lateralization hypothesis. First, in a behavioral study, Luks, Nusbaum and Levy (1998) presented 32 subjects with utterances that were pronounced either as a question or a statement in the dichotic listening paradigm. Participants were instructed to categorize the utterances as a question or statement. No ear advantage was found. In a second experiment, 50 subjects had to categorize the same utterances but now emotionally (whether the utterances sounded surprised or neutral). This time a LEA was found. The authors concluded that the ear advantage could be modulated by task demands alone. However, these results are only partially in line with the functional lateralization hypothesis, as the expected REA for the linguistic prosody categorization task was not found. Second, Wildgruber et al. (2004) contrasted discrimination of sentential focus (linguistic prosody perception) to discrimination of expressiveness of the same stimuli (which can be argued to fall in the category of emotional prosody) in a functional imaging study including 10 participants. When contrasting the two tasks directly, a cluster of activity was observed in the left dorsolateral prefrontal cortex (DLPFC) for perception of linguistic prosody while bilateral orbitofrontal activation was noted for the perception of emotional prosody. Although this seems to provide evidence in favor of a greater left- than right-hemispheric contribution to linguistic prosody perception, no formal tests on hemispheric asymmetry were performed, leaving it unclear whether the left DLPFC was indeed significantly more active than its right-hemispheric counterpart. Also, the expected right lateralized activity for the emotional task was not found. Lastly, a recent ERP-study using the cross-splicing paradigm with a sample size of 20 participants also failed to find (functional) hemispheric specialization effects for emotional versus linguistic prosody perception (Paulmann, Jessen, & Kotz, 2012). Therefore, based on the scarce evidence to date, we conclude that the evidence for functional lateralization in prosody perception is weak and future high-powered studies that manipulate the function of prosody independently of acoustics while measuring neural activity over the two hemispheres are required to clarify this issue.

4.2. Task effects

Although no significant interaction between task demands and hemispheric asymmetry was found, emotional task demands enhanced the amplitude of all ERP components across both hemispheres relative to the linguistic task. Two explanations may account for these effects. On the one hand, directing attention to the emotional dimension of the stimuli might have enhanced processing (or the emotional dimension might have interfered with linguistic processing during the linguistic task). Indeed, previous studies have found evidence for enhancement of ERP amplitudes by emotional task demands (e.g., for the LPP see Hajcak et al., 2006; Naumann et al., 1992), interpreted as prioritized processing of emotional information. On the other hand, task differences could be explained by differences in task difficulty, as the emotional task was easier than the linguistic task in the present investigation. Indeed, previous ERP studies have found enhanced amplitudes when difficulty of a task is reduced (e.g., for the LPP, see Molnár, 1999; Gaál, Csuhaj, & Molnár, 2007; but see Combs & Polich, 2006 for conflicting evidence). These two explanations might be non-mutually exclusive as processing of emotional information might be easier than linguistic processing because of its comparatively large salience. Future studies that vary emotional versus non-emotional task demands while controlling for difficulty (or vice versa) are needed to shed light on this issue.

Lastly, early negativity amplitudes were larger across both hemispheres for the left ear than the right for the emotional task, while the reverse was true for the linguistic task. Although it is tempting to interpret this interaction as an electrophysiological correlate of LEA and REA respectively, these effects were unrelated to behavioral EA effects.

4.3. Electrophysiological correlates of the Ear Advantage (EA)

Although we did not find an overall EA for linguistic or emotional prosody perception, the relatively large sample size of the present investigation permitted an additional analysis to explore the electrophysiological correlates of the EA. Participants could be divided into a REA or LEA group independently of whether the emotional or linguistic task was performed, allowing for an analysis of possible underlying electrophysiological determinants of the EA. This analysis revealed that only hemispheric asymmetry in the latency of the early negativity for lateral sites could account for REAs, with REAs associated with shorter latency in the left than right hemisphere. For LEAS, however, such a latency difference was not found.

The 'structural model' of the ear advantage proposed by Kimura (see Della Penna et al., 2007) suggests that EAs arise in the DL-situation because ipsilateral neural routes from the ear to the hemisphere are suppressed. When the specialized hemisphere is ipsilateral to the stimulated ear, the signal has to be transferred over the corpus callosum to reach it and reaction time is delayed. The present results are in line with the structural model as REAs (but not LEAs) were associated with a delayed response of the ipsilateral hemisphere. Our results suggest that REAs can be explained by relatively early specialization at the level of the auditory cortex of the contralateral hemisphere, as the locus of the auditory N1 has been suggested to lay in the auditory cortex (Sandmann et al., 2007). Such early ERP correlates of the EA have been reported before in the literature (Eichele, Nordby, Rimol, & Hugdahl, 2005, but for conflicting evidence see Greenwald & Jerger, 2003). However, other studies have found later ERP components to be correlated with the EA (Ahoniska, Cantell, Tolvanen, & Lyytenen, 1993; Teder, Alho, Reinikainen, & Näätänen, 1993; Erhan, Borod, Tenke, & Bruder, 1998). As suggested by Erhan, Borod, Tenke and Bruder (1996), the N1 asymmetry in favor of the left hemisphere might not be related to prosodic processing but rather to the detection of phonemes. Thus, although speculative, the REAs found in the present study may represent relatively early left hemispheric specialization for phonetic processing during prosody processing.

4.4. Strengths and weaknesses

The present study is the first to systematically test the functional lateralization hypothesis of prosody perception using both behavioral evidence (EAs) and ERPs and with the highest statistical power to date. Despite relatively high statistical power, we failed to find evidence for functional lateralization, which we have argued may reflect the modest magnitude of this effect. The task differences found were confounded with a task difficulty effect, rendering it uncertain whether these differences reflect differences in the linguistic versus emotional mode of processing or differential difficulty of the tasks. Lastly, we included only negative emotions, restricting inferences about emotion effects to negative valence.

4.5. Conclusion

The present investigation did not find evidence for the functional lateralization hypothesis of prosody perception despite relatively high statistical power, suggesting that functional lateralization effects are small. Evidence was found in favor of the idea that REAs can be explained by a speed of processing advantage of the contralateral auditory cortex, in line with the structural model of dichotic listening.

ACKNOWLEDGEMENTS

We thank Jos Pacilly for doing programming work for this study.

5. References

Ahonniska, J., Cantell, M., Tolvanen, A., & Lyytinen, H. (1993). Speech perception and brain laterality: the effect of ear advantage on auditory event-related potentials. *Brain and Language, 45,* 127-146.

Becker, F., & Reinvang, I. (2007). Successful syllable detection in aphasia despite processing impairments as revealed by event-related potentials. *Behavioral and Brain Functions, 3*: 6.

Brechmann, A., & Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cerebral Cortex, 15*, 578-587.

Briggs, K. E., & Martin, F. H. (2009). Affective picture processing and motivational relevance: Arousal and valence effects on ERPs in an oddball task. *International Journal of Psychophysiology, 72*, 299-306.

Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews, 8,* 383-403.

Combs, L. A., & Polich, J. (2006). P3a from auditory white noise stimuli. *Clinical Neurophysiology, 117*, 1106-1112.

Cutler, A. (2005). Lexical stress. In D. B. Pisoni & R. E. Remez (Eds.). *The Handbook of Speech Perception* (264-289). Oxford: Blackwell.

Davidson, R. J., & Hugdahl, E. K. (Eds.). (1995). *Brain Asymmetry.* Cambridge, MA: MIT Press.

Dellla Penna, S., Brancucci, A., Babiloni, C., Franciotti, R., Pizzella, V., Rossi, D… Romani, G. L. (2007). Lateralization of dichotic speech stimuli is based on specific auditory pathway interactions: Neuromagnetic evidence. *Cerebral Cortex, 17*, 2303-2311.

Eichele, T., Nordby, H., Rimol, L. M., & Hugdahl, K. (2005). Asymmetry of evoked potential latency to speech sounds predicts the ear advantage in dichotic listening. *Cognitive Brain Research, 24,* 405-412.

Erhan, H., Borod, J. C., Tenke, C. E., & Bruder, G. E. (1998). Identification of emotion in a dichotic listening task: Event-related brain potential and behavioral findings. *Brain and Cognition, 37*, 286-307.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191

Fitzgerald, P. G., & Picton, T. W. (1983). Event-related potentials recorded during the discrimination of improbable stimuli. *Biological Psychology, 17*, 241-276.

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

Gaál, Z. A., Csuhaj, R., & Molnár, M. (2007). Age-dependent changes of auditory evoked potentials--effect of task difficulty. *Biological Psychology, 76,* 196-208.

Gamble, M. L., & Luck, S. J. (2011). N2ac: An ERP component associated with the focusing of attention within an auditory scene. *Psychophysiology, 48*, 1057-1068.

Greenwald, R. R., & Jerger, J. (2003). Neuroelectric correlates of hemispheric asymmetry: spectral discrimination and stimulus competition. *Journal of the American Academy of Audiology, 14*, 434-443.

Grimshaw, G. M., Kwasny, K. M., Covell, E., & Johnson, R. A. (2003). The dynamic nature of language lateralization: effects of lexical and prosodic factors. *Neuropsychologia, 41*, 1008-1019.

Hajcak, G., Moser, J. S., & Simons, R. F. (2006). Attending to affect: Appraisal strategies modulate the electrocortical response to arousing pictures. *Emotion, 6*, 517-522.

Kotz, S. A., & Paulmann, S. (2011). Emotion, Language, and the Brain. *Language and Linguistics Compass, 5*, 108-125.

Luks, T. L., Nusbaum, H. C., & Levy, J. (1998). Hemispheric involvement in the perception of syntactic prosody is dynamically dependent on task demands. *Brain and Language, 65*, 313-332.

Molnár, M. (1999). The dimensional complexity of the P3 event-related potential: Area-specific and task-dependent features. *Clinical Neurophysiology, 110*, 31-38.

Nager, W., Rosenthal, O., Bohrer, I., Teder-Sälejärvi, W. A., & Münte, T. F. (2001). Human event-related potentials and distraction during selective listening. *Neuroscience Letters, 297*, 1-4.

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.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97-113

Paulmann, S., Jessen, S., & Kotz, S. (2012). It's special the way you say it: An ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia, 50*, 1609-1620.

Pell, M. D. (1998). Recognition of prosody following unilateral brain lesion: Influence of functional and structural attributes of prosodic contours. *Neuropsychologia, 36*, 701-715.

Polich, J., & Bondurant, T. (1996). P300 Sequence effects, probability, and interstimulus interval. *Physiology and Behavior, 61,* 843-849.

Sandmann, P., Eichele, T., Specht, K., Jäncke, L., Rimol, L. M., Nordby, H., & Hugdahl, K. (2007). Hemispheric asymmetries in the processing of temporal acoustic cues in consonant-vowel syllables. *Restorative Neurology and Neuroscience, 25*, 227-240.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*, 227-256.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10*, 24-30.

Sussman, E., Winkler, I., Huotilainen, M., Ritter, W., & Näätänen, R. (2002). Top-down effects can modify the initially stimulus-driven auditory organization. *Cognitive Brain Research, 13*, 393-405.

Teder, W., Alho, K., Reinikainen, K., & Näätänen, R. (1993). Interstimulus interval and the selective-attention effect on auditory ERPs: "N1 enhancement" versus processing negativity. *Psychophysiology, 30*, 71-81.

Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Papers in linguistics: international journal of human communication, 13*, 201-277.

Van Lancker, D., & Sidtis, J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research, 35,* 963-970.

Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., & Ackermann, H. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex, 14*, 1384-1389.

Witteman, J., Van IJzendoorn, M. H., Van de Velde, D., Van Heuven, V. J., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. *Neuropsychologia, 49*, 3722-3738.

Witteman, J. Van Heuven, V. J., & Schiller, N. O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia, 50,* 2752-2763.

Appendix A

Pseudowords used in the experiments:

dundon
kaldun
duldin
kanpal
daldan
kuldul
kuldil
duldun
dalpal
kulpul

Chapter 2.3

A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception

Witteman, J., Van Heuven, V. J., & Schiller, N. O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia, 50,* 2752-2763.

Abstract

With the advent of neuroimaging considerable progress has been made in uncovering the neural network involved in the perception of emotional prosody. However, the exact neuroanatomical underpinnings of the emotional prosody perception process remain unclear. Furthermore, it is unclear what the intrahemispheric basis might be of the relative right-hemispheric specialization for emotional prosody perception that has been found previously in the lesion literature. In an attempt to shed light on these issues, quantitative meta-analyses of the neuroimaging literature were performed to investigate which brain areas are robustly associated with stimulus-driven and task-dependent perception of emotional prosody. Also, lateralization analyses were performed to investigate whether statistically reliable hemispheric specialization across studies can be found in these networks. A bilateral temporofrontal network was found to be implicated in emotional prosody perception, generally supporting previously proposed models of emotional prosody perception. Right-lateralized convergence across studies was found in (early) auditory processing areas, suggesting that the right hemispheric specialization for emotional prosody perception reported previously in the lesion literature might be driven by hemispheric specialization for non-prosody-specific fundamental acoustic dimensions of the speech signal.

1. Introduction

For group-dwelling primates including humans the ability to recognize and strategically react to the emotions of con-specifics is essential as it promotes reproductive success (Silk, 2007). The 'social brain hypothesis' even proposes that selective pressure on such social processing capacities drove the relative expansion of the neocortex in the course of primate evolution (Dunbar, 1998). The importance of the ability to read the (emotional) mind of others in modern human society is exemplified by the devastating effects that disorders of 'theory of mind' processing such as autism spectrum disorders can have on daily functioning of affected individuals (for a review on the evolution and disorders of theory of mind see e.g. Brüne & Brüne-Cohrs, 2006).

How something is being said is an important source of information regarding the emotional state of the speaker. This emotional prosodic (or suprasegmental, i.e., pertaining to a phonological element whose domain is something larger than a single segment, Trask, 1996, p. 343) layer of speech uses a variety of acoustic cues such as intensity, speaking rate, F_0 and spectral balance to signal emotional states (Scherer, 2003). Cross-cultural studies have found better-than-chance decoding accuracy between cultures for basic vocal emotions such as anger and sadness (Scherer, Banse, & Wallbott, 2001; Van Bezooijen, Otto, & Heenan, 1983), suggesting that the human brain might indeed harbor a dedicated emotional voice recognition system that is rooted in our evolutionary history.

Our knowledge of how the brain processes emotional prosody has been greatly advanced with the advent of neuroimaging techniques. Using this technique there generally have been two approaches to delineating the neural network involved in emotional prosody perception. On the one hand, it has been tested which regions in the brain show a stronger response to emotional prosodic categories (such as anger) than to neutral prosody: this network includes areas involved in the extraction of acoustic cues that are important for emotional prosody perception and therefore has been called 'stimulus-driven' (Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009). On the other hand, it has been investigated which brain areas are more active when attention is directed to emotional prosody as compared to direction of attention to a non-emotional prosodic dimension of the same stimulus material (such as the segmental structure or semantics). As this network of regions is sensitive to emotional prosodic task demands independently of acoustics, it has been called 'task-dependent' (Wildgruber et al., 2009). By comparing these stimulus-driven and task-dependent networks, insight can be gained in the regional sensitivity of the brain to earlier and later stages of the emotional prosodic perception process respectively despite the relatively low temporal resolution of functional neuroimaging.

Imaging studies employing stimulus-driven contrasts have generally found a network of areas comprising the superior temporal cortex (Beaucousin et al., 2007; Brück, Kreifelts, Kaza, Lotze, & Wildgruber, 2011; Dietrich, Hertrich, Alter, Ischebeck, & Ackermann, 2008; Ethofer et al., 2007; Ethofer, Van de Ville, Scherer, & Vuilleumier, 2009; Ethofer et al., 2011; Frühholz, Ceravolo, & Grandjean, 2011; Grandjean et al., 2005; Imaizumi et al., 1997; Kotz et al., 2003; Leitman et al., 2010; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Morris, Scott, & Dolan, 1999; Mothes-Lasch, Mentzel, Miltner, & Straube, 2011; Sander et al., 2005; Wiethoff et al., 2007; Wittfoth et al., 2009) and frontal cortical areas (Brück et al., 2011b; Ethofer et al., 2009; Frühholz et al., 2011; Grandjean et al., 2005; Kotz et al., 2003; Leitman et al., 2010; Mitchell et al., 2003; Morris et al., 1999). Additionally, many of these studies find involvement of subcortical structures such as the insula (Ethofer et al., 2009; Leitman et al., 2010; Morris et al., 1999; Mothes-Lasch et al., 2011), basal ganglia (Ethofer et al., 2009; Kotz et al., 2003; Leitman et al., 2010; Morris et al., 1999) and to a lesser extent the amygdala (Ethofer et al., 2009; Mothes-Lasch et al., 2011; Sander et al., 2005; Wiethoff, Wildgruber, Grodd, & Ethofer, 2009).

Studies using task-dependent contrasts tend to find frontal activations (Bach et al., 2008; Brück, Kreifelts, Kaza, Lotze, & Wildgruber, 2011; Buchanan et al., 2000; Ethofer et al., 2006; Ethofer et al., 2009; George et al., 1996; Imaizumi et al., 1997; Mitchell et al., 2003; Wildgruber et al., 2004) in addition to superior temporal activation (Bach et al., 2008; Ethofer et al., 2006; Ethofer et al., 2009; Frühholz, Ceravolo, & Grandjean, 2011; Gandour et al., 2003; Wildgruber et al., 2005). This pattern of results seems to suggest that superior temporal regions are primarily involved in the initial stages of the emotional prosodic perception process, while frontal areas are involved in later stages.

This literature together with electrophysiological and clinical data has inspired two rather similar multi-stage models (Schirmer & Kotz, 2006; Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009) of emotional prosody perception. Both models propose that in an initial phase there is extraction of basic acoustic properties from the speech signal in the primary and secondary auditory cortex. In the model of Wildgruber et al. (2009) the mid-superior temporal gyrus (mid-STG) is explicitly mentioned as an important brain region participating in this phase. Furthermore, both models suggest that, although processing is bilateral in this stage, there might be relative superiority of the right auditory cortex. Both models postulate a second phase in which there is integration of acoustic information into an emotional percept or 'gestalt' (or where meaningful suprasegmental representations of acoustic sequences are realized; Wildgruber et al., 2009) but the models differ in the exact neuroanatomical underpinnings. While the model of Schirmer and Kotz (2006) and its further elaboration by Kotz and Paulmann (2011) suggest that this integration of acoustic cues is achieved by moving anteroventrally along the auditory 'what'-pathway to the anterior superior temporal sulcus (STS), the model by Wildgruber et al. (2009) and its further expansion by Brück et al. (2011a) proposes that this stage is supported by the posterior superior temporal cortex. Both models suggest right-hemispheric specialization in this second phase. Finally, both models propose a third phase in which emotional prosody is explicitly evaluated and integrated with other cognitive processes. While the model by Brück et al. (2011a) assumes that both (left and right) inferior frontal gyri (IFG) have a similar evaluative function in this phase, the Kotz and Paulmann (2011) model further specifies that the right IFG and orbitofrontal cortex (OFC) are involved in the explicit evaluation of emotional prosody and that the left IFG is responsible for integrating emotional prosodic information with other layers of the speech signal (such as semantics). Concerning the role of subcortical structures, the model of Kotz and Paulmann (2011) proposes that the amygdala and basal ganglia (BG) are important for the detection of emotional salience and evaluation (stage 3) of emotion prosody respectively. The model by Brück et al. (2011a) proposes that the amygdala plays a role in implicit emotional prosodic perception but leaves the role of the BG unspecified.

A longstanding question in the neuropsychological literature is whether there is hemispheric specialization within this emotional prosodic perception network. An important source of evidence regarding hemispheric specialization is provided by the lesion literature as it gives information on the necessity of cerebral structures for emotional prosody perception. A recent meta-analysis of the lesion literature (Witteman, Van IJzendoorn, Van de Velde, Van Heuven, & Schiller, 2011) indeed showed that while damage to both cerebral hemispheres degrades emotional prosody perception performance, damage to the right hemisphere compromised performance more, suggesting that there is *relative* right-hemispheric specialization for emotional prosody perception. However, it remains unclear through which mechanism such right hemispheric specialization for emotional prosodic perception might emerge: hemispheric specialization might emerge in any of the three stages proposed by the two emotional prosody perception models.

Indeed, on the one hand the *cue-dependent lateralization hypotheses* propose that the right-hemisphere specialization for emotional prosody perception can be traced back to a (non-prosody specific) advantage of the right hemisphere for early acoustic processing such as spectral processing (Van Lancker & Sidtis, 1992) as spectral parameters are important for decoding of emotional prosody (Scherer, 2003). The idea that there is hemispheric specialization for processing of elementary dimensions of speech such as temporal vs. spectral information (Zatorre & Belin, 2001) or integration of the speech signal over smaller vs. longer time windows (Poeppel, 2003) in the left and right hemisphere respectively has recently indeed been gaining support (see e.g. Boemio, Fromm, Braun, & Poeppel, 2005; Jamison, Watkins, Bishop, & Matthews, 2006). Both proposed models of emotional prosody perception (Brück, Kreifelts, & Wildgruber, 2011a; Kotz and Paulmann, 2011) indeed assume relative right hemispheric specialization for the first (acoustic) stage of emotional prosody perception in the auditory cortex.

On the other hand, the *functional lateralization hypothesis* proposes that the right hemisphere is specialized for the processing of the emotional communicative *function* of emotional prosody (Van Lancker, 1980). This proposal might correspond to the more abstract processing of emotional prosody in stage 2 (integration of acoustic cues into an emotional percept) or stage 3 (explicit evaluation of emotional prosody) in the models of emotional prosody perception described above. A closer look at the proposed models of emotional prosody perception reveals that while both models assume right-hemispheric specialization for stage 2 emotional prosody perception, the Kotz and Paulmann (2011) model also allows for right-hemispheric specialization in the last evaluative stage while the Brück et al. (2011a) model assumes bilateral processing. As a consequence of the multi-stage nature of emotional prosodic perception, right-hemispheric specialization for the whole *process* of emotional prosody perception (as frequently assessed through behavioral tasks in lesion studies) might thus depend on right-hemispheric specialization for just one stage or could be additive or even interactive (for instance, there might be top-down modulated right-hemispheric specialization for acoustic processing based on task demands) across multiple stages.

In sum, although considerable progress has been made in delineating the neural network involved in emotional prosody perception some questions remain unresolved. First, recently proposed models disagree on the exact neuroanatomical substrate of the proposed second stage (integration of acoustic cues into an emotional percept or representation of meaningful suprasegmental acoustic sequences) of emotional prosody perception. Second, is it unclear where in the process of emotional prosody perception the right-hemispheric specialization arises that has been found in the lesion literature.

The exact pattern of results found in fMRI studies is dependent on many variables including the scanner hardware, scanning parameters, exact task instructions and analysis methods and hence results may vary considerably between studies (Bennett & Miller, 2010). Therefore, there seems to be merit in determining which cerebral structures are reliably activated across studies that employ a variety of slightly different tasks to measure the same hypothesized underlying cognitive process. This will provide information about which neural network is generally involved (without study-specific idiosyncratic effects) in the hypothesized cognitive process. To the best of our knowledge, no such quantitative meta-analysis of the neuroimaging literature of emotional prosodic perception has as yet been undertaken. Therefore, in the present investigation activation likelihood estimation (ALE) meta-analysis, a commonly used and valid coordinate based voxel-wise quantitative meta-analysis method for neuroimaging data (Eickhoff et al., 2009; Laird et al., 2005; Turkeltaub et al., 2011; Turkeltaub, Eden, Jones, & Zeffiro, 2002), was used to address following questions:

(1) Which neural networks are reliably involved in stimulus-driven and task-dependent perception of emotional prosody?
(2) Can statistically robust lateralization of activation probability be demonstrated for these networks?

Two ALE meta-analyses were performed in an attempt to answer these questions: one on stimulus-driven processing of emotional prosody and another on task-driven processing of emotional prosody. It was hypothesized that these meta-analyses would yield the bilateral temporofrontal network with subcortical involvement that has been proposed by recent models of emotional prosody perception. Additionally, for each of the two meta-analyses subtraction analyses were performed to investigate whether there was statistically reliable lateralization of ALE clusters. It was hypothesized that there would be multiple lateralized clusters, with clusters in early auditory processing areas (reflecting cue-dependent lateralization) and in multimodal temporal or frontal areas (possibly corresponding to functional lateralization).

2. Methods

2.1. Study and experiment selection

The PubMed database (www.ncbi.nlm.nih.gov/pubmed/) was searched for relevant publications until September 2011 using the following search string for the title or the abstract (where the asterisk denotes a wildcard): (brain OR neural) AND emotion* AND (prosod* OR vocal OR voice). Additionally, the reference lists of articles and recent reviews (Brück et al., 2011a; Kotz & Paulmann, 2011) were checked for potentially relevant publications. The following inclusion criteria were applied: (1) an emotional prosodic perception task was employed (i.e. studies using nonverbal vocal emotional sounds such as laughter were excluded) (2), emotional prosody was the only relevant independent variable that was varied (i.e. studies contrasting emotional prosody that was congruent with emotional lexical semantics to neutral prosody in words with neutral lexical semantics were excluded, since emotional semantics and emotional prosody were confounded in that case), (3) subjects were healthy adult participants, (4) whole brain analyses were reported (ROI analyses were excluded), (5) peak coordinates of significantly activated areas were reported in the MNI (Evans et al.,

1993) or Talairach (Talairach & Tournoux, 1988) standard stereotactic space, and (6) the article was published in a peer-reviewed international journal.

Experiments (contrasts) were selected from each paper as follows. For the meta-analysis on the stimulus-driven network of prosodic perception, the active condition had to be emotional prosody (i.e. emotional expression superimposed on an utterance containing speech segments such as voiced nasals, vowels, words or pseudowords) and the control condition either neutral prosody or speech material devoid of emotional prosody (synthesized speech). For the meta-analysis on the task-dependent network of emotional prosodic perception, attention had to be directed to the emotional prosody in the active condition, while in the control condition the same material had to be presented without attention directed at the emotional prosody. If a study reported multiple experiments that were eligible for inclusion in one of the two meta-analyses, the contrast was chosen that best answered the question "which network in the brain is involved in the processing of emotional prosody?" (e.g. when foci were presented for individual emotions separately and for all emotions combined, the latter experiment was included as it is more likely that it engages the full emotional prosodic perception network) to prevent one study from driving the meta-analysis.

All included studies used fMRI except George et al. (1996) and Imaizumi et al. (1997), which used PET. As can be observed in Table 1, a total of 16 studies (N = 296 subjects; 93 foci) was included in the meta-analysis on the stimulus-driven network of emotional prosodic perception. All of the included studies compared speech with emotional prosody to speech with neutral prosody or to synthesized speech without prosody.

Table 2 lists the 11 studies (N = 153 subjects; 50 foci) that were included in the meta-analysis on the task-driven network of emotional prosodic perception. All the included studies compared emotional prosody processing to the processing of a different dimension (such as speaker gender or lexical semantics) of the same speech stimuli. Note that Ethofer et al. (2006) was excluded from the task-driven analysis, as a predefined ROI was used in the analyses (i.e. no whole-brain corrected coordinates were reported).

Table 1. Overview of studies included in the meta analysis on stimulus-driven effects.

Study	**N (males)**	**Active condition**	**Control condition**	**Emotions**	**Task(s)**	**N foci**
Bach 2008	16 (8)	Emotional	Neutral	Anger, fear, neutral	• Categorize gender • Categorize emotional prosody	9
Beacousin 2007	23 (11)	Emotional	No prosody (text to Speech)	Anger, sadness, joy, no emotional intonation	• Categorize emotional semantics	10
Brück 2011b	24 (12)	Emotional	Neutral	Anger, joy, neutral	• Identify emotional prosody • Identify vowel • Identify word	4
Ethofer 2007	24 (12)	Emotional	Neutral	Anger, fear, joy, eroticism, neutral	• Passive listening	6
Ethofer 2009	24 (12)	Anger	Neutral	Anger, neutral	• Categorize emotional prosody • Categorize word class	7
Ethofer 2011	22 (13)	Emotional	Neutral	Anger, joy, sadness, relief, neutral	• Categorize gender	2
Frühholz 2011	17 (3)	Anger	Neutral	Anger, neutral	• Discriminate emotional prosody • Discriminate gender	13
Grandjean 2005	15 (8)	Anger	Neutral	Anger, neutral	• Categorize gender	8
Kotz 2003	12 (4)	Anger	Neutral	Joy, anger, neutral	• Categorize emotional prosody	12
Leitman 2010	19 (19)	Emotional	Neutral	Joy, anger, fear, neutral	• Categorize emotional prosody	4
Mitchell 2003	13 (13)	Emotional	Neutral	Joy, sadness, neutral	• Passive listening	6
Mothes-Lasch 2011	24 (16)	Anger	Neutral	Anger, neutral	• Categorize gender	1
Rota 2008	10 (10)	Emotional	Neutral	Joy, sadness, anger, neutral	• Categorize emotional prosody	0
Sander 2005	15 (8)	Anger	Neutral	Anger, neutral	• Categorize gender	8
Schirmer 2008	14 (14)	Anger	Neutral	Anger, neutral	• Watch movie and listen passively	0
Wiethoff 2007	24 (12)	Emotional	Neutral	Joy, eroticism, anger, fear, sadess, disgust, neutral	• Passive listening	3
Total	**296**					**93**

Table 2. Overview of studies included in the meta analysis on task driven effects.

Study	**N** (male)	**Active condition**	**Control condition**	**Emotion**	**Tasks**	**N foci**
Bach 2008	16 (8)	Categorize prosody	Categorize gender	Anger, fear, neutral.	• Categorize gender • Categorize emotional prosody	11
Brück 2011	24 (12)	Identify prosody	Identify vowel + Identify words	Anger, joy, neutral	• Identify emotional prosody • Identify vowel • Identify word	1
Buchanan 2000	10 (10)	Detect emotional prosody	Detect phoneme	Anger, joy, sad, neutral	• Detect joy • Detect sadness • Detect phoneme	3
Ethofer 2009	24 (12)	Categorize emotional prosody	Categorize word class	Anger, neutral	• Categorize emotional prosody • Categorize word class	4
Frühholz 2011	17 (3)	Discriminate emotional prosody	Discriminate gender	Neutral, Anger	• Discriminate emotional prosody • Discriminate gender	3
Gandour 2003	10 (5)	Discriminate emotional prosody	Passive listening	Anger, joy, sadness	• Discriminate emotional prosody • Passive listening	7
George 1996	13 (8)	Categorize emotional prosody of sentence	Categorize word	Joy, sadness, anger, neutral.	• Categorize emotional prosody • Categorize word	1
Imaizumi 1997	6 (6)	Categorize emotional prosody of words	Categorize speaker identity	Surprise, disgust, joy anger	• Categorize emotional prosody • Categorize speaker identity	13
Mitchell 2003	13 (13)	Attend prosody	Attend semantics	Joy, sadness, neutral	• Attend prosody • Attend semantics	3
Wildgruber 2004	10 (6)	Discriminate emotional prosody	Discriminate linguistic prosody	Emotional expressiveness	• Discriminate emotional prosody • Discriminate linguistic prosody	2
Wildgruber 2005	10 (5)	Categorize emotional prosody	Categorize vowels	Joy, anger, fear, sadness, disgust	• Categorize emotional prosody • Categorize vowels	2
Total	**153**					**50**

2.2. ALE meta-analyses

Activation Likelihood Estimation (ALE) is a widely used and valid coordinate-based meta-analysis procedure that tests for consistent brain activation across studies for a given set of (similar) tasks (Laird et al., 2005; Turkeltaub, Eden, Jones, & Zeffiro, 2002). In brief, ALE meta-analysis uses the peak-coordinates of activation clusters reported in the previous literature and models the location uncertainty of these coordinates as 3D-Gaussian probability density distributions. The voxelwise convergence of these distributions is then tested against a null-distribution and corrected for multiple comparisons to test which voxels show above-chance convergence across studies. As a last step cluster analysis is performed with a user-specified minimum cluster volume. Recently, the ALE-algorithm has been further optimized by adopting a random-effects approach (allowing for generalization of the results beyond the experiments analyzed) and more precisely modeling location uncertainty (Eickhoff et al., 2009) and by minimizing the cumulative effects of multiple neighboring within experiment foci (Turkeltaub et al., 2011). All ALE-analyses were performed using the latest version of the ALE-algorithm (Turkeltaub et al., 2011) at the time of writing, as implemented in GingerALE 2.1.1 (www.brainmap.org/ale).

As discussed in the Introduction, an important unresolved question in the literature is whether there is hemispheric specialization for emotional prosodic perception. Therefore, we additionally applied a lateralization analysis as reported previously by Turkeltaub and Coslett (2010) for each of the two meta-analyses. To formally test for significant lateralization of ALE-values, we multiplied x-coordinates of the input foci by −1 (effectively left-right flipping each coordinate) and performed a subtraction analysis where we subtracted the flipped foci from the original (unflipped) foci.

All meta-analyses were performed in Talairach space. When foci were reported in MNI space, they were converted to Talairach space using the Lancaster transform (Lancaster et al., 2007) as implemented in the GingerALE software. For the two main meta-analyses on stimulus and task-driven emotional prosody perception, above-chance voxel wise convergence was tested corrected for multiple comparisons using the false discovery rate (FDR) approach (Genovese, Lazar, & Nichols, 2002) as implemented in GingerALE with $q = 0.05$ and a minimum cluster extent of 100 mm^3. These FDR corrected images were subsequently used as input images for the subtraction analyses to test for lateralization effects. At present there is no established method to correct ALE subtraction maps for multiple comparisons (Eickhoff et al., 2011, p. 941). Further, for smaller data-sets with 10-20 experiments as the present, an uncorrected threshold with a small extent threshold has been suggested to provide sufficient protection against false positive ALE-clusters (Gobras, Beaton, & Eickhoff, 2012, p. 439). Therefore subtraction results are reported with a 100 mm^3 extent threshold in combination with an uncorrected threshold at a more lenient $p = 0.05$ and a more conservative $p = 0.001$ level for ALE-values. Statistical maps were overlaid onto the Talairach anatomical template and visualized using Mango software (http://ric.uthscsa.edu/mango/). Additionally, similarly to previous studies (Turkeltaub & Coslett, 2010; Van der Laan, De Ridder, Viergever, & Smeets, 2011) we applied one further conservativeness criterion. To avoid discussing clusters to which only one or two studies directly contributed, at least three of the included studies had to contribute directly to a significant ALE-cluster in order to be discussed in the discussion-section (all significant clusters are reported in the results section).

3. Results

3.1. Main analyses

3.1.1. Stimulus-driven network

The ALE-analysis of experiments contrasting emotional prosody to neutral prosody or speech without prosody revealed 15 clusters of above chance convergence (see Table 3 and Figure 1). Five clusters met our additional criterion of having at least three directly contributing studies. The most concurrent ALE-cluster lay in the right mid superior temporal gyrus and continued medially into the transverse temporal gyrus (peak Talairach coordinates (44, −24, 8), ALE-value = $2.67 \bullet 10^{-3}$, volume = 2696 mm^3). In the left temporal cortex two significant ALE-clusters were found. The cluster with the highest convergence was located in the mid-STG (peak Talairach coordinates (−58, −22, 2), ALE-value = $1.57 \bullet 10^{-3}$, volume = 2696 mm^3), while the other ALE-cluster was located in the posterior STG (peak Talairach coordinates (−42, −34, 10), ALE-value = $1.52 \bullet 10^{-3}$, volume = 848 mm^3). Two ALE-clusters were located in the inferior frontal gyri. One was located in the lateral left IFG (peak Talairach coordinates (−54, 20, 2), ALE-value = $1.23 \bullet 10^{-3}$, volume = 680 mm^3). The other cluster was located in the right IFG ALE-cluster (peak Talairach coordinates (48, 28, 6), ALE-value = $1.23 \bullet 10^{-3}$, volume = 976 mm^3) and extended into the dorsal anterior insula.

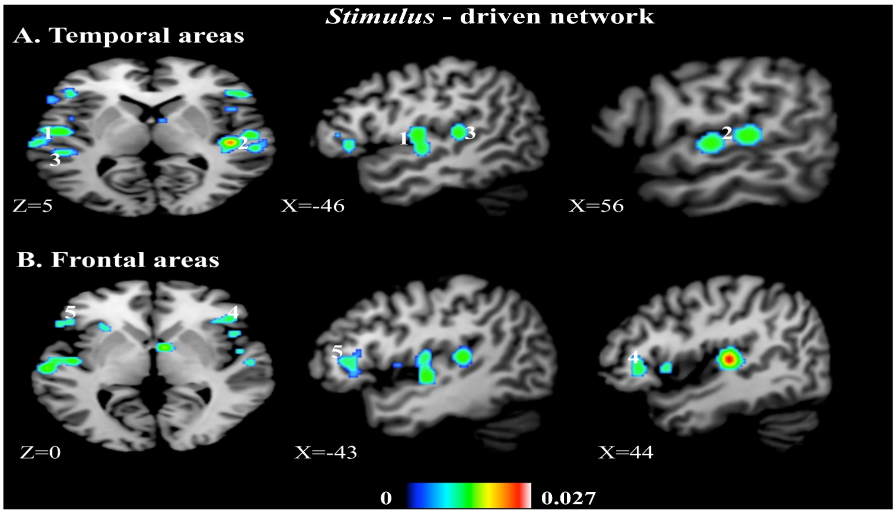

Figure 1. Results of the meta-analysis of studies contrasting emotional prosody to neutral prosody or synthesized speech. All ALE-clusters were corrected for multiple comparisons at an FDR of 0.05 and had a minimum cluster extent of 100 mm^3. For display purposes the cluster-uncorrected statistical maps are shown. Clusters to which at least three studies have directly contributed are marked with a number that corresponds to the cluster number in Table 3. In the top-panel A, 1 = left STG, 2 = right STG and 3 = left STG. In the bottom-panel B, 4 = right IFG, 5 = left IFG.

Table 3. Locations (Talairach space) of significant ALE-clusters.

Clu ster	**Cluster size (mm^3)**	**Anatomical Label**	**Peak voxel coordinates**			**ALE-value ($\bullet 10^{-3}$)**	**No. stu dies**
			x	y	z		
***Stimulus*-driven analysis:**							
1	2696	L Superior Temporal Gyrus	−58	−22	2	1.57	7
		L Superior Temporal Gyrus	−48	−12	6	1.56	
		L Insula	−42	−14	−4	1.48	
2	2696	R Superior Temporal Gyrus	44	−24	8	2.67	7
		R Superior Temporal Gyrus	56	−30	8	1.55	
		R Superior Temporal Gyrus	54	−16	4	1.54	
3	848	L Superior Temporal Gyrus	−42	−34	10	1.52	3
4	808	R Inferior Frontal Gyrus	48	28	6	1.23	3
		R Inferior Frontal Gyrus	44	28	2	1.18	
		R Inferior Frontal Gyrus	36	24	0	0.91	
5	680	L Inferior Frontal Gyrus	−54	20	2	1.23	3
		L Inferior Frontal Gyrus	−42	28	6	1.08	
		L Inferior Frontal Gyrus	−46	24	0	1.02	
6	528	L Inferior Frontal Gyrus	−34	6	−12	1.74	2
7	504	R Medial Globus Pallidus	8	−2	−2	1.66	2
8	400	R Medial Frontal Gyrus	16	40	14	1.52	2
9	400	L Superior Occipital Gyrus	−32	−72	30	1.52	2
10	392	R Middle Occipital Gyrus	24	−86	18	1.48	2
11	392	L Inferior Frontal Gyrus	−52	20	24	1.47	2
12	224	R Insula	46	12	2	1.05	2
13	192	R Superior Temporal Gyrus	50	−4	−4	1.02	2
14	152	R Subcallosal Gyrus	26	4	−10	0.91	2
		R Putamen	26	−2	−6	0.89	2
15	152	L Claustrum	−26	18	−2	0.99	2
***Task*-driven analysis:**							
1	880	R Superior Temporal Gyrus	46	−34	6	1.35	3
2	512	R Inferior Frontal Gyrus	40	34	2	0.93	3
		R Inferior Frontal Gyrus	36	34	−6	0.81	
3	184	L Middle Frontal Gyrus	−42	34	0	0.90	1
4	168	R Superior Frontal Gyrus	10	20	50	0.86	1
5	128	R Middle Frontal Gyrus	48	26	24	0.85	1

L = Left hemisphere; *R* = Right hemisphere. All reported ALE-clusters were thresholded at $q < 0.05$ corrected and an extent threshold of 100 mm^3

3.1.2. Task-driven network

The ALE-analysis of experiments contrasting processing of emotional prosody to processing a different dimension of the same material revealed 5 clusters of above chance convergence (see Table 3 and Figure 2). Two of these clusters fulfilled our additional criterion of having at least three directly contributing studies. The cluster of highest convergence lay in the right posterior STG (peak Talairach coordinates (46, −34, 6), ALE-value = $1.35 \cdot 10^{-3}$, volume = 880 mm^3). The second cluster was located in the right lateral IFG (peak Talairach coordinates (40, 34, 2), ALE-value = $.93 \cdot 10^{-3}$, volume = 416 mm^3). Although significant ALE-voxels were also noted in the left IFG, these voxels did not survive the additional criterion of a minimum cluster volume of 100 mm^3.

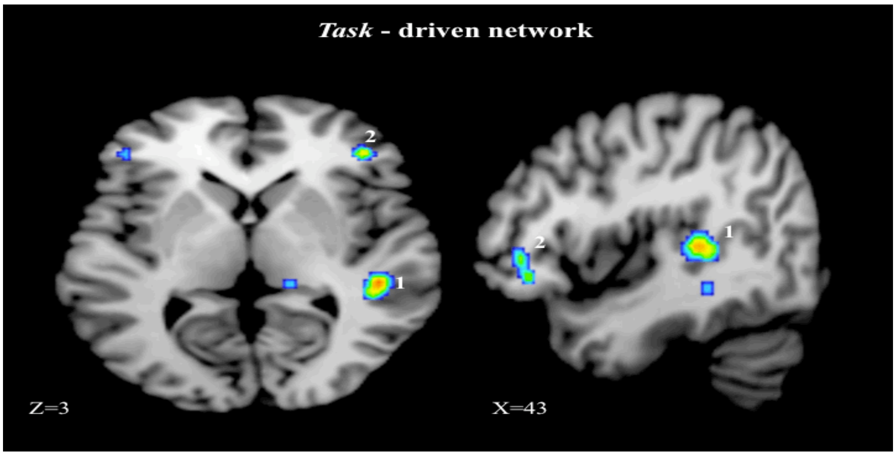

Figure. 2. Results of the meta-analysis of studies contrasting processing of emotional prosody to processing of a different dimension of the same stimulus material. All ALE-clusters were corrected for multiple comparisons at a FDR of 0.05 and had a minimum cluster extent of 100 mm^3. For display purposes the cluster-uncorrected statistical maps are shown. Clusters to which at least three studies have directly contributed are marked with a number that corresponds to the cluster number in Table 4. 1 = right posterior STG, 2 = right IFG.

3.2. Lateralization analyses

An important issue in the literature is whether there is hemispheric specialization for emotional prosodic perception. For the stimulus-driven analysis, convergence seemed to be higher in the right temporal lobe than the left. For the task-driven analysis, both significant ALE-clusters were right lateralized. These qualitative impressions were followed up with a formal test of lateralization of ALE-clusters by subtracting ALE-maps that were flipped around the x-axis (i.e. right/left flipping the ALE-map) from the original ALE-maps. One weakness of the ALE meta-analysis is that activation magnitude (effect size) and spatial extent of clusters are not considered. Therefore, similarly to Turkeltaub and Coslett (2010), we checked all included studies for whether the effect size (Z or T-value) and spatial extent of peak (maximally active) clusters in the vicinity of lateralized ALE clusters was larger in the left or right hemisphere. To test whether lateralization of cluster activation magnitude and extent was statistically significant between the hemispheres, a binomial test was performed.

3.2.1. Stimulus-driven network

Table 4 lists the two ALE-clusters that showed significant lateralization in the subtraction analysis at $p = 0.05$ uncorrected for the stimulus-driven contrasts. As can be seen in Figure 3, the only ALE cluster that fulfilled our additional criterion of having at least three directly contributing studies was located in the right transverse temporal gyrus (peak Talairach coordinates (50,−24,10), $Z = 1.95$, volume = 624 mm^3). No ALE-cluster survived the more conservative $p = 0.001$ threshold. Although peak activation of temporal lobe clusters tended to be greater in the right than left hemisphere, this effect was not statistically reliable (P(binomial) = 0.27). However, there was a trend for peak cluster extent to be larger in the right than left temporal lobe (P(binomial) = 0.065).

3.2.2. Task-driven network

As can be observed in Table 4, the lateralization analysis on the task-driven network of emotional prosodic perception revealed only one significantly lateralized ALE-cluster that was located in the posterior right STG (peak Talairach coordinates (44.25, −38.25, 8.75), $Z = 2.14$, volume = 880 mm^3) at $p = 0.05$ uncorrected. No ALE-cluster survived the more conservative $p = 0.001$ threshold. For all included studies reporting clusters in the temporal lobe, peak activation and extent were greater in the right than the left hemisphere (P(binomial) < 0.05).

Table 4. Locations of significant clusters for the *lateralization* analyses.

Cluster	Cluster size (mm^3)	Anatomical Label	Peak voxel coordinates			*Z*-value	No. studies
			x	y	z		
***Stimulus*-driven lateralization analysis:**							
1	624	R Transverse Temporal Gyrus	50	−24	10	1.95	3
		R STG	42	−26	4	1.87	
		R Transverse Temporal Gyrus	40	−24	12	1.82	
		R STG	38	−26	6	1.80	
2	192	L Insula	−42	−14	4	2.01	0
***Task*-driven lateralization analysis:**							
1	880	R STG	44.25	−38.25	8.75	2.13	3

L = left hemisphere; *R*=Right hemisphere. All reported ALE-clusters were thresholded at $p < 0.05$ uncorrected and a cluster extent of 100 mm^3

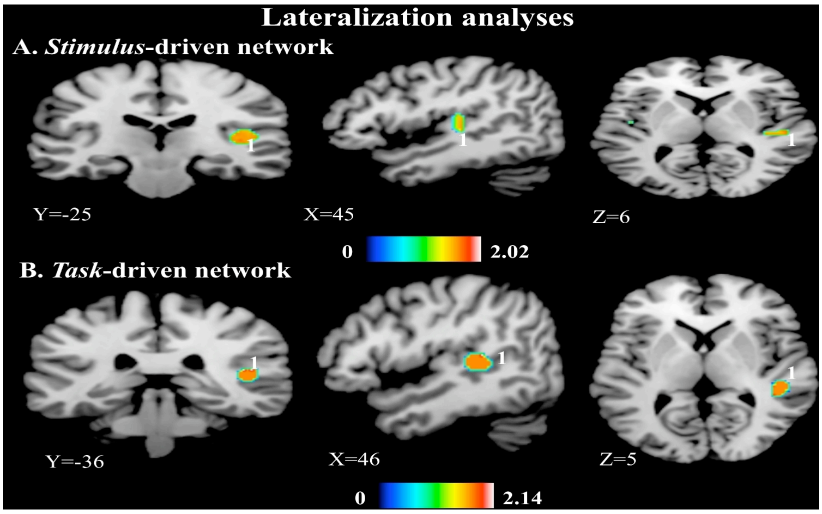

Figure 3. Results of the lateralization analyses. All ALE-clusters were significant at p = 0.05 uncorrected and had a minimum cluster size of 100 mm^3. Clusters to which at least three studies have directly contributed are marked with a number which corresponds to the cluster number in Table 4.

4. Discussion

Using activation-likelihood estimation (ALE) meta-analysis, we investigated which neural networks are reliably engaged in stimulus-driven and task-driven emotional prosody perception and to which extent these networks show hemispheric asymmetry. Although all reported clusters in the main analyses passed the criterion of a less than 5% family-wise probability that the convergence found could have occurred by chance, the maximum percentage of experiments directly contributing to a cluster was 41. Exactly the same maximum percentage of contributing experiments was found in a recent ALE-meta analysis of visual food-cue processing, which the authors qualified as a 'moderate' level of convergence as compared to other ALE-meta analyses (Van der Laan, De Ridder, Viergever, & Smeets, 2011). As these authors pointed out, a recent review reported that the test-retest reliability of fMRI results as expressed by the intraclass correlation coefficient is 0.5 (Bennett & Miller, 2010) which implies an even lower correspondence of results across *different* experiments (tasks) that intend to measure the same neurocognitive process (as has been done in the present investigation). Hence, a moderate degree of convergence across experiments as found in the present meta-analysis could have been expected.

In addition to heterogeneity induced by non-prosody specific variables that differed between included studies such as scanning parameters (e.g. whether measures were taken to reduce echo planar imaging noise) and analysis details, there was variation in factors specific to the subject at hand that might have reduced convergence. For instance, in the case of stimulus-driven processing it can be observed in Table 1 that there is variation in the exact task instruction, varying from passive listening to discrimination and categorization. These different task instructions might engage the different components of the stimulus-driven network, lowering overall convergence. This also holds for the task-driven analysis: as can bee seen in Table 2 there was variation in the exact task instructions for both the active and control conditions, likely inducing variation in the exact locations of activation clusters found across studies. Another source of variation is whether stimulus-driven processing was implicit or explicit: the few studies that have directly compared implicit to explicit stimulus-driven emotional prosody processing, suggest that these two modes of processing might tax different components of the emotional prosody perception network (Bach et al., 2008; Frühholz, Ceravolo, & Grandjean, 2011). A third source of variation might be the number and kind of emotions included in the study, which varied from only angry and neutral to seven different emotional categories, as can been seen in Table 1. As there is evidence for emotion specific processing in the auditory cortex (Ethofer, Van de Ville, Scherer, & Vuilleumier, 2009) and since variation in the number of emotions in categorization tasks might induce variation in the degree to which the working memory system is taxed (Hoekert, Kahn, Pijnenborg, & Aleman, 2007) this factor might have compromised convergence of activation clusters across studies.

Unfortunately, the number of neuroimaging studies on emotional prosody perception published to date was too low to permit a reliable formal meta-analytic contrast of these moderating variables. In order to only discuss the part of the emotional prosody perception network that converges robustly across studies (and therefore represents the general network relatively independent of differential engagement of specific neural structures associated with the mentioned moderators), in the following only ALE-clusters will be addressed to which at least three of the included experiments directly contributed.

4.1. Stimulus-driven network

Contrasting the perception of emotional prosodic categories to neutral prosody (or speech devoid of prosody) engages areas involved in the processing of acoustic properties that differ between emotional and non-emotional prosody but likely also engages systems involved in more abstract processing of emotional information.

The clusters of highest convergence for the stimulus-driven contrast lay in the bilateral mid-STG, extending posteriorly and laterally with respect to HG. In the right hemisphere the STG cluster also extended medially into the transverse temporal (Heschl's) gyrus proper, which likely houses the human primary auditory cortex (Galaburda & Sanides, 1980). Indeed, as the mid-STG is located within the zone of purportedly speech-specific superior temporal cortex (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000) and shows sensitivity to the emotional quality of speech, the mid-STG (and its posterolateral extension) has recently been coined the 'emotional voice area (EVA)' (Ethofer et al., 2011). Previous studies have shown that the mid-STG is sensitive to multiple acoustic properties relevant for emotional prosody perception (Wiethoff et al., 2007) and is active independently of task demands (Ethofer et al., 2009). This suggests that the mid-STG (and its posterolateral extension) is involved in the automatic integration of emotional prosodic acoustic cues relevant for the perception of emotional prosody.

In both hemispheres significant convergence in the STG extended from the mid-STG to the posterior STG, which produced a separate ALE-cluster in the left hemisphere. According to the Brück, Kreifelts, and Wildgruber (2011) model of emotional prosody perception, the posterior superior temporal cortex (STC) is involved in more abstract representation of emotional prosody (stage 2). Moreover, as will be elaborated further in the section on the task-dependent emotional prosody perception network, it appears that activity in the posterior STC during auditory processing can be influenced by task demands (e.g. Brechmann & Scheich, 2005) making it likely that this area participates later in the prosodic perception process than the mid-STG in line with the model proposed by Brück et al. (2011a).

As could be expected from both models of emotional prosody perception, significant clusters of convergence were found in the bilateral IFG. Both models of emotional prosody perception implicate the bilateral IFG in a final multimodal stage where emotional prosody is explicitly evaluated and integrated with other layers of speech. Lesion studies (Breitenstein, Daum, & Ackermann, 1998; Hornak et al., 2003; Hornak, Rolls, & Wade, 1996) and a more a recent TMS-study (Hoekert, Vingerhoets, & Aleman, 2010) have furthermore indicated that the inferior frontal cortex is not only sufficient but also necessary for adequate emotional prosody perception. If the IFG are indeed involved in a final multimodal explicit evaluative phase, it would be expected that the bilateral IFG are involved in the explicit evaluation of emotion in other modalities such as emotional faces or emotional scenes as well, which has been confirmed by recent meta-analyses of the respective neuroimaging literature (for emotional faces see Fusar-Poli et al., 2009a; for emotional faces and scenes see Sabatinelli et al., 2011).

In the right inferior frontal gyrus the ALE-cluster extended into the dorsal anterior insula. Of the two proposed models of emotional prosody perception proposed by Kotz and Paulmann (2011) and Brück, Kreifelts and Wildgruber (2011) only the latter proposes a role for the anterior insula, namely in automatic (implicit) processing of emotion. However, insula activation is not consistently found in fMRI studies during implicit processing of emotional prosody (Bach et al., 2008; Ethofer et al., 2011; Frühholz, Ceravolo, & Grandjean, 2011; Grandjean et al., 2005) but see (Morris, Scott, & Dolan, 1999; Mothes-Lasch, Mentzel, Miltner, & Straube, 2011). Furthermore, the neuroimaging literature suggests that there is functional differentiation within the anterior insula. For instance, a recent large scale meta-analysis of the functional neuroimaging literature reports that the dorsal insula might be an integration zone important for general maintenance of the task-set while the ventral insula is specifically involved in the recognition of emotion and empathy (Kurth, Zilles, Fox, Laird, & Eickhoff, 2010). Although the exact role of insular clusters in the emotional prosody perception process is admittedly speculative at this time, the dorsal insula might control general prosodic task performance while the ventral insula might specifically support emotion recognition and empathy during emotional prosody perception.

Although two ALE clusters were found in the basal ganglia (BG), more specifically the right medial globus pallidus and caudate body and the left claustrum, these clusters did not survive our additional criterion that at least three studies should have directly contributed to a cluster. Given the relatively low number of experiments in the analysis, this might have been a result of power being just too low to detect these clusters above threshold. Notably, no significant convergence was found in the amygdala, a structure often claimed to be essential for emotional processing. Although both models of emotional prosody perception implicate the amygdala in emotional prosody perception, namely in the detection of emotional salience (Kotz & Paulmann, 2011) and in automatic (implicit) emotional prosody perception (Brück, Kreifelts, & Wildgruber, 2011) lesion studies have failed to find an effect of amygdala damage on emotional prosody perception task performance (Adolphs, 2001; Adolphs & Tranel, 1999), suggesting that the amygdala might not be crucial for emotional prosody perception. Possibly the amygdala is only engaged in the emotional prosody perception process under specific task conditions, preventing robust convergence of peak activity in the amygdala across studies as analyzed by the present meta-analysis. Another possibility is that the amygdala habituates quickly to emotional prosody stimulation (Wiethoff, Wildgruber, Grodd, & Ethofer, 2009) reducing the ability to detect the amygdala activity over prolonged periods of stimulation as is typical for fMRI experiments. Lastly, the amygdala response to emotional prosody may vary considerably between individuals depending on personality traits as has been shown in a recent study (Brück, Kreifelts, Kaza, Lotze, & Wildgruber, 2011).

4.2. Task-driven network

Contrasting active processing of emotional prosody to processing of the same acoustic material while directing attention to a different attribute (such as semantics or speaker identity) captures more abstract later stages of the emotional prosody perception process. Both proposed models of emotional prosody perception implicate the bilateral IFG in the evaluation of emotional prosody (stage 3) but differ in the exact neuroanatomical underpinnings of stage 2 (integration of acoustic information into an emotional percept or representation of meaningful suprasegmental acoustic sequences). While the Brück, Kreifelts and Wildgruber (2011) model proposes that stage 2 processing occurs in the posterior superior temporal cortex, the Kotz and Paulmann model (2011) postulates that this more abstract stage of processing is supported by the anterior STS.

The most concurrent ALE-cluster for the task-driven network lay in the posterior STG/STS immediately posterior to HG, a region that is sometimes referred to as the 'planum temporale' (Westbury et al., 1999). Contrarily, no significant convergence across studies was found in the anterior STS (not even sub-threshold) for task-driven emotional prosody processing. This difference in convergence does not seem attributable to methodological factors such as a difference in susceptibility artifact in the BOLD signal (Devlin et al., 2000) as both regions lay in the superior temporal region, clear from air filled cavities. Therefore, stage 2 emotional prosody processing is more likely to be supported by the posterior STG/STS as proposed by Brück, Kreifelts, and Wildgruber (2011) than the anterior STS as proposed by Kotz and Paulmann (2011). The posterior auditory association cortex in the posterior STG seems (given its associative nature) well equipped for the proposed integrative function of stage 2 emotional prosody processing. Furthermore, a recent meta-analysis has shown that the posterior auditory association cortex is specialized in the processing of spectral information (Samson, Zeffiro, Toussaint, & Belin, 2010). As spectral parameters are important for the recognition of emotional prosody (Scherer, 2003) the posterior auditory association cortex might contribute to emotional prosody perception through enhanced spectral processing. Furthermore, neuroimaging studies have shown that activity of the posterior STG can be (top-down) modulated by task demands (Ahveninen et al., 2006; Bunzeck, Wuestenberg, Lutz, Heinze, & Jancke, 2005; Jäncke, Mirzazade, & Shah, 1999), which is clearly compatible with a role for the posterior STG in task-dependent emotional prosody perception as found in the present analysis and as proposed by Brück et al. (2011a).

In line with both models, significant convergence across studies was found in the right IFG for task-dependent processing of emotional prosody. Although significant ALE-values were also found in the left IFG homolog, these voxels did not survive cluster thresholding which might be a consequence of the relatively low power of the task-dependent analysis. As was already pointed out in the discussion of the IFG clusters in the stimulus-driven network, the IFG likely plays a role in explicit multimodal emotional evaluation and the integration of emotional prosody with other layers of speech. Interestingly, a recent imaging study suggests that there is functional connectivity between the right posterior temporal cortex and the right IFG during emotional prosody processing (Ethofer et al., 2006). Although in their dynamic causal model a forward projection from the posterior temporal cortex to the IFG was more probable than vice versa, another interesting possibility might be that activity in the posterior auditory association cortex can be selectively enhanced to flexibly augment processing of task-relevant (spectral) acoustic properties through top-down mediation by the IFG as has been suggested by Leitman et al. (2010).

4.3. Hemispheric specialization for emotional prosody

As outlined in the introduction, both proposed models of emotional prosody perception allow for hemispheric specialization in multiple stages of the emotional prosody perception process. More specifically, both models postulate (relative) right-hemispheric specialization for the first and second stage of emotional prosody perception while only the model by Kotz and Paulmann (2011) additionally leaves open the possibility of right-hemispheric specialization for emotional evaluation in the final stage. Indeed, a recent meta-analysis of the lesion literature found a statistically robust greater emotional prosodic performance degradation following right than left hemispheric damage (Witteman, Van IJzendoorn, Van de Velde, Van Heuven, & Schiller, 2011), suggesting that the right cerebral hemisphere is of relatively greater importance than the left in the decoding of emotional prosody. However, due to the typically limited spatial specificity of acquired lesions this meta-analysis could not address at which intrahemispheric loci this right-hemispheric specialization might arise, an issue the present meta-analysis is better equipped to address.

The lateralization analysis for the stimulus-driven network revealed right-lateralized convergence of ALE-clusters in the medial transverse temporal gyrus. However, the cluster only emerged at a more lenient threshold and therefore has to be interpreted with caution. Nevertheless, this result suggests that hemispheric specialization might emerge very early at the level of the (medial) HG (where the human PAC is likely to be located). Previous studies outside the prosody domain have indeed found functional lateralization at the level of HG. For instance, it has been found that the left and right HG may already show an advantage for temporal (Jamison, Watkins, Bishop, & Matthews, 2006; Warrier et al., 2009) and spectral (Mathys, Loui, Zheng, & Schlaug, 2010; Warrier et al., 2009; Zatorre, 1988) processing, respectively. As spectral variation is an important acoustic dimension for recognizing emotional prosody (Scherer, 2003), the right HG might make a relatively greater contribution to emotional prosody perception than its left hemispheric homolog through enhanced spectral processing.

The task-driven lateralization analysis revealed a right-lateralized ALE-cluster in the posterior STG, but again only at the more lenient threshold warranting caution in interpreting this result. Previous studies have shown that the right posterior STG might be specialized in pitch processing (Hyde, Peretz, & Zatorre, 2008; Johnsrude, Penhune, & Zatorre, 2000). Hence the right posterior STG, similarly to the right HG, might contribute disproportionally as compared to the left posterior STG to the perception of emotional prosody through its superior pitch-processing capability. Furthermore, in line with a task-dependent role of the posterior STG in emotional prosody perception as found in the present investigation, previous studies have shown that lateralization of posterior STG activity during auditory processing is dependent on task demands (Brechmann & Scheich, 2005; Geiser, Zaehle, Jancke, & Meyer, 2007). For instance, Brechmann and Scheich (2005) showed that when subjects had to categorize the pitch direction of frequency-modulated tones, the right posterior STG was more active than the left, but when the duration of the same material had to be categorized the hemispheric specialization shifted to the left STG. As these authors suggest, the left or right posterior STG might be recruited through top-down mediation to dynamically enhance task-relevant auditory processing. In the case of emotional prosodic processing, the right posterior STG might be similarly recruited to enhance spectral processing to support emotional prosody perception.

The present meta-analyses therefore suggest that the statistically robust relative right-hemispheric specialization for emotional prosody perception found previously in the lesion literature (Witteman, Van IJzendoorn, Van de Velde, Van Heuven, & Schiller, 2011) might be explained by multiple (relative) right-hemispheric asymmetries in the primary and associative auditory cortex, possibly corresponding to the proposed first and second stage of emotional prosody perception, respectively. In the latter case increased activity in the right posterior auditory association cortex might reflect increased effort to extract acoustic properties relevant to emotional prosody perception through top-down modulation by the right IFG (Leitman et al., 2010), which would render hemispheric specialization for emotional prosody perception an interactive dynamic process, although this hypothesis admittedly requires further explicit evaluation. These hemispheric asymmetries early in the emotional prosody perception process are in line with cue-dependent hypotheses (e.g. Van Lancker & Sidtis, 1992) of hemispheric specialization for emotional prosody perception. More generally, the results suggest that hemispheric specialization for emotional prosody can be explained by specialization of the left and right cerebral hemisphere for the processing of more basic (non-prosody-specific) acoustic dimensions such as spectral versus temporal processing (Zatorre & Belin, 2001) or integration over small versus large temporal windows (Poeppel, 2003), respectively. Such an early auditory bottom-up account of hemispheric specialization for emotional prosody has, in fact, been suggested previously (Ethofer et al., 2011; Obleser, Eisner, & Kotz, 2008; Robin, Tranel, & Damasio, 1990) and can also explain why no right hemispheric specialization has been found for visual emotional (i.e. facial) expression in a recent large-scale meta-analysis of the neuroimaging literature (Fusar-Poli et al., 2009b).

4.4. Strengths and limitations

To the best of the authors' knowledge, this is the first quantitative meta-analysis of the neuroimaging literature on emotional prosody perception. Individual imaging studies generally suffer from small sample size and the results reflect study specific-details in the experimental design. Quantitative meta-analysis alleviates these problems and enables one to study which neural network is reliably associated with a neurocognitive process with high statistical power. A limitation of ALE meta-analysis is that only the peak-coordinates of activation clusters are considered and not the effect size or extend of activation. However, when an attempt was made to compensate for this weakness by manually checking the extent and magnitude of lateralized clusters for the lateralization analyses, it was found that these measures largely confirmed the coordinate-based analyses. A further limitation was that due to the limited number of neuroimaging papers on emotional prosody perception published to date, it was not possible to formally test for the influence of moderating variables that have been suggested in the literature. Hence, the exact role of factors such as appraisal level and specific emotions in the neural processing of emotional prosody warrants further investigation.

4.5. Conclusion

In sum, the present meta-analyses implicate a bilateral temporofrontal network of areas including the PAC, mid-STG, post-STG and the IFG in the perception of emotional prosody, largely confirming recently proposed models of emotional prosody perception. Activation likelihood was larger in the posterior-STG than the anterior-STS, suggesting that the proposed second stage of emotional prosody perception is more likely to be supported by the posterior-STG, as suggested by Brück, Kreifelts and Wildgruber (2011) than the anterior-STS as suggested by Kotz and Paulmann (2011). Concerning hemispheric specialization, multiple right lateralized ALE-clusters were found in the auditory cortex but only at a more lenient threshold. These results suggest that the relative right-hemispheric specialization for emotional prosody perception found previously in the lesion literature (Witteman, Van IJzendoorn, Van de Velde, Van Heuven, & Schiller, 2011) might be a multi-stage (and possibly interactive) process that might be best explained by relatively early hemispheric specialization for prosody-relevant acoustic processing. However, future meta- (or mega-) analyses with a larger number of experiments are needed to conclusively test this hypothesis.

6. References[1]

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

Adolphs, R., & Tranel, D. (1999). Intact recognition of emotional prosody following amygdala damage. *Neuropsychologia*, 37, 1285-1292.

Ahveninen, J., Jaaskelainen, I. P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., . . . Belliveau, J.W. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proceedings of the National Academy of Science of the United States of America*, *103,* 14608-14613.

*Bach, D. R., Grandjean, D., Sander, D., Herdener, M., Strik, W. K., & Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *NeuroImage, 42*, 919-927.

*Beaucousin, V., Lacheret, A., Turbelin, M. R., Morel, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). FMRI study of emotional speech comprehension. *Cerebral Cortex, 17*, 339-352.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*, 309-312.

Bennett, C. M., & Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Annals of the New York Academy of Sciences, 1191*, 133-155.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience, 8*, 389-395.

Brüne, M., & Brüne-Cohrs, U. (2006). Theory of mind--evolution ontogeny brain mechanisms and psychopathology. *Neuroscience & Biobehavioral Reviews, 30*, 437-455.

Brechmann, A., & Scheich, H. (2005). Hemispheric shifts of sound representation in auditory cortex with conceptual listening. *Cerebral Cortex, 15,* 578-587.

Breitenstein, C., Daum, I., & Ackermann, H. (1998). Emotional processing following cortical and subcortical brain damage, contribution of the fronto-striatal circuitry. *Behavioural Neurology, 11*, 29-42.

Brück, C., Kreifelts, B., & Wildgruber, D. (2011a). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Phyics of Life Reviews, 8,* 383-403.

*Brück, C., Kreifelts, B., Kaza, E., Lotze, M., & Wildgruber, D. (2011b). Impact of personality on the cerebral processing of emotional prosody. *NeuroImage, 58,* 259-268.

*Buchanan, T.W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., & Jankce, L. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research, 9*, 227-238.

Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H. J., & Jäncke, L. (2005). Scanning silence: Mental imagery of complex sounds. *NeuroImage, 26,* 1119-1127.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E.,. . . Tyler, L. K. (2000). Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task. *Neuroimage, 11,* 589-600.

Dietrich, S., Hertrich, I., Alter, K., Ischebeck, A., & Ackermann, H. (2008). Understanding the emotional expression of verbal interjections: A functional MRI study. *Neuroreport, 8,* 1751-1755.

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology, 6,* 178-190.

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: Random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping, 30,* 2907-2926.

Eickhoff, S. B., Bzdok, D., Laird, A. R., Roski, C., Caspers, S., Zilles, K., & Fox, P. T. (2011). Co-activation patterns distinguish cortical modules, their connectivity and functional differentiation. *Neuroimage,* 57, 938-949.

*Ethofer, T., Wiethoff, S., Anders, S., Kreifelts B., Grodd, W., & Wildgruber, D. (2007). The voices of seduction: Cross-gender effects in processing of erotic prosody. *Social Cognitive and Affective Neuroscience, 2,* 334-337.

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., Grodd, W., & Wildgruber, D. (2006). Cerebral pathways in processing of affective prosody, A dynamic causal modeling study. *NeuroImage, 21,* 580-587.

*Ethofer, T., Bretscher, J., Geschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012). Emotional voice areas, anatomic location functional properties and structural connections revealed by combined fMRI/DTI. *Cerebral Cortex, 22,* 191-200.

*Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009). Differential influences of emotion task and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience, 21,* 1255-1268.

Ethofer, T., Van de Ville, D., Scherer, K., & Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology, 19,* 1028-1033.

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., & Petersen, S. E. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. *Nuclear Science Symposium and Medical Imaging Conference, 3*, 1813-1817.

*Frühholz, S., Ceravolo, L., Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex*, *22,* 1107-1117.

Fusar-Poli P., Placentino A., Carletti, F., Allen, P., Landi, P., Abbamonte, M., . . . Politi, P. L. (2009b). Laterality effect on emotional faces processing, ALE meta-analysis of evidence. *Neuroscience Letters, 452,* 262-267.

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., . . . & Politi, P. (2009a). Functional atlas of emotional faces processing: A voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of Psychiatry and Neuroscience, 34,* 418-432.

Galaburda, A., & Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. *Journal of Comparative Neurology, 190,* 597-610.

*Gandour, J., Wong, D., Dzemidzic, M., Lowe, M., Tong, Y., & Li, X. (2003). A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. *Human Brain Mapping, 18,* 149-157.

Geiser, E., Zaehle, T., Jäncke, L., & Meyer, M. (2007). The neural correlate of speech rhythm as evidenced by metrical speech processing. *Journal of Cognitive Neuroscience, 20,* 541-552.

Genovese, C. R., Lazar, N. A., Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, *15,* 870-878.

*George, M. S., Parekh, P. I., Rosinsky, N., Ketter, T. A., Kimbrell, T. A., Heilman, K., . . . Post, R. M. (1996). Understanding emotional prosody activates right hemisphere regions. *Archives of neurology, 53,* 665-670.

Gobras, M., Beaton, S., & Eickhoff, S. B. (2012). Brain regions involved in human movement perception: A quantitative voxel-based meta-analysis. *Human Brain Mapping, 33,* 431-454.

*Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience, 8,* 145-146.

Hoekert, M., Kahn, R. S., Pijnenborg, M., & Aleman, A. (2007). Impaired recognition and expression of emotional prosody in schizophrenia: Review and meta-analysis. *Schizophrenia Research, 96,* 135-145.

Hoekert, M., Vingerhoets, G., & Aleman, A. (2010). Results of a pilot study on the involvement of bilateral inferior frontal gyri in emotional prosody perception, an rTMS study. *BMC Neuroscience, 11,* 93.

Hornak, J., Bramham, J., Rolls, E. T., Morris. R. G., O'Doherty, J., Bullock P. R., & Polkey., C. E. (2003). Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. *Brain, 126,* 1691-1712.

Hornak, J., Rolls, E. T., & Wade, D. (1996). Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia, 34,* 247-261.

Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia, 46,* 632-639.

*Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., Itoh, K., Kato, T., . . . Nakamura, K. (1997). Vocal identification of speaker and emotion activates different brain regions. *NeuroReport, 8,* 1809-2812.

Jäncke, L., Mirzazade, S., Shah, J. N. (1999). Attention modulates activity in the primary and the secondary auditory cortex: A functional magnetic resonance imaging study in human subjects. *Neuroscience Letters, 266,* 125-128.

Jamison, H. L., Watkins, K. E., Bishop, D. V. M., & Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cerebral Cortex ,16,* 1266-1275.

Johnsrude, I. S., Penhune, V. B., & Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain, 123,* 155-163.

*Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language, 86,* 366-376.

Kotz, S. A., & Paulmann, S. (2011). Emotion language and the brain. *Language and Linguistics Compass, 5,* 108-125.

Kurth, F., Zilles, K., Fox, P., Laird, A., & Eickhoff, S. (2010). A link between the systems: Functional differentiation and integration within the human insula revealed by meta-analysis. *Brain Structure and Function, 214,* 519-534.

Laird, A. R., Fox, P. M., Price, C. J., Glahn, D.C., Uecker, A. M., Lancaster, J. L., . . . Fox, P. T. (2005). ALE meta-analysis: Controlling the false discovery rate and performing statistical contrasts. *Human Brain Mapping, 25,* 155-164.

Lancaster, J.L., Tordesillas-Gutiérrez, D., Martinez, M., Salinas, F., Evans, A., Zilles, K., . . . Fox, P. T. (2007). Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping, 28,* 1194-1205.

*Leitman, D. I., Wolfm D. H., Ragland, J. D., Laukka, P., Loughead, J., Valdez, J. N., . . . Gur, R. C. (2010). "It's not what you say but how you say it": a reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience*, *19,* 1.

Mathys, C., Loui, P., Zheng, X., & Schlaug, G. (2010). Non-invasive brain stimulation applied to Heschl's gyrus modulates pitch discrimination. *Frontiers in Psychology, 1*, 193.

*Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. R. (2003). The neural response to emotional prosody as revealed by functional magnetic resonance imaging. *Neuropsychologia, 41,* 1410-1421.

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

*Mothes-Lasch, M., Mentzel, H. J., Miltner, W. H. R, & Straube, T. (2011). Visual attention modulates brain activation to angry voices. *The Journal of Neuroscience, 31,* 9594-9598.

Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *The Journal of Neuroscience, 28,* 8116-8123.

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

Robin, D. A., Tranel, D., & Damasio, H. (1990). Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions. *Brain and Language, 39*, 539-555.

*Rota, G., Veit, R., Nardo, D., Weiskopf, N., Birbaumer, N., & Dogil, G. (2008). Processing of inconsistent emotional information: An fMRI study. *Experimental Brain Research, 186,* 401-407.

Sabatinelli, D., Fortune, E. E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W. T., Beck, S., & Jeffries, J. (2011). Emotional perception: Meta-analyses of face and natural scene processing. *NeuroImage, 54,* 2524-2533.

Samson, F., Zeffiro, T. A., Toussaint, A., & Belin, P. (2010). Stimulus complexity and categorical effects in human auditory cortex: An activation likelihood estimation meta-analysis. *Frontiers in Psychology, 1*, 241.

*Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *NeuroImage, 28,* 848-858.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40,* 227-256.

Scherer, K. R., Banse, R., & Wallbott, H.G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32,* 76-92.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10,* 24-30.

*Schirmer, A., Escoffier, N., Zysset, S., Koester, D., Striano, T., & Friederici, A. D. (2008). When vocal processing gets emotional: An the role of social orientation in relevance detection by the human amygdala *NeuroImage, 40,* 1402-1419.

Silk, J. B. (2007). Social components of fitness in primate groups. *Science, 317,* 1347-1351.

Talairach, J., & Tournoux, P. (1988). Co-planar stereotaxic atlas of the human brain: 3-d proportional system: An approach to cerebral imaging. Stuttgart: Thieme.

Turkeltaub, P. E., & Coslett, B. H. (2010). Localization of sublexical speech perception components. *Brain and Language, 114*, 1-15.

Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage, 16,* 765-780.

Turkeltaub, P.E., Eickhoff, S. B., Laird, A. R., Fox, M., Wiener, M., & Fox, P. (2011). Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Human Brain Mapping*, *33,* 1-13.

Trask, R. L. (1996). *A dictionary of phonetics and phonology*. London: Routledge.

Van Bezooijen R., Otto, S.A., & Heenan, T.A. (1983). Recognition of vocal expressions of emotion. *Journal of Cross-Cultural Psychology, 14*, 387-406.

Van der Laan, L. N., De Ridder, D. T. D., Viergever, M. A., & Smeets, P. A. M. (2011). The first taste is always with the eyes: A meta-analysis on the neural correlates of processing visual food cues. *NeuroImage, 55*, 296-303.

Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. Papers in Linguistics: *International Journal of Human Communication, 13*, 201-277.

Van Lancker, D., & Sidtis, J. J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research, 35,* 963-970.

Warrier, C., Wong, P., Penhune, V., Zatorre, R., Parrish, T., Abrams, D., & Kraus, N. (2009). Relating structure to function: Heschl's gyrus and acoustic processing. *The Journal of Neuroscience, 29,* 61-69.

Westbury, C. F., Zatorre, R. J., & Evans, A.C. (1999). Quantifying variability in the planum temporale: A probability map. *Cerebral Cortex, 9,* 392-405.

*Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., & Ethofer, T. (2008). Cerebral processing of emotional prosody--influence of acoustic parameters and arousal. *NeuroImage, 39,* 885-893.

Wiethoff, S., Wildgruber, D., Grodd, W., & Ethofer, T. (2009). Response and habituation of the amygdala during processing of emotional prosody. *NeuroReport, 20,* 1356-1360.

*Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., & Ackermann H (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral Cortex,14,* 1384-1389.

*Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., & Ackermann, H. (2005). Identification of emotional intonation evaluated by fMRI. *NeuroImage, 24,* 1233-1241.

Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology, 11,* 277-281.

Witteman J, Van IJzendoorn, M. H, Van de Velde D, Van Heuven, V. J., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. *Neuropsychologia, 49,* 3722-3738.

Wittfoth, M., Schröder, C., Schardt, D. M., Dengler, R., Heinze, H. J., & Kotz, S. A. (2010). On emotional conflict: Interference resolution of happy and angry prosody reveals valence-specific effects. *Cerebral Cortex, 20,* 383-392.

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *The Journal of the Acoustical Society of America, 84*, 566-572.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*, 946-953.

[1] References marked with an asterisk (*) were included in the meta-analyses

Chapter 2.4

Selective automatic activation of the superior temporal gyrus for the perception of angry prosody

Witteman, J., Veer, I. M., Van Steenbergen, H., Rombouts, S. A, Van Heuven, V. J., & Schiller, N. O. (Submitted). Selective automatic activation of the superior temporal gyrus for the perception of angry prosody.

Abstract

Significant progress has been made in delineating the neural network involved in the perception of emotional prosody. However, it is still unclear to what extent there is differential involvement of neural structures in the intentional (explicit) vs. unintentional (implicit) perception of emotional prosody. Further, it is unknown whether automaticity of emotional prosody perception found in previous investigations is specific for threatening prosody or a general property of emotional prosody perception. Therefore, in the present study we used functional magnetic resonance imaging to investigate the neural networks involved in the intentional and unintentional perception of surprise and anger prosody. The core temporofrontal prosody perception network was found to be active independently of instruction, in line with recent models of emotional prosody perception. For explicit emotional prosody perception, additional activation was found in areas previously implicated in 'Theory of Mind' processing. Hemispheric specialization within the core emotional prosody perception network showed a rightward asymmetry for implicit prosody, but shifted to symmetric activation during explicit prosody perception, showing that hemispheric asymmetry for emotional prosody perception is relative and can be modulated by task demands only. Last, persistent supra-threshold activation of the right superior temporal gyrus was found for unintentional perception of anger but not for surprise, supporting models that postulate phylogenetically prepared prioritized processing of threat signals in the brain.

1. Introduction

Accurate perception of non-verbal emotional signals is essential for social interactions. Through the voice, the emotional state of the speaker can be conveyed by means of prosody (the suprasegmental layer of speech), using various acoustic parameters such as speaking rate, fundamental frequency and spectral balance (e.g., Van Heuven & Sluijter 1996; Scherer, 2003). It can be argued that in every-day life most emotional cues are processed spontaneously and only in exceptional situations do we deliberately evaluate the emotional state of the speaker through decoding of non-verbal signals (Critchley et al., 2000). Thus, implicit perception of emotional prosody may represent the rule and explicit perception the exception. The present investigation examined the neural network underlying implicit versus explicit perception of emotional prosody.

Neural models of emotional prosody perception (Kotz & Paulmann, 2011; Bruck et al., 2011; for meta-analytic evidence see Witteman, Van Heuven, & Schiller, 2012) have proposed that emotional meaning is extracted from the speech signal in three stages. First, extraction of basic acoustic properties from the speech signal takes place in the primary and secondary auditory cortex, i.e., Heschl's gyrus (HG) and the mid-superior temporal gyrus (mid STG), where the terms 'anterior', 'mid' and 'posterior' STG will henceforth be used relative to the location of HG (i.e., 'mid' lies parallel to HG, see Turkeltaub, Branch, & Coslett, 2010). Further, both models propose that in a second stage these acoustic properties are integrated into an 'emotional gestalt', but disagree whether its anatomical substrate lies in the posterior STG (Bruck et al., 2011, for meta-analytic support of this model, see Witteman et al., 2012) or anterior STS (Kotz & Paulmann, 2011). Last, both models propose a third stage, in which emotional prosody is evaluated and integrated with other layers of speech (such as the lexico-syntactic meaning of the utterance) in lateral frontal areas. Additionally, it has been suggested that subcortical areas are involved in emotional prosody perception such as the amygdala and basal ganglia (for reviews see Kotz & Paulmann, 2011; Bruck et al., 2011). However, upon closer inspection, the amygdala may not be essential for emotional prosody perception since lesion studies have repeatedly failed to find compromised emotional prosody perception after amygdala damage (Adolphs & Tranel, 1999; Bach, Hurlemann, & Dolan, 2013). The precise role of the various basal ganglia nuclei in the emotional prosody perception process remains to be determined.

Implicit versus explicit processing of emotional processing may tax the various components of the core prosody perception network differentially. Additionally, as explicit processing of prosody demands an extra stage of information processing as compared to implicit emotional prosody processing, additional areas may be recruited during explicit processing. Previous neuroimaging studies have examined this issue in two ways. First, by comparing a condition in which participants actively evaluate emotional prosody with a condition where subjects evaluate a different dimension of the same stimuli (such as gender identification), the components of the explicit emotional prosody perception network can be extracted. These contrasts have generally revealed enhanced processing in the posterior STG and IFG for explicit emotional prosody processing (for meta-analytic evidence, see Witteman, Van Heuven, & Schiller, 2012). Note, however, that these components may additionally reflect other differences between the two tasks used than just the explicit versus implicit processing mode.

Alternatively, this confound can be eliminated by comparing emotional prosody vs. neutral prosody perception under an explicit condition to the same stimulus-driven contrast under an implicit processing condition (Frühholz, Ceravolo, &

Grandjean, 2012). When qualitatively comparing studies using such a stimulus-driven contrast under an explicit condition (Kotz et al., 2003; Leitman et al., 2010; Frühholz et al., 2012) with studies that used a stimulus-driven task under an implicit processing condition (Gandjean et al., 2005; Sander et al., 2005; Frühholz et al., 2012; Mothes-Lasch, Mentzel, Miltner, & Straube, 2011) the explicit studies sometimes find involvement of the middle temporal gyrus MTG (Leitman et al., 2010) and the temporoparietal junction (Fruhholz et al., 2012), while the implicit studies do not. Interestingly, the MTG and temporoparietal junction (TPJ) have been suggested to be involved in 'Theory of Mind' (ToM) processing in the previous literature (Bzdok et al., 2012) and may therefore be part of an extended emotional prosody perception network for deliberate (explicit) analysis of non-verbal emotional information. Crucially, the only two studies that have compared explicit to implicit processing directly using stimulus-driven contrasts, suggest involvement of the anterior cingulate and the putamen in explicit processing of emotional prosody (Bach et al., 2008: Frühholz et al., 2012). Thus, in the present investigation we similarly contrasted explicit and implicit processing of a stimulus-driven contrast to examine which regions of the emotional prosody perception network are involved in implicit and explicit processing of emotional prosody.

It has been suggested that emotional processing might be relatively 'automatic' (Anderson, Christoff, Panitz, De Rosa, & Gabrieli, 2003; Sander et al., 2005, for a recent review see Straube, Mothes-Lasch & Miltner, 2011) and that such automaticity might reflect a phylogenetically prepared mechanism for biologically relevant stimuli to get privileged access to the limited processing capacity of the brain (Öhman & Mineka, 2001) due to selective pressure for processing of emotional signals in evolutionary history (Dunbar, 1998; Silk, 2007). Upon closer examination, the terms 'implicit' and 'automatic' overlap to such an extent that it has been argued that we can replace the former with the latter, as the latter is better defined (Moors & De Houwer, 2006). According to Moors and De Houwer, automaticity is a gradual concept that comprises several core features such as uncontrollability, efficiency and lack of awareness of the process – the more of these features are present the more 'automatic' a process can be considered to be. One aspect of the uncontrollability aspect of automaticity is whether a process is initiated even when the subject does not have the goal to initiate the process (i.e., unintentionally of emotional processing). Thus, in the present investigation we examined whether emotional prosody processing persists despite lack of the participant's intention to engage in emotional prosody processing. On the neural level, we defined the persistence of emotional prosody processing as the continued presence of supra-threshold neural activity during unintentional emotional prosody perception, as compared to intentional emotional prosody processing (similarly to Anderson et al., 2003).

Finally, it has been suggested that there might have been particular selective pressure for automatic processing of negative emotions and more specifically threat (e.g., anger) since rapid threat detection can prevent harm to the organism and this promote fitness (Öhman & Mineka, 2001; Anderson, Christoff, Panitz, De Rosa, & Gabrieli, 2003; Schupp et al., 2004). Indeed, some evidence for automaticity of the perception of threatening (i.e. anger) prosody has been found previously (Sander et al., 2005; Grandjean et al., 2005). Further, there is evidence from the visual modality that negative emotional signals in particular attract additional processing resources (Holmes, Bradley, Kragh, Nielsen, & Mogg, 2009; Ito, Larsen, Smith, & Cacioppo, 1998; Schupp et al. 2004; but for conflicting evidence see Schröder et al., 2006) as compared to positive emotional stimuli. Thus, processing of threat may be relatively automatic as compared to processing non-threatening emotional information. To test this hypothesis, it is required to examine whether the persistence of neural activity under unintentional emotional processing is particularly evident for negative emotions as compared to positive emotions, which has not been examined for emotional prosody yet. Thus, as a last goal of the present investigation, we examined the persistence of supra-threshold neural activity during task-irrelevant processing of positive versus negative (threatening) emotional prosody.

2. Methods

2.1. Participants

A total of 25 participants took part in the experiment. Five participants were excluded because of lower than chance level performance and one more due excessive motion (average displacement > 2.5 mm), resulting in a total of 19 participants (mean age 24.92, SD = 5.65; 6 male, 13 female). All participants were right-handed as determined by the Edinburgh Handedness Inventory (M = 88.11, SD = 11.32, minimum = 67), were native speakers of Dutch, had normal or corrected-to-normal vision and no current psychiatric disorder or a history of neurological disease. Furthermore, none of the participants reported hearing problems or substance dependence. The study was approved by the local ethics committee and conducted in accordance with the Declaration of Helsinki, and informed consent was obtained from all participants prior to the experiment. Participants received € 20 for their participation in the two hour fMRI session.

2.2. Materials

Pseudowords (see appendix A) with a disyllabic structure were generated. All pseudowords obeyed Dutch phonotactics and were checked for the absence of semantic content. All pseudowords were expressed with neutral, (pleasantly) surprised and angry prosody and with stress either on the first or second syllable by two professional actors (one male, one female) and recorded at 16 bit resolution and a 44.1 kHz sampling rate in a sound-proofed booth. Surprised and angry prosody were chosen in order to sample both positive and negative emotions - both of which are considered to be approach emotions (thus, there was no confounding role of the approach-withdrawal dimension). Items were intensity-normalized and had a mean duration of 756 ms (SD = 65 ms).

In line with previous literature (e.g., Scherer, 2003), the emotional categories differed from neutral prosody in mean F_0 and F_0 variability; anger differed additionally in intensity variability (see Table 1). Note that the three categories did not differ in mean intensity as the stimuli had been intensity-normalized. To verify the validity of the intended prosodic contrasts, a panel of five healthy volunteers classified each prosodic stimulus (in addition to happy and sad prosodic stimuli that were not used in the present study) in a forced choice task. Only pseudowords were selected for which all of the emotional prosodic categories (neutral, surprised and angry intonation) were classified correctly by at least 4 out of 5 panel members for both actors.

Two categorization tasks were created that presented identical stimuli under an implicit and explicit emotional condition. For the *explicit* (intentional) emotional condition, participants were instructed to decide whether the stimulus sounded neutral, angry or surprised. For the *implicit* (unintentional) emotional task participants were required to categorize a non-emotional dimension of the same stimuli (whether metrical stress was on the first or the second syllable). From the pool of validated stimuli, 32 items of each emotional category were selected, with half of the words bearing metrical stress on the first and the other half on the second syllable. Speaker gender was balanced across all items.

Table 1. Acoustic properties of the emotional categories.

	Neutral	Anger	Surprise
Mean intensity (dB)	79.45	79.27	80.64
Mean variation (SD) intensity	8.84	10.74	8.83
Mean F_0 (Hz)	180.73	281.35	282.46
Mean variation (SD) F_0	44.78	78.56	101.19
Mean total duration (s)	0.79	0.76	0.72

2.3. Procedure

Each participant performed both tasks, each of which took 12 minutes to complete. The implicit emotional task was always presented first to reduce participants' awareness of the emotional dimension of the stimuli (and hence to prevent intentional processing of emotional prosody, allowing us to investigate the intentionality dimension of automaticity in emotional prosody perception (see Moors & De Houwer, 2006). Participants were instructed that they would hear a nonsense word and asked to categorize the task-relevant dimension (emotion, metrical stress) as fast and accurately as possible with a right hand button press. Assignment of individual categories to response buttons was counterbalanced across subjects. Participants were instructed that they could respond while the stimulus was still playing (i.e., reaction times [RTs] were recorded from the onset of the stimulus). All other instructions were exactly the same for both tasks except for words describing the task-specific (emotional versus non-emotional) categories.

Each categorization task started with 12 practice trials. Participants practised the tasks in the scanner with simulated scanner noise and kept practising until a performance level of at least 75% correct was reached. Subsequently, the 96 test trials (32 items per emotional category) were performed.

Throughout the experiment a black fixation cross was presented in the center of a grey background. Auditory stimuli were presented binaurally through MR-compatible headphones and a trial ended 2000 ms after stimulus onset. Stimuli were presented in an event-related fashion with a jittered inter stimulus interval (between 4000 and 8000 ms). Stimulus presentation order was (pseudo-)random with the restriction that no more than two consecutive presentations of the same stimulus category were allowed. Participants were instructed to fix their gaze on the fixation cross throughout the experiment.

Stimulus presentation was controlled using E-prime 1.2 and stimulus material was presented at 16 bit resolution and a 44.1 kHz sampling frequency at a comfortable intensity level. Subjects reported that the stimuli could be perceived clearly despite of the scanner noise.

2.4. FMRI data acquisition

Imaging data were acquired on a Philips 3.0-T Achieva MRI scanner using an eight-channel SENSE head coil for radiofrequency transmission and reception (Philips Medical Systems, Best, The Netherlands). For each task, whole-brain fMRI data were acquired using T_2* - weighted gradient-echo echo-planar imaging (EPI) with the following scan parameters: 355 volumes (the first 5 volumes were discarded to reach signal equilibrium); 38 axial slices scanned in ascending order; repetition time (TR) = 2200 ms; echo time (TE) = 30 ms; flip angle = 80°; FOV = 220 × 220 mm; 2.75 mm isotropic voxels with a 0.25 mm slice gap. A high-resolution anatomical image (T_1-weighted ultra-fast gradient-echo acquisition; TR = 9.75 ms; TE = 4.59 ms; flip angle = 8°; 140 axial slices; FOV = 224 × 224 mm; in-plane resolution .875 × .875 mm; slice thickness = 1.2 mm), and a high-resolution T_2*- weighted gradient echo EPI scan (TR = 2.2 s; TE = 30 ms; flip angle = 80°; 84 axial slices; FOV = 220 × 220 mm; in-plane resolution 1.96 × 1.96 mm, slice thickness = 2 mm) were additionally acquired for registration to standard space.

2.5. FMRI data preprocessing

Prior to analysis, all fMRI data sets were submitted to a visual quality control check to ensure that no gross artefacts were present in the data. Next, data were analyzed using FSL Version 4.1.3 (Smith et al., 2004). The following preprocessing steps were applied to the EPI data sets: motion correction, removal of non-brain tissue, spatial smoothing using a Gaussian kernel of 8 mm full width at half maximum (FWHM), grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor, and a high pass temporal filter of 70s (i.e., ≈0.014 Hz). The dataset was registered to the high resolution EPI image, the high resolution EPI image to the T1-weighted image, and the T1-weighted image to the 2 mm isotropic MNI-152 standard space image (T1-weighted standard brain averaged over 152 subjects; Montreal Neurological Institute, Montreal, QC, Canada). The resulting transformation matrices were then combined to obtain a native to MNI space transformation matrix and its inverse (MNI to native space).

2.6. Behavioral analysis

Repeated-measures analyses of variance (RM-ANOVA) were performed with Task as within-subject factor and proportion correct responses and mean reaction time for correct responses as dependent variables.

2.7. FMRI-data analysis

2.7.1. Whole brain analyses

Data analysis was performed using FEAT (FMRI Expert Analysis Tool) Version 5.98, part of FSL (FMRIBs Software Library, http://www.FMRIb.ox.ac.uk/fsl). In native space, the FMRI time-series was analyzed in an event-related fashion using the General Linear Model with local autocorrelation correction applied (Woolrich, Ripley, Brady, & Smith, 2001). For both runs, onsets of each of the stimulus categories (neutral, angry, surprised) were modeled as events with 800 ms duration in separate regressors. Regressors were convolved with a double gamma hemodynamic response function, and temporal derivatives of each of the three regressors were added to the model. If errors were present, these trials were included in the model with a separate regressor. At first level, every emotional category (neutral, surprise, anger) was contrasted against baseline and against the other emotional categories.

At second level, three whole-brain analyses were performed. First, the main effect of emotional (anger and surprise) versus neutral prosody across both tasks was analyzed to examine the task-independent effect of emotional prosody. Second, the effect of emotional versus neutral prosody was analyzed under the explicit and implicit condition separately to examine activations associated with explicit and implicit emotional prosody perception. Last, the processing of emotional prosody versus neutral prosody was directly compared between the two tasks to examine which brain regions might be specifically engaged during implicit versus explicit emotional prosody perception. All whole-brain analyses were cluster-corrected for multiple comparisons using an initial cluster-forming threshold of $Z > 2.3$ ($p < .01$), and a corrected significance threshold of $p < .05$ (Worsley, 2001) was applied.

2.7.2. Regions of Interest (ROI) analyses

The previous literature suggests that the superior temporal gyrus (STG) and the inferior frontal gyrus (IFG) are particularly important in the perception of emotional prosody (Schirmer & Kotz, 2006; Brück, Kreifelts, & Wildgruber, 2011; Witteman, Van Heuven, & Schiller, 2012). Therefore, the whole-brain analyses were followed up with an a priori region-of-interest (ROI) analysis on the STG and IFG. Anatomical ROIs were created for the IFG (pars opercularis) and STG (posterior part) at the 50% probability threshold, as defined by the Harvard-Oxford cortical probability atlas (http://www.fmrib.ox.ac.uk/fsl/data/atlas-descriptions.html#ho). For these ROIs, the mean *Z*-values were extracted for each stimulus category (anger, surprise, neutral) against baseline and for each emotional category (surprise, anger) against neutral, for each task and each participant using Featquery (http://fsl.fmrib.ox.ac.uk/fsl/fsl4.0/feat5/featquery.html). These *Z*-transformed parameter estimates indicate how well the mean signal of each ROI is explained by the model (i.e., by the presence of the particular stimulus category). Using these mean *Z*-values as dependent variables, two RM-ANOVAs were performed. First, to examine the effect of prosodic category (anger, surprise, neutral), a RM-ANOVA was performed with the Z-values against baseline as dependent variables and the factors Task, ROI, Hemisphere and Emotion as within-subject factors. Secondly, to specifically examine lateralization effects for emotional prosody an additional RM-ANOVA was performed with the Z-values of each emotional category (anger, surprise) against neutral as dependent variable and the factors Task, ROI, Hemisphere and Emotion as within-subject independents. For this analysis only main effects of and interactions with the factor Hemisphere are reported. In case of sphericity-violations, Greenhouse-Geisser corrected *p*-values are reported.

3. Results

3.1. Behavioral results

As depicted in Figure 1, participants performed well above chance level for the explicit task ($M = 88\%$, $SD = 7.8\%$) and the implicit task ($M = 83\%$, $SD = 8.5\%$). Furthermore, a RM-ANOVA with the proportion correct responses as dependent variable, Task and Emotion as the within-subject factors revealed a main effect of Task ($F(1,18) = 6.21$, $p = .02$), which was qualified by a Task × Emotion interaction ($F(2,17) = 10.74$, $p = .001$). Follow up RM-ANOVAs for each task with Emotion as independent variable showed that there were no significant differences in accuracy between the emotional categories for the explicit task (for all, $p > .08$, Sidak corrected). For the implicit task however, accuracy was lower when the task-irrelevant emotional category was surprise as compared to neutral ($p = .03$, Sidak corrected) and to anger ($p < .001$, Sidak corrected), while there was no difference between anger and neutral ($p = .21$).

A RM-ANOVA with reaction time (RT) for correct categorization as the dependent variable revealed a main effect of Task ($F(1,18) = 4.83$, $p = .04$) with slower responses for the implicit task ($M = 1174$; $SD = 133$) than the explicit task ($M = 1111$; $SD = 100$). Further, there was a main effect of Emotion ($F(1,18) = 19.54$, $p < .001$) which was qualified by a Task × Emotion interaction ($F(2,36) = 7.81$, $p = .007$). Follow-up RM-ANOVAs for each task, indicated that for the explicit task surprise ($p = .02$, Sidak corrected) and anger ($p <.001$, Sidak corrected) were categorized faster than neutral, while anger and surprise were not significantly different ($p = 0.07$, Sidak corrected). For the implicit task categorization was slower when the task irrelevant emotional category was surprise as compared to neutral ($p = .04$, Sidak corrected) and anger ($p < .001$, Sidak corrected) while there was no difference between neutral and anger ($p = .69$).

In sum, the behavioral results suggest that participants perceived the stimuli well and that performance was better for anger and surprise in the explicit task than neutral, and that performance improved when anger was the task irrelevant emotional category during the implicit task.

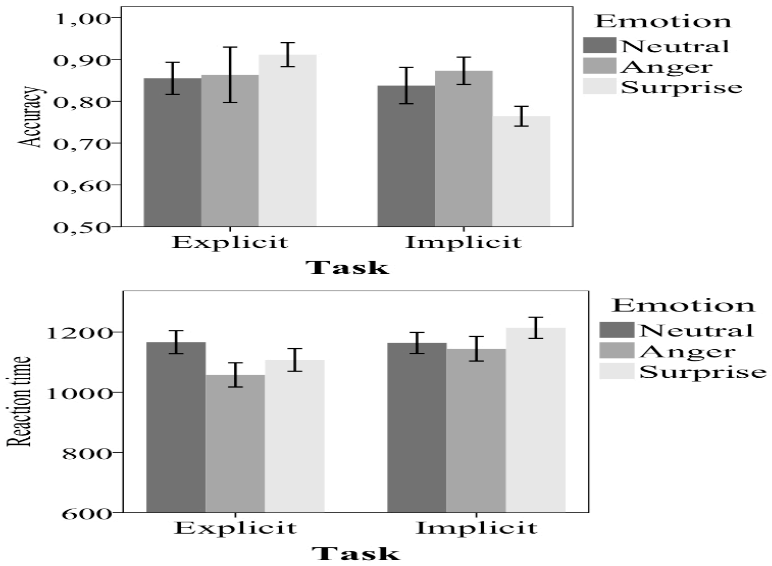

Figure 1. Behavioral results. Mean accuracy (upper graph) and reaction time (lower graph) for each task and each emotional category. Error bars indicate the 95% confidence interval.

3.2. FMRI results

3.2.1. Whole-brain analyses

All effects for the whole-brain analyses can be found in Table 2. The whole-brain cluster-corrected ($p < 0.05$) analysis for [emotion > neutral] across both tasks revealed one cluster with a peak in the right anterior STG/STS (peak MNI coordinates (60, −10, −6), peak $Z = 4.48$, extent = 673 voxels), extending both anteriorly and posteriorly with respect to HG. As shown in Figure 2, additional activation in the left STG and IFG was observed at a more liberal threshold ($p < .001$, uncorrected). For [anger > neutral] the cluster-corrected analysis revealed a cluster with a peak in the right posterior STG/STS (peak MNI coordinates (63, −30, 2), peak $Z = 4.49$, extent = 1215 voxels) and the left mid STG/STS (−54, −22, 0), peak $Z = 4.12$, extent = 326 voxels), extending posteriorly.

As can be observed in Figure 3 [emotion > neutral] for the explicit task revealed a large cluster with a peak in the left mid STS (−58, −26, −4), peak $Z = 3.56$, extent = 1381 voxels), extending into the anterior STG and posterior MTG. A second cluster was observed in the left temporoparietal junction (TPJ) with a peak in the supramarginal gyrus (−52, −54, 32), peak $Z = 3.75$, extent = 1131 voxels), extending into the angular gyrus and the posterior STG. As can be seen in Figure 4 contrasting anger against neutral [anger > neutral] revealed a cluster with a peak in the right posterior STS (60, −30, 0), peak $Z = 3.6$, extent = 1250 voxels), extending into the STG anteriorly and posteriorly with respect to HG. Contrasting surprise to neutral [surprise > neutral] revealed a large cluster with a peak in the left posterior MTG (−60, −54, −2), peak $Z = 3.81$, extent = 2579 voxels), extending anteriorly into the MTG/STS/STG and dorsally into TPJ.

For the implicit task, contrasting both emotional categories to neutral did not reveal any significant clusters. However, as can be seen in Figure 4, contrasting anger to neutral [anger > neutral] revealed a large cluster with a peak in the right mid STG/STS (70, −24, 4), peak $Z = 3.54$, extent = 1351 voxels), extending both anteriorly and posteriorly along the STG/STS. As can be observed in Figure 4 this cluster overlapped to a considerable extent with the cluster found for anger processing in the explicit condition. Lastly, contrasting surprise to neutral [surprise > neutral] resulted in a cluster with a peak in the medial SFG (0, 12, 58), peak $Z = 3.59$, extent = 99 voxels), extending ventrally into the cingulate gyrus.

Directly contrasting emotional prosody processing (versus neutral) in the explicit task with processing of the same emotional categories in the implicit condition (and vice versa) did not reveal any significant clusters of activation.

In sum, the whole-brain analysis indicated that parts of the left and right STG were involved in emotional prosody perception. For the explicit task, TPJ was additionally activated and surprise activated the left MTG during the explicit task and the medial SFG in the implicit task. Interestingly, anger activated the right STG during both tasks while surprise did not. Lastly, no area was significantly more active during emotional prosody perception in the implicit task as compared to the explicit task or vice versa.

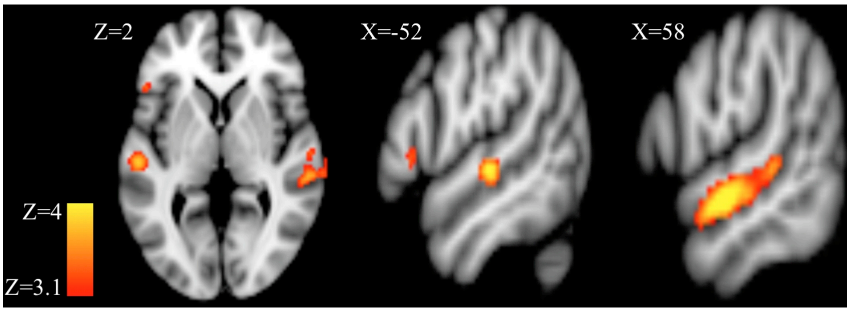

Figure 2. Activation map for emotional > neutral prosody across tasks overlaid on the MNI standard brain. Activation is thresholded at *p* < .001 uncorrected for illustration purposes. The left and right side of the image represent the left and right side of the brain.

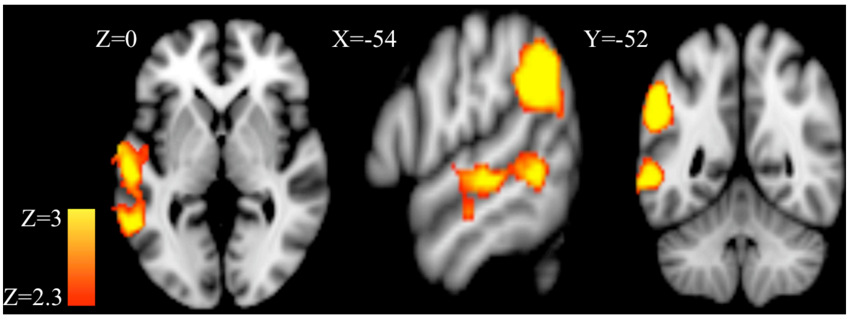

Figure 3. Activation clusters for emotional > neutral in the explicit task overlaid on the MNI standard brain. Clusters of significant activation (p < .05, cluster corrected) were observed in the left mid-STS (extending into the anterior STG and posterior MTG) and left TPJ (cluster peak in the supramarginal gyrus).

Table 2. Peak activations for the whole brain analyses.

Effect	**Contrast**	**Brain region (peak)**	**Peak MNI (x,y,z)**	**Z-value**	**Cluster size in voxels**
Main effect	[Emotion > neutral]	R STG	60, −10, −6	4.48	673
	[Anger > neutral]	R STG; L STG	63,−30, 2; −54,−22, 0	4.49; 4.12	1215; 326
Explicit task	[Emotion > neutral]	L STG; L SMG;	−58, −26, −4; −52, −54, 32	3.56; 3.75	1381; 1131
	[Anger > neutral]	R STG	60,−30, 0	3.6	1250
	[Suprise > neutral]	L MTG	−60, −54, −2	3.81	2579
Implicit task	[Anger > neutral]	R STG	70,−24,4	3.54	1351
	[Surprise > neutral]	M SFG	0, 12, 58	3.59	999

L=left, *R*=Right, *M*=midline. STG=superior temporal gyrus, MTG = middle temporal gyrus, SFG = superior frontal gyrus, SMG = supramarginal gyrus. Coordinates are given in Montreal Neurological Institute (MNI) standard space. All reported clusters surive a $p<0.05$ cluster correction.

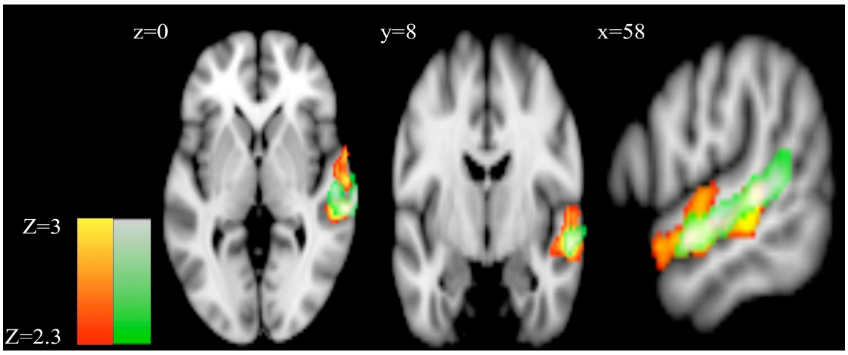

Figure 4. Activation clusters for anger > neutral in the explicit task (red to yellow) and the implicit task (green to white). A significant cluster of activation ($p < .05$, cluster corrected) was found in the right STG/STS for both tasks, but activation extended more anteriorly for the explicit task and more posteriorly for the implicit task.

3.2.2. ROI-analyses

ROI analyses were performed to further characterize the effect of emotional prosody in the core temporofrontal network, including the STG and IFG, which was robustly associated with emotional prosody perception in previous studies. The mean *Z*-values for each ROI in each condition against baseline are shown in Figure 5. A RM-ANOVA with Task, ROI, Hemisphere and Emotion as within-subject factors and the *Z*-values against baseline as dependent variables revealed a main effect of Task ($F(1,18) = 12.89$, $p = .002$), ROI ($F(1,18) = 53.66$, $p < .001$), Hemisphere ($F(1,18) = 6.01$, $p = .02$) and Emotion ($F(1,18) = 4.04$, $p = .04$). The Task, ROI and Emotion main effects were qualified by a three-way Task × ROI × Emotion interaction ($F(1,17) = 21.46$ $p < .001$). Decomposing the interaction with separate RM-ANOVAs for each ROI revealed a significant Task × Emotion interaction for the STG (($F(2,36) = 4.47$, $p = .04$) but not for the IFG ($p = .12$). Following the interaction up for the STG with separate RM-ANOVAS for each task revealed a main effect of Emotion in the explicit task ($F(1,18) = 7.63$, $p = 0.003$) and in the implicit task ($F(1,18) = 10.66$, $p = .001$) for the STG. Post-hoc tests (Sidak corrected) indicated that for the explicit task, both anger ($p = .03$) and surprise ($p = .004$) activated the STG more than neutral while anger and surprise were not significantly different ($p = 1$, NS). For the implicit task, anger activated the STG more than neutral ($p = .002$) and surprise ($p = .01$) while surprise and neutral were not significantly different ($p = .91$, NS).

To further characterize lateralization effects of emotional prosody, the *Z*-score of each emotion (anger, surprise) against the neutral condition was used as dependent variable. The analysis revealed a Task × Hemisphere interaction (($F(1,18) = 5.86$, $p = .03$) and a ROI × Hemisphere × Emotion interaction (($F(1,18) = 5.41$, $p = .03$). Following up the Task × Hemisphere interaction with a separate ANOVA for each task, revealed that for the explicit task there was no significant difference between the hemispheres ($p = .39$, NS) while for the implicit task right-hemispheric activation was significantly larger than left-hemispheric activation (($F(1,18) = 6.39$, $p = .02$). However, in the uncorrected statistical map significant activity was noted in the left STG, showing that the right-hemispheric advantage was relative. Decomposing the ROI × Hemisphere × Emotion interaction with separate ANOVAs for each ROI did not reveal a significant Hemisphere × Emotion interaction for the STG (($F(1,18) = .34$, $p = .57$) but a significant Hemisphere × Emotion interaction in the IFG (($F(1,18) = 7.57$, $p = .01$). Follow up RM-ANOVAs for each emotional category revealed a trend for a right-hemispheric advantage for anger in the IFG (($F(1,18) = 4.15$, $p = .06$) but no difference between the hemispheres for surprise (($F(1,18) = 1.57$, $p = .23$)

In sum, the ROI analyses suggest that the bilateral STG was sensitive to both emotional categories (versus neutral) when attention was directed at the emotional prosody. However, when attention was diverted away from the emotional prosodic dimension during the implicit task, only anger maintained more activation than neutral in the STG while surprise did not. Thus, diverting attention away from the emotional prosody decreased processing of surprise as compared to neutral below the statistical significance threshold (although the decrease in activation between tasks was not significant) in the STG, while above-threshold activation for anger processing was sustained. Lastly, we found a modulation of hemispheric asymmetry by task demands in the temporofrontal network, with bilateral activation for emotional (against neutral) prosody in the explicit task and relative right lateralization in the implicit task.

Therefore, the unilateral activation clusters found for the whole-brain analyses in the STG seem largely due to thresholding effects.

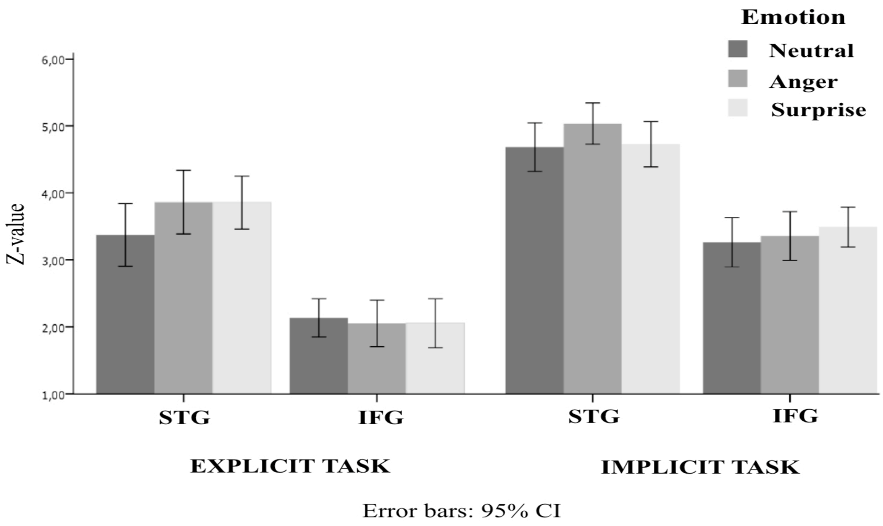

Figure 5. The mean *Z*-value for each task and each emotion is shown for the IFG and the STG. Error bars indicate the 95% confidence interval.

4. Discussion

The aim of the present study was to investigate the brain areas involved in the implicit (unintentional) and explicit (intentional) perception of emotional prosody. Further, it was investigated whether automaticity for emotional prosody processing could be demonstrated, and if so, whether it might be particularly evident for anger. First, we found activation of the core temporofrontal prosody perception network (comprising the STG and IFG) independently of task demands. Further, for explicit perception of emotional prosody additional activation was found in the left middle temporal gyrus, angular gyrus and the supramarginal gyrus. However, comparing both tasks directly for both emotions against neutral revealed no differential activation between both modes of processing. Last and importantly, anger processing showed sustained supra-threshold activation of the right STG under both task conditions while surprise only showed supra-threshold STG activation in the explicit condition.

In line with recent models of emotional prosody perception (Kotz & Paulmann, 2011; Brück, Kreifelts, & Wildgruber, 2011) and meta-analytic evidence (Witteman, Van Heuven, & Schiller, 2012), activation was found along the STG/STS for stimulus driven (i.e., emotion contrasted against neutral) emotional prosody perception and additionally in the IFG (uncorrected) independently of task demands. Thus, this result adds to the growing evidence base in support of a core temporofrontal network that is involved in the (at least) three stage emotional prosody perception process, involving extraction of acoustic features relevant to emotional prosody, integration of those features into more abstract representations and finally evaluation of these representations and integration with other layers of the speech signal.

Our results further suggest that intentional processing of emotional prosody may recruit additional areas beyond the core emotional prosody perception network. During intentional emotional prosody perception, additional activation was found in the left MTG and the left TPJ extending into the supramarginal and angular gyrus, while during unintentional (implicit) emotional prosody processing these areas were not significantly activated (however these areas were not more active during the explicit task as compared to the implicit task). Involvement of the MTG or TPJ in intentional emotional prosody perception has been observed in previous work (Wildgruber, Pihan, Ackermann, Erb, & Grodd, 2002; Ethofer et al., 2009; Leitman et al., 2010; Frühholz et al., 2012, but see Kotz et al., 2003) and within the context of emotional processing the MTG has been suggested to be involved in abstract multimodal processing of emotion (Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003; Pourtois, De Gelder, Bol, & Crommelinck, 2005; Johnstone, Van Reekum, Oakes, & Davidson, 2006). Further, a recent meta-analysis (Bzdok et al., 2012) of the neuroimaging literature suggests that the MTG and TPJ play a key role in 'Theory of Mind', the ability to contemplate about others' thoughts and feelings in an abstract manner. Thus, the MTG and TPJ may be optionally recruited in addition to the core emotional prosody perception network for more abstract and intentional analysis of others' feelings as communicated through the voice. However, the involvement of these additional structures during intentional prosody perception may be a matter of degree rather than type, as the direct statistical comparison between both tasks did not reveal any clusters differentiating between both tasks. Further, although admittedly this is accepting the null hypothesis, the direct comparison between the two tasks suggests that the core prosody perception network is also similarly active during intentional (explicit) and unintentional (implicit) emotional prosody perception.

Regarding hemispheric specialization for emotional prosody perception, previous meta-analytic evidence has suggested that there is relative right-hemispheric specialization (Witteman, Van IJzendoorn, Van de Velde, Van Heuven, & Schiller, 2011) that might be driven by relatively early specialization for the processing of acoustic cues relevant to emotional prosody perception (such as F_0) in the primary and secondary auditory cortex (Witteman, Van Heuven, & Schiller, 2012). Although in the present study the whole brain analysis for task independent effects revealed a cluster in the right STG, inspection of the uncorrected activation maps revealed bilateral activation in the STG. Further, ROI analysis of the STG and IFG showed overall bilateral activation independent of task, bilateral activation during the explicit task and relatively greater activation in the right than the left hemisphere during implicit (unintentional) emotional prosody perception. A problem with interpreting the previous neuroimaging evidence regarding hemispheric specialization for emotional prosody perception is that most studies do not perform a formal statistical test for lateralization and only report (lateralized) clusters of supra-threshold activation. With this approach the possibility cannot be excluded that in reality there is no meaningful difference in activity between the hemispheres. Indeed, previous neuroimaging studies that did perform a formal test of hemispheric lateralization either found no hemispheric asymmetry (Grandjean et al., 2005) or relative right hemispheric asymmetry to bilateral activation, depending on the specific emotional category analyzed (Leitmann et al., 2010). Thus, in line with these studies and meta-analytic work (Witteman et al., 2011; Witteman et al., 2012), our results suggest that right hemispheric specialization for emotional prosody perception is relative at best. Further, in accordance with previous work (e.g. Luks, Nusbaum, & Levy, 1998; Gandour et al., 2004) the present results suggest that the hemispheric asymmetry during prosody perception can be modulated by task demands only (as acoustic material was held constant between the tasks), with greater activation of the right hemisphere than the left during implicit emotional prosody perception shifting to symmetric activation during explicit emotional prosody perception. Although speculative, a greater contribution of the left hemisphere during explicit emotional prosody perception might be related to the added verbal load required for verbal labeling of emotional prosody (Ross et al., 1997).

Of note, no significant activation of the amygdala was found in any of the activation maps. Although the amygdala is often claimed to be essential for adequate emotional prosody perception, a recent meta-analysis of the neuroimaging literature of emotional prosody perception did not find significant convergence of activity across studies in the amygdala (Witteman et al., 2012). Further, lesion studies have repeatedly failed to find significant disturbance of emotional prosody perception in patients with amygdala lesions as compared to controls (Adolphs & Tranel, 1999; Adolphs, Tranel, & Damasio, 2001; Bach, Hurlemann, & Dolan, 2013), suggesting that the amygdala might not be necessary for emotional prosody perception. However, it has been suggested that amygdala involvement in the emotional prosody perception process might be hard to detect with a whole brain analysis because there is differential involvement of the various subnuclei (Frühholz et al., 2012) and because the amygdala rapidly habituates to emotional prosody (Wiethof, Wildgruber, Grodd, & Ethofer, 2009; but see Scheuerecker et al., 2007), preventing its detection with typical fMRI designs that involve sustained stimulation, as the present.

We were additionally interested in whether automaticity for emotional prosody perception could be demonstrated, and if so, whether automaticity would be particularly evident for threat processing, as has been predicted on evolutionary grounds (Öhman & Mineka, 2001; Anderson, Christoff, Panitz, De Rosa, & Gabrieli, 2003; Schupp et al., 2004). Building on the conceptual work on automaticity by Moors and De Houwer (2006), we specifically investigated whether the 'uncontrollability' aspect of automaticity could be demonstrated for emotional prosody perception – i.e., whether emotional prosody processing would persist even when participants do no intend to initiate the emotional prosodic perception process. On the neural level, we adopted the criterion for automaticity proposed by Anderson et al. (2003) – i.e., the persistence of supra-threshold activation of neural structures during unintentional emotional prosody perception as compared to intentional emotional prosody perception. Indeed, in the present study the right STG remained significantly active when subjects did not intentionally analyze emotional prosody for anger (as compared to neutral) but not for surprise. Activation of the right temporal cortex during unintentional emotional prosody perception has been found previously for anger as compared to neutral prosody (Grandjean et al., 2005; Sander et al., 2005). However, in these previous studies anger was the only emotional prosody category employed, rendering it uncertain the possibility that such automaticity is specific for anger processing or that it may be a general property of emotional prosody perception. Thus, the present results extent this previous work by providing initial evidence that automaticity of emotional prosody perception seems to be specific to threat. Further, the results support models postulating prioritized processing of negative and particularly threatening emotional signals in the brain (i.e., 'negativity bias', see Ito, Larsen, Smith, & Cacioppo, 1998), possibly through a phylogenetically prepared mechanism (e.g., Grossman, Striano, & Friederici, 2005; for a seminal review on preparedness of threat perception, see Öhman & Mineka, 2001). However, it should be noted that even for threat perception automaticity of processing is likely to be relative in the sense that if attentional resources are sufficiently depleted, activation during unintentional emotional perception is eventually reduced (Pessoa, McKenna, Gutierrez, & Ungerleider, 2002; Mothes-Lasch, Miltner, & Straube, 2012; for a review on automaticity of emotional processing and its relativity, see Straube, Mothes-Lasch, & Miltner, 2011). Thus, future investigations should investigate whether specific automatic activation for anger prosody can be replicated and persists under further reduced levels of attention.

Strengths and limitations

The present study had comparably high statistical power as compared to previous investigations of implicit and explicit emotional prosody perception (Bach et al., 2008; Frühholz et al., 2012). As has been noted before (Straube, Mothes-Lasch, & Miltner, 2011; Mothes-Lasch, Miltner, & Straube, 2012), research on automaticity of emotional prosody perception is still in its infancy and has so far exclusively relied on negative emotions. Thus, the present study was the first to include a positive emotional category in addition to anger and neutrality under intentional and unintentional emotional prosody perception, allowing for the first time a test of the specificity of automatic activation during threat perception.

Our study also had some limitations. First, the unintentional emotional perception task was always performed first. Therefore, we cannot exclude the possibility that differences between the two tasks may have been influenced by habituation or sensitization effects. However, as the main goal of the present investigation was to test the specificity of automatic processing of angry prosody, the fixed task order was deemed necessary, as has been suggested in previous work (Bach et al., 2008) in order to minimize intentional emotional prosody processing during the unintentional task. However, we cannot rule out the possibility that participants did intentionally analyze emotional prosody during the unintentional perception task after all. As suggested in previous work, however (Bach et al., 2008), note that if this had been the case, it would have had the net effect of a more conservative estimation of differences between the two tasks. Second, there were two response categories during the implicit task while there were three during the explicit task. Thus, we cannot exclude the possibility that the use of an extra finger during the intentional emotional perception task may have contributed to differences in activation between the two task conditions.

Conclusion

The present study aimed to investigate the brain areas involved in the implicit and explicit perception of emotional prosody. Activation of the core temporofrontal emotional prosody perception network was found independently of task demands, in line with recent models of emotional prosody perception. For intentional (explicit) perception of emotional prosody, additional activation was found in the MTG and TPJ, possibly reflecting abstract and intentional analysis of others' emotions. Direct comparison between the two tasks, however suggested that the same network is active during both intentional and unintentional emotional prosody perception. Further, in line with recent meta-analytic evidence, the present results suggest that there is relative right-hemispheric specialization for emotional prosody perception at best. The results further suggest that hemispheric asymmetry during prosody perception can be modulated by task demands only. Last, significant persistence of supra-threshold activation during unintentional (implicit) emotional prosody perception was found in the right STG for anger but not surprise, supporting evolutionary grounded models that postulate prioritized processing of threat signals in the brain.

5. References

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

Adolphs, R., & Tranel, D. (1999). Intact recognition of emotional prosody following amygdala damage. *Neuropsychologia, 37,* 1285-1292.

Anderson, A. K., Christoff, K., Panitz, D., De Rosa, E., & Gabrieli, J. D. E. (2003). Neural correlates of the automatic processing of threat facial signals. *The Journal of Neuroscience, 23*, 5627-5633.

Bach, D. R., Grandjean, D., Sander, D., Herdener, M., Strik, W. K., & Seifritz, E. (2008). The effect of appraisal level on processing of emotional prosody in meaningless speech. *NeuroImage, 42,* 919-927.

Bach, D. R., Hurlemann, R., & Dolan, R. J. (2013). Unimpaired discrimination of fearful prosody after amygdala lesion. *Neuropsychologia, 51,* 2070-2074.

Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews, 8,* 383-403.

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A., Langner, R., & Eickhoff, S. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function, 217,* 783-796.

Critchley, H., Daly, E., Phillips, M., Brammer, M., Bullmore, E., Williams, S., . . . Murphy, D. (2000). Explicit and implicit neural mechanisms for processing of social information from facial expressions: A functional magnetic resonance imaging study. *Human Brain Mapping, 9,* 93-105.

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews, 6,* 178-190.

Ethofer, T., Kreifelts, B., Wiethoff, S., Wolf, J., Grodd, W., Vuilleumier, P., & Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience, 21,* 1255-1268.

Frühholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex, 22,* 1107-1117.

Gandour, T., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu,. . . . Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *NeuroImage, 23,* 344-357.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience, 8,* 145-146.

Grossmann, T., Striano, T., & Friederici, A. D. (2005). Infants' electric brain responses to emotional prosody. *Neuroreport, 16,* 1825-1828.

Holmes, A., Bradley, B. P., Kragh Nielsen, M., & Mogg, K. (2009). Attentional selectivity for emotional faces: Evidence from human electrophysiology. *Psychophysiology, 46,* 62-68.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology, 75,* 887-900.

Johnstone, T., Van Reekum, C. M., Oakes, T. R., & Davidson, R. J. (2006). The voice of emotion: an FMRI study of neural responses to angry and happy vocal expressions. *Social Cognitive and Affective Neuroscience, 1,* 242-249.

Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language, 86,* 366-376.

Kotz, S. A., & Paulmann, S. (2011). Emotion, Language, and the Brain. *Language and Linguistics Compass, 5,* 108-125.

Leitman, D. I., Wolf, D. H., Ragland, J. D., Laukka, P., Loughead, J., Valdez, J. N., . . . Gur, R. C. (2010). "It's not what you say, but how you say it": A reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience, 4,* 19.

Luks, T. L., Nusbaum, H. C., & Levy, J. (1998). Hemispheric involvement in the perception of syntactic prosody is dynamically dependent on task demands. *Brain and Language, 65,* 313-332.

Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia, 41,* 1410-1421.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132,* 297-326.

Mothes-Lasch, M., Mentzel, H. J., Miltner, W. H. R., & Straube, T. (2011). Visual attention modulates brain activation to angry voices. *The Journal of Neuroscience, 31,* 9594-9598.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*, 483-522.

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences, 99,* 11458-11463.

Pourtois, G., De Gelder, B., Bol, A., & Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex, 41,* 49-59.

Ross, E. D., Thompson, R. D., & Yenkosky, J. (1997). Lateralization of affective prosody in brain and the callosal integration of hemispheric language functions. *Brain and Language, 56,* 27-54.

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *NeuroImage, 28,* 848-858.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40,* 227-256.

Scheuerecker, J., Frodl, T., Koutsouleris, N., Zetzsche, T., Wiesmann, M., Kleemann, A. M., . . . Meisenzahl, E. M. (2007). Cerebral differences in explicit and implicit emotional processing--an fMRI study. *Neuropsychobiology, 56,* 32-39.

Schröder, C., Möbes, J., Schütze, M., Szymanowski, F., Nager, W., Bangert, M., . . . Dengler, R. (2006). Perception of emotional speech in Parkinson's Disease. *Movement Disorders, 21,* 1774-1778.

Schupp, H. T., Öhman, A., Junghöfer, M., Weike, A. I., Stockburger, J., & Hamm, A. O. (2004). The facilitated processing of threatening faces: An erp analysis. *Emotion, 4,* 189-200.

Shtyrov, Y., Kujala, T., Lyytinen, H., Kujala, J., Ilmoniemi, R. J., & Näätänen, R. (2000). Lateralization of speech processing in the brain as indicated by mismatch negativity and dichotic listening. *Brain and Cognition, 43,* 392-398.

Silk, J. B. (2007). Social components of fitness in primate groups. *Science, 317,* 1347-1351.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., . . . Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage, 23,* S208-S219.

Straube, T., Mothes-Lasch, M., & Miltner, W.H. (2011). Neural mechanisms of the automatic processing of emotional information from faces and voices. *British Journal of Psychology, 102,* 830-848.

Turkeltaub, P. E., & Branch Coslett, H. (2010). Localization of sublexical speech perception components. *Brain and Language, 114,* 1-15.

Van Heuven, V. J., & Sluijter, A. M. (1996). Notes on the phonetics of word prosody. In R. Goedemans, E. Van der Hulst, & E. Visch. (Eds.). *Stress patterns in the world, part 1: Background* (233-269). The Hague: Holland Institute of Generative Linguistics, Leiden/Holland Academic Graphics.

Wiethoff, S. W., Wildgruber, D., Grodd, W., & Ethofer, T. (2009). Response and habituation of the amygdala during processing of emotional prosody. *Neuroreport, 20,* 1356-1360.

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,* 856-869.

Witteman, J., Van Heuven, V. J., & Schiller, N. O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia, 50,* 2752-2763.

Witteman, J., Van IJzendoorn, M. H., Van de Velde, D., Van Heuven, V. J., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. *Neuropsychologia, 49,* 3722-3738.

Woolrich, M. W., Ripley, B. D., Brady, M., & Smith, S. M. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *NeuroImage, 14,* 1370-1386.

Worsley, K. J. (2001). Statistical analysis of activation images. In P. Jezzard, P. Mathews & S. Smith (Eds.). *Functional MRI: An Introduction to Methods* (251-270). New York: Oxford University Press.

APPENDIX A

Pseudowords used in the experiments:

konpon
dinpil
duldin
kondon
duldun
daldan
paldan
dalpan

Chapter 2.5

The nature of affective priming in music and prosody

Contributions of first and second author:

Conceived the experiments: J. W. and K. G.
Designed the study materials and tasks: J. W.
Data collection: J. W. (pilot studies) and K. G. (EEG)
Analyzed the data: J. W. (behavioral data) and K. G (EEG)
Wrote the first draft of the report: K.G.
Commented on the first draft: J. W.

Goerlich, K. S., Witteman, J., Schiller, N. O., Van Heuven, V. J., Aleman, A., & Martens, S. (2012). The nature of affective priming in music and speech. *Journal of Cognitive Neuroscience, 24,* 1725-1741.

Abstract

The phenomenon of affective priming has caught scientific interest for over 30 years, yet the nature of the affective priming effect remains elusive. This study investigated the underlying mechanism of cross-modal affective priming and the influence of affective incongruence in music and speech on negativities in the N400 time-window. In Experiment 1, participants judged the valence of affective targets (affective categorization). We found that music and speech targets were evaluated faster when preceded by affectively congruent visual word primes, and vice versa. This affective priming effect was accompanied by a significantly larger N400-like effect following incongruent targets. In this experiment, both spreading of activation and response competition could underlie the affective priming effect. In Experiment 2, participants categorized the same affective targets based on nonaffective characteristics. However, as prime valence was irrelevant to the response dimension, affective priming effects could no longer be attributable to response competition. In Experiment 2, affective priming effects were observed neither at the behavioral nor electrophysiological level. The results of this study indicate that both affective music and speech prosody can prime the processing of visual words with emotional connotations, and vice versa. Affective incongruence seems to be associated with N400-like effects during evaluative categorization. The present data further suggest a role of response competition during the affective categorization of music, prosody, and words with emotional connotations.

1. Introduction

In daily life, the ability to adequately process emotional information from our environment is of vital importance. Central to research on affective processing is the long standing 'primacy of emotion' debate, which can be traced back to the founder of experimental psychology (Wundt, 1907). The affective primacy hypothesis assumes that humans are endowed with an evaluative decision mechanism allowing them to automatically evaluate affective stimulus information (e.g., Arnold, 1960; Bartlett, 1932; Zajonc, 1980, 1984). In line with this hypothesis, Fazio, Sanbonmatsu, Powell and Kardes (1986) showed that participants need less time to judge the affective valence of a target stimulus (e.g., 'ugly') if it is preceded by an affectively related prime (e.g., 'hate'). This affective priming effect is thought to be an early, fast-acting, automatic process that can occur outside of conscious awareness (for a review, see Fazio, 2001; Klauer & Musch, 2003). Affective priming has been found for a variety of stimuli, such as pictures, prosody (i.e., melodic and rhythmic aspects of speech), music, and odors. The affective priming effect has caught scientific interest for more than three decades, yet the exact nature of the mechanism causing affective priming remains unclear.

Early explanations proposed spreading of activation as the mechanism underlying affective priming. The spreading of activation account (e.g., Bargh, Chaiken, Raymond, & Hymes 1996; De Houwer, Hermans, & Spruyt, 2001; Spruyt, Hermans, De Houwer, & Eelen, 2002) assumes an associative network of interconnected concept nodes, in which affective valence is represented. An affective prime pre-activates the representations of affectively related targets at the conceptual level by spreading of activation through this network and thereby leads to faster encoding of targets with the same valence as the prime.

More recently, researchers have begun to conceptualize the affective priming effect in terms of conflict at the response stage of processing rather than at the conceptual level (e.g., De Houwer, Hermans, Rothermund, & Wentura, 2002; Wentura, 1999, 2000). According to this account, an affective prime automatically triggers a response tendency that corresponds to its valence. This leads to response facilitation for targets with the same valence as the prime, and to response inhibition when the valence of the target is different from the valence of the prime (Stroop-like interference process). Note that the two accounts need not be mutually exclusive.

Music, the "language of the emotions" (Scherer, 1995) and speech prosody, the vocal expression of emotions ("melody of speech"), have long been assumed to share a common ancestry (Brown, 2000; Dissanayake, 2000; Helmholtz, 1863/1954; Joseph, 1988; Pinker, 1995; Rousseau, 1761/1986; Spencer, 1857). Recent studies on the expression of emotion through music and speech have shown that both involve similar emotion-specific acoustic cues (such as pitch, tempo, and intensity) through which emotion is conveyed in similar ways (Ilie & Thompson, 2006; Juslin & Laukka, 2003; Zatorre, 2002). Such acoustic attributes are thought to be connected with affective connotations that are used to communicate discrete emotions in both vocal and musical expression of emotion (Ilie & Thompson, 2006; Juslin & Laukka, 2003). Recent studies have suggested that both affective speech prosody and music may influence the processing of visual affective stimuli (e.g., Bostanov & Kotchoubey, 2004; Schirmer & Kotz, 2003; Schirmer, Kotz, & Friederici, 2002; Schirmer, Kotz, & Friederici, 2005 for prosody; Daltrozzo & Schoen, 2009; Jolij & Meurs, 2011; Koelsch, Kasper, Sammler, Schulze, Gunter, & Friederici, 2004; Kotz & Paulmann, 2007; Logeswaran & Bhattacharya, 2009; Sollberger et al., 2003; Steinbeis & Koelsch, 2008, 2011 for music).

Electrophysiological studies employing affective priming paradigms found that the N400, an event-related potential (ERP) known to be elicited by semantic mismatches (for a review, see Kutas & Federmeier, 2011), also occurs for mismatches in affective meaning between speech prosody and visually presented words (e.g., Schirmer et al., 2002, 2005; Schirmer & Kotz, 2003). Recently, the N400 has additionally been observed for mismatches in affective meaning between music and linguistic stimuli (Daltrozzo & Schoen, 2009; Steinbeis & Koelsch, 2008, 2011; see Koelsch, 2011 for a review on musical meaning processing). In the present study, we employed both speech prosody and music to compare the capability of both media of interacting with the processing of linguistic stimuli in a cross-modal affective priming paradigm.

The N400 in response to affectively incongruent prosody and music has generally been interpreted to reflect spreading of activation (e.g., Daltrozzo & Schoen, 2009; Steinbeis & Koelsch, 2008, 2011). However, studies investigating the underlying mechanisms of affective priming have pointed out the importance of response competition in such affective evaluation tasks (e.g., Bartholow, Riordan, Saults, & Lust, 2009; Klauer, Roßnagel, & Musch, 1997; Wentura, 1999). One method to test the contribution of response competition to affective priming is to have subjects categorize affective stimuli based on non-affective characteristics or to name the affective targets. As the prime does not convey response-relevant information in these situations, conflict at the response level is eliminated, while spreading of activation remains as a possible mechanism of affective priming. Interestingly, affective priming effects have been observed less reliably in studies using non-affective categorization of affective targets and naming paradigms than in the affective evaluation paradigm (e.g., De Houwer, Hermans, & Eelen, 1998; De Houwer et al., 2002; Klauer & Musch, 2001; Klinger, Burton, & Pitts, 2000; Spruyt, Hermans, Pandelaere, De Houwer, & Eelen, 2004; but see De Houwer & Randell, 2004; Hermans, De Houwer, & Eelen, 2004; Spruyt, De Houwer, Hermans, & Eelen, 2007a, for positive evidence of affective priming in such tasks). Such failures to find affective priming effects during non-affective categorization tasks illustrate the role of response conflict during affective priming. Taken together, a number of studies using behavioral paradigms have provided evidence for a contribution of response competition to affective priming in the evaluative categorization task. It is important to note, however, that affective priming effects may not be exclusively explained by either spreading of activation or response competition, but may also depend on contextual and attentional factors. In fact, Gawronski and coworkers recently showed that affective priming effects in a paradigm based on response conflict (affective evaluation) depended on participants' attention to the category membership of the primes (Gawronski et al., 1010). This demonstrates that affective evaluations as assessed by implicit measures may not be as rigid and inflexible as previously assumed, but may vary with depth of processing and attention under task-specific conditions.

At the electrophysiological level, two previous studies have investigated the contribution of conflict at the response level to affective priming effects during evaluative categorization using visual primes and targets. Bartholow and colleagues (2009) showed that an important factor in driving affective priming effects between words with positive and negative connotations indeed lies in the response system: after prime onset, preferred response activation occurred in motor cortex, as the lateralized readiness potential (LRP) indicated. In addition, increased N2 amplitudes in affectively incongruent conditions suggested that response conflict occurred when the response activated by the prime differed from the target response. Eder and coworkers tested the contribution of semantic priming versus response priming on affective priming between pictures and words (Eder, Leuthold, Rothermund, & Schweinberger, 2011). These authors likewise used the LRP to measure prime-induced response activations and further tested the effect of affective incongruence on the P300 and the N400. Their findings show an earlier occurring stimulus-locked LRP in affectively congruent conditions, and increased amplitudes of the N400 in affectively incongruent conditions, whereas the P300 remained unaffected by affective congruence. The authors concluded that both semantic priming and response priming are likely to constitute affective priming effects in the evaluative categorization task.

In summary, the view that affective priming is driven only by spreading of activation has recently been challenged by studies employing electrophysiology in combination with behavioral measures. Using unimodal visual priming paradigms, these studies suggested that conflict at the response level contributes to affective priming in the evaluative categorization task. The present study tested for the first time the mechanisms contributing to cross-modal affective priming between auditory and visual stimuli by systematically varying the possibility of response conflict to occur between two otherwise identical experiments. Furthermore, we aimed to compare the capability of speech prosody and music of affectively priming visually presented linguistic stimuli. To this end, we employed a cross-modal paradigm to test affective priming effects between music, speech prosody, and visually presented words with affective connotations at the behavioral level as well as the impact of affective congruence on negativities in the N400 time-window.

The study comprised two experiments: Experiment 1 aimed to test the occurrence of cross-modal affective priming by emotional music and speech on visually presented word targets, and vice versa. Participants judged the affective valence of the targets (affective categorization task). In this experiment, spreading of activation as well as response competition may cause the affective priming effect. We hypothesized to find a behavioral affective priming effect (longer reaction times) for affectively incongruent music, speech, and word targets, accompanied by increased negativities in the N400 time-window in affectively incongruent compared to congruent conditions.

Experiment 2 employed the same stimuli as experiment 1. However, participants were now asked to categorize the targets based on non-affective characteristics (non-affective categorization task), excluding response competition to occur while still allowing for spreading of activation. If the affective priming effect and N400-like effect in experiment 1 were indeed caused by response competition, no affective priming effect and no negativities in the N400 range should be found in experiment 2.

2. Methods

2.1. Participants

Thirty-two students (16 male, mean age 23.8, SD 4.4) from the University of Groningen participated in experiment 1, 49 different students (24 male, mean age 23.3, SD 4.9) in experiment 2. All participants were right-handed native speakers of Dutch, had normal or corrected-to-normal vision and no hearing impairment. None of the participants were professional musicians. Subjects received € 20 for their participation in the two-hour EEG session. Informed consent was obtained from all participants prior to the study. The study was approved by the local ethics committee of the BCN-Neuroimaging Center Groningen and conducted in accordance with the Declaration of Helsinki.

2.2. Stimuli

The stimulus set comprised 48 words for visual presentation (24 positive, 24 negative, with 50 % denoting persons and 50 % denoting objects), 48 pseudo-words spoken in happy (24) and sad (24) prosody, and 48 music segments expressing happy (24) or sad (24) emotion. All stimuli were validated in three separate pilot studies prior to the experiment.

In the visual word pilot, ten independent raters of Leiden University judged the words with emotional connotations on a 9-point Likert scale (-4 = very negative, 0 = neutral, 4 = very positive). Only words rated 3 or higher by 9 out of 10 raters were included as positive word stimuli, only words rated -3 or lower by 9 out of 10 raters were included as negative word stimuli. Table 1 shows the positive and negative words used as experimental stimuli in both experiments.

For the prosody pilot, bisyllabic pseudo-words that obeyed Dutch phonotactics were recorded with the help of an actress, cut to a length of approximately 600 ms and amplitude normalized using the Praat speech processing software (Boersma & Weenink, 1996). The normalization procedure amplified every stimulus item such that the digitalized sample with the maximum amplitude was set at the maximum positive or negative value of the converter range, and all other samples were scaled proportionally. As a result, all stimuli had about equal intensity. Ten independent raters at Leiden University judged the pseudo-words on a 9-point Likert scale (-4 = very sad, 0 = neutral, 4 = very happy) with the additional option to choose "other" if another emotion than happy or sad was perceived. Only pseudo-words rated 3 or higher for happy prosody and -3 or lower for sad prosody by 9 out of 10 raters were included in the study.

Music excerpts were created from a number of piano and guitar compositions by Western classical music composers (e.g., Bach, Beethoven, Chopin, Mendelssohn; for a full list of compositions see Table 2). From these compositions, segments with a length of 600 ms were excerpted in Praat (cut at zero-crossings), amplitude normalized and subsequently judged by 13 independent raters at the University of Groningen on a 9-point Likert scale (-4 = very sad, 0 = neutral, 4 = very happy) with the additional option to choose "other" if another emotion than happy or sad was perceived. Only music segments rated 3 or higher for happy music and -3 or lower for sad music by 11 out of 13 raters were included in the study.

Table 1. Affective word stimuli (Dutch originals and English translation) used in experiment 1 and 2.

	Positive Words	**English Translation**	**Negative Words**	**English Translation**
1	Bloem	*Flower*	Beul	*Hangman*
2	Bonbon	*Candy*	Bom	*Bomb*
3	Cake	*Cake*	Braaksel	*Vomit*
4	Echtpaar	*Married Couple*	Dief	*Thief*
5	Expert	*Expert*	Galg	*Gallows*
6	Genie	*Genius*	Graf	*Grave*
7	Geschenk	*Present*	Hoer	*Whore*
8	Goedzak	*Good Soul*	Junk	*Junk*
9	Held	*Hero*	Klootzak	*Asshole*
10	Honing	*Honey*	Monster	*Monster*
11	Ijsje	*Ice Cream*	Pijnbank	*Rack*
12	Kanjer	*Hunk*	Pis	*Piss*
13	Lieverd	*Darling*	Pus	*Pus*
14	Maatje	*Buddy*	Racist	*Racist*
15	Paleis	*Palace*	Sadist	*Sadist*
16	Parel	*Perl*	Schijt	*Shit*
17	Roos	*Rose*	Slet	*Slut*
18	Satijn	*Satin*	Sloerie	*Slut*
19	Schatje	*Baby*	Tiran	*Tyrant*
20	Snoep	*Candy*	Tumor	*Tumor*
21	Vriend	*Friend*	Vandaal	*Vandal*
22	Vriendin	*Girlfriend*	Vergif	*Poison*
23	Winnaar	*Winner*	Vetkwab	*Fat Roll*
24	Zon	*Sun*	Viespeuk	*Dirt Bag*

Table 2. Piano and guitar compositions used as a basis for the musical stimuli.

Instrument	**Valence**	**Composer**	**Composition**
Piano	Happy	J. S. Bach	Violin Partita in E major
		Mendelssohn	TN iii/7: *Scherzo*
		Chopin	Op. 7: *Leicht und Luftig*
		Sibelius	Etude in G flat major
			Five Pieces for Piano, Op. 85: *Bellis*
Piano	Sad	Beethoven	Op. 27 No. 2
		Chopin	"Moonlight": *Adagio sostenuto*
		Mendelssohn	Sonata No. 8 in C minor, Op. 13
			Nocturne Op. 27 No.1
			Nocturne B I 49:
			Op. 6: *Sehnsuchtig*
Guitar	Happy	Torroba	Sonatina: *Allegretto*
			Sonatina: *Allegro*
			Castillos de Espańa: *Turégano*
			Castillos de Espańa: *Olites*
			Aires de la Mancha
		J.S. Bach	BWV 1003: *Allegro*
			BWV 1003: *Fuga*
		Albéniz	*Sevilla*
		Grieg	Op. 12 No. 6
		Sor	*Aire Venezolano*
		Tárrega	*Maria*
Guitar	Sad	Mompou	Suite Compostelana: *Canción*
			Suite Compostelana: *Cuna*
			Suite Compostelana: *Coral*
		Albéniz	*Mallorca*
		J. S. Bach	Sonata in A minor, BWV 1003: *Grave*
		Tárrega	*Endecha*
		De Visée	*Changing my tune*

Only piano segments served as experimental stimuli. Guitar segments were additionally included as fillers in experiment 2 for the purpose of an instrument categorization task (piano vs. guitar).

2.3. Procedure

The cross-modal affective priming paradigm used in experiment 1 and 2 included four main conditions (see Figure 1): *MusicTarget* (music target preceded by visual word prime), *ProsodyTarget* (prosody target preceded by visual word prime), *MusicPrime* (visual word target preceded by music prime), and *ProsodyPrime* (word target preceded by prosody prime). Each main condition comprised two congruent and two incongruent subconditions (congruent: positive prime – positive target, negative prime, negative target, incongruent: positive prime – negative target, negative prime – positive target). Each of the four main conditions (*MusicTarget, ProsodyTarget, MusicPrime, ProsodyPrime*) consisted of 96 trials. Overall, each word, prosody, and music stimulus was presented twice, once congruent and once incongruent, eliminating stimulus characteristics as an explanation of priming effects. Prime – target pairs were created in a randomized fashion. The order in which prime – target pairs were presented was random with the restriction that consecutive targets or primes were never the same and no more than 3 targets were presented consecutively. The four main conditions were presented in four separate blocks, whose order of presentation was counterbalanced according to a Latin square.

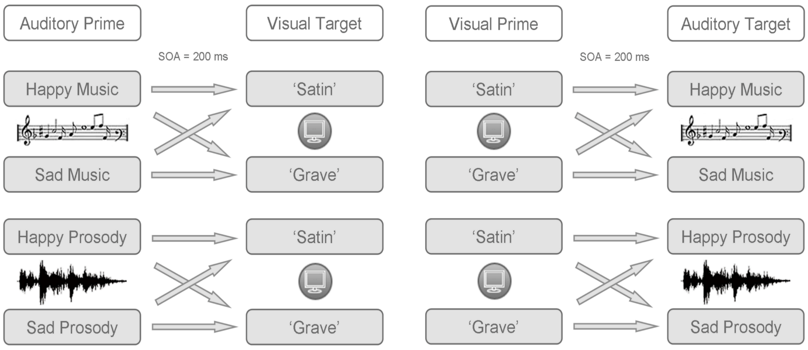

Figure 1. The cross-modal affective priming paradigm.

Stimulus presentation was controlled using E-Prime (1.2). Each trial started with a black fixation cross in the middle of the screen (1500 ms), followed by a red fixation cross (500 ms) signaling the occurrence of the prime. To reduce blink artifacts, participants were instructed to blink when the fixation cross was black, and not to blink anymore when it turned red. When the red fixation cross disappeared, the prime was presented, followed by the target after 200 ms. The stimulus onset asynchrony (SOA) of 200 ms was chosen based on findings that the affective priming effect dissipates after 300 ms (Hermans, De Houwer, & Eelen, 2001). Reaction time was recorded from the onset of the target.

In experiment 1, participants were to judge the pleasantness of the target as fast and accurately as possible (affective categorization task). In experiment 2 (non-affective semantic/phonological categorization task), participants judged whether a visual target word was an object or a person, whether a spoken pseudo-word contained a monophthong (pure vowel) or a diphthong (gliding vowel), or whether a music segment was played by a piano or a guitar.

2.4. ERP Recordings

Electroencephalogram (EEG) was recorded from 64 tin electrodes mounted in an elastic electro cap organized according to the international 10/20 system. EEG data were recorded with a linked mastoid physical reference and were re-referenced using an average reference. Bipolar vertical and horizontal electrooculograms (EOGs) were recorded for artifact rejection purposes. The ground electrode was applied to the sternum. Impedance of all electrodes was kept below 5 kΩ for each participant. EEG was continuously recorded with a sampling rate of 500 Hz, amplified, and off-line digitally low-pass filtered with a cut-off frequency of 30 Hz. Participants were seated in front of a monitor at a distance of approximately 50 cm in a dimly lit, electrically shielded and sound-attenuated booth. Music and speech stimuli were presented via loudspeakers placed at the left and right side of the participant at approximately 70dB.

2.5. Data Analysis

The EEG data were analysed with Brain Vision Analyzer (version 1.05). Prior to averaging, trials with eye-movement and blink artifacts were excluded from analysis. Criteria for artifact rejection within an epoch were a maximal voltage step of 50 μV, a maximal difference between two values in a segment of 100 μV, and a minimum and maximum amplitude of -100 μV and 100 μV, respectively.

For experiment 1, all 32 subjects (16 male) were included in the analysis. For experiment 2, one subject was excluded due to noisy ERP data, leaving a total of 48 subjects (25 male) for analysis. ERP epochs for each subject were computed in a 1000 ms time-window and aligned to a 100 ms pre-stimulus baseline. Mean amplitudes for music, speech, and word targets were computed at the N400 time-window (400 to 500 ms after target-onset) for affectively congruent and incongruent conditions in each participant. This time-window was chosen based on previous N400 literature and on visual inspection of the data, which showed consistent differences between conditions for affectively congruent and incongruent targets between 400 and 500 ms.

Regional repeated-measures Analysis of Variance (RM-ANOVA) was conducted in SPSS (17.0) using 30 electrodes in six regions (anterior, central, posterior) in the left and right hemisphere. The left anterior region included electrodes F3, F5, F7, FC3, and FC5, the right anterior region electrodes F4, F6, F8, FC4, and FC6. The left central region included electrodes C3, C5, CP3, CP5, and T7, the right central region electrodes C4, C6, CP4, CP6, and T8. The left posterior region included electrodes P3, P5, P7, PO3, and PO5, the right posterior region electrodes P4, P6, P8, PO4, and PO6. Figure 2 depicts an electrode map with the six regions of electrodes used for analysis.

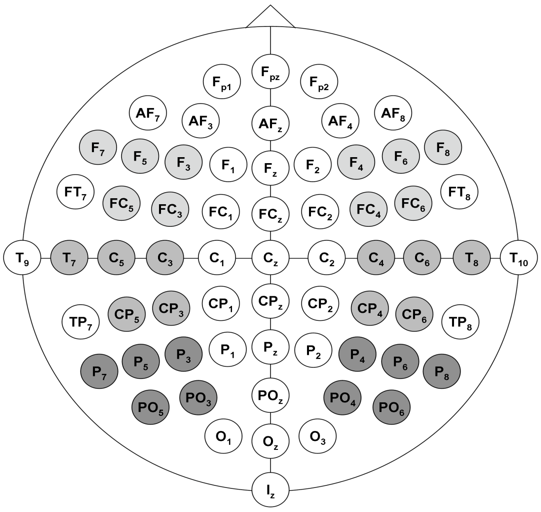

Figure 2. Electrode array used in the experiment. Gray circles indicate the electrodes used.

To test for the effect of prime valence on target processing, the factors *Prime Valence* (positive vs negative) and *Target Valence* (positive vs negative) were entered into the analysis separately. A significant interaction between Prime Valence and Target Valence, indicating ERP differences between affectively congruent and incongruent conditions was interpreted as an affective priming effect. As the main goal of the present study was to examine affective priming effects, only the results of the Prime Valence × Target Valence interaction (i.e., the affective priming effect) and factors qualifying this interaction at the behavioral and electrophysiological level are presented. In case of sphericity violations, Greenhouse-Geisser corrected *p*-values are reported

3. Results

3.1. Behavioral Results

Behavioral data analysis showed that performance was higher than 90% in all conditions, indicating ceiling effects. Therefore, only the results of the reaction time (RT) analyses on correctly identified targets are reported. Significant interactions between prime and target indicate an affective priming effect. Table 3 shows the mean reaction times to positive and negative targets in affectively congruent and incongruent conditions in both experiments. Further information on the percentage of affective priming effects is provided by Table 4 for experiment 1 (N = 32), and by Table 5 for experiment 2 (N = 49).

First, a full ANOVA including the factors *Condition* (MusicTarget vs ProsodyTarget vs MusicPrime vs ProsodyPrime), *Prime Valence* (positive vs negative), and *Target Valence* (positive vs negative) was conducted, with *Sex* (male vs female) and *Experiment* (1 vs 2) as between-subjects factors. Results showed that there were significant prime-target interactions for RT in experiment 1 (see Figure 3) but not in experiment 2 (see Figure 4), indicated by a significant 3-way interaction *Prime Valence* × *Target Valence* × *Experiment*, $F(1,77) = 24.29$, $p < 0.001$ (see Figure 5 for a comparison of the overall affective priming effect between the two experiments). Follow-up ANOVAs were subsequently performed in each condition for experiment 1 (affective categorization) and experiment 2 (non-affective categorization).

3.1.1. Experiment 1: Affective Categorization

MusicTarget

Participants evaluated music segments preceded by affectively congruent visual word primes significantly faster than music segments preceded by affectively incongruent word primes. RM-ANOVA revealed a significant 2-way *Prime Valence* × *Target Valence* interaction [$F(1,30) = 27$, $p < 0.001$], indicating a behavioral affective priming effect for congruent music targets.

ProsodyTarget

Prosody targets were evaluated significantly faster when preceded by affectively congruent visual word primes compared to prosody targets preceded by incongruent word primes. RM-ANOVA revealed a significant 2-way *Prime Valence* × *Target Valence* interaction for reaction time [$F(1,30) = 13.1$, $p < 0.001$], indicating a behavioral affective priming effect for congruent prosody targets.

Table 3. Reaction times (means and standard deviations) in response to positive and negative targets in congruent compared to incongruent conditions in experiment 1 (affective categorization) and experiment 2 (non-affective categorization).

Experiment 1

		Congruent Conditions		Incongruent Conditions	
Condition	**Target Valence**	**Mean (ms)**	**SD (ms)**	**Mean (ms)**	**SD (ms)**
Music Target	Positive	589.39	84.24	625.66	85.31
	Negative	615.21	80.23	644.66	100.33
Prosody Target	Positive	691.13	128.89	736.66	129.28
	Negative	700.00	140.29	715.99	122.89
Music Prime	Positive	578.85	63.91	601.08	64.61
	Negative	591.26	71.13	586.50	53.58
Prosody Prime	Positive	612.42	94.11	637.85	96.52
	Negative	628.72	90.34	631.97	96.28

Experiment 2

		Congruent Conditions		Incongruent Conditions	
Condition	**Target Valence**	**Mean (ms)**	**SD (ms)**	**Mean (ms)**	**SD (ms)**
Music Target	Positive	707.68	122.03	697.59	116.91
	Negative	771.69	128.75	780.85	128.68
Prosody Target	Positive	903.98	146.82	890.03	155.47
	Negative	913.24	162.15	919.63	144.12
Music Prime	Positive	599.98	78.15	606.10	81.65
	Negative	617.38	81.84	614.21	80.29
Prosody Prime	Positive	651.00	97.18	650.57	91.87
	Negative	668.49	94.01	660.45	98.69

Table 4. Frequency of affective priming effects in the 32 participants of experiment 1. Mean and standard deviations (SD) of the difference in reaction times between affectively congruent and incongruent conditions are shown.

Condition	**Frequency (%)**	**Mean (ms)**	**SD (ms)**
Music Target	87.50	41.31	29.44
Prosody Target	75.00	47.98	42.16
Music Prime	56.25	26.56	20.71
Prosody Prime	75.00	23.71	13.58

Table 5. Frequency of affective priming effects in the 49 participants of experiment 2. Mean and standard deviations (SD) of the difference in reaction times between affectively congruent and incongruent conditions are shown.

Condition	**Frequency (%)**	**Mean (ms)**	**SD (ms)**
Music Target	55.00	13.35	14.32
Prosody Target	41.00	25.55	20.56
Music Prime	57.00	11.25	12.67
Prosody Prime	55.00	25.84	19.85

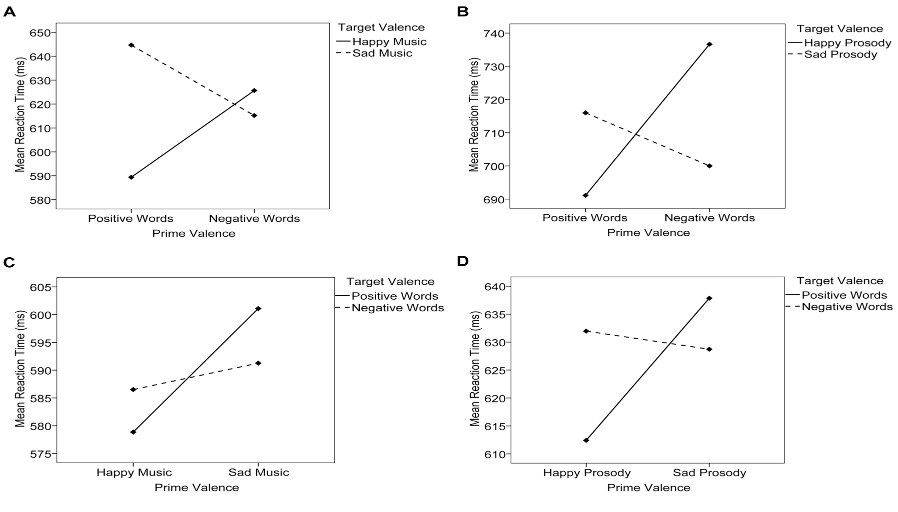

Figure 3. Behavioral results for experiment 1. Shown are reaction times for prime valence and target valence for MusicTarget (A), ProsodyTarget(B), MusicPrime (C) and ProsodyPrime (D).

MusicPrime

There was a trend to evaluate visual word targets faster when preceded by affectively congruent music primes compared to incongruent music primes. RM-ANOVA revealed a trend of the *Prime Valence* × *Target Valence* interaction for reaction time [$F(1,30) = 3.4$, $p = 0.073$]. Music excerpts as primes elicited behavioral affective priming effects in only 56 % of the participants (Table 4).

ProsodyPrime

When preceded by affectively congruent prosody primes, visual words were evaluated significantly faster than words preceded by affectively incongruent prosody primes. RM-ANOVA revealed a significant 2-way *Prime Valence* × *Target Valence* interaction for reaction time [$F(1,30) = 14.6$, $p < 0.001$]. Table 4 shows the frequency of affective priming effects in the four conditions of experiment 1.

3.1.2 Experiment 2: Non-Affective Categorization

MusicTarget

Repeated measures ANOVA revealed that the 2-way *Prime Valence* × *Target Valence* interaction was not significant for reaction time [$F(1,47) < 1$], indicating the absence of an affective priming effect for music targets preceded by word primes during non-affective semantic/phonological categorization.

ProsodyTarget

RM-ANOVA revealed a non-significant 2-way *Prime Valence* × *Target Valence* interaction for RT [$F(1,47) < 1$], indicating the absence of an affective priming effect for prosody targets preceded by word primes during semantic/phonological categorization.

MusicPrime

No affective priming effect during semantic/phonological categorization was found for word targets preceded by music primes, as a non-significant 2-way *Prime Valence* × *Target Valence* interaction [$F(1,47) < 1$] indicated.

ProsodyPrime

No affective priming effect was found during semantic/phonological categorization of target words preceded by prosody primes. RM-ANOVA revealed a non-significant 2-way *Prime Valence* × *Target Valence* interaction for RT [$F(1,47) = 1.7$, $p = 0.200$]. Table 5 shows that the frequency of affective priming effects in the four conditions of experiment 2 was at chance level.

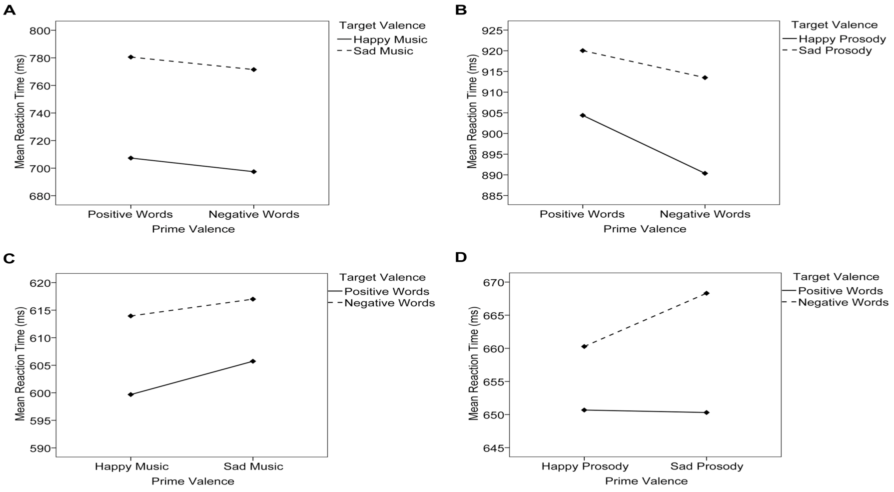

Figure 4. Behavioral results for experiment 2. Shown are reaction times for prime valence and target valence for MusicTarget (A), ProsodyTarget(B), MusicPrime (C) and ProsodyPrime (D).

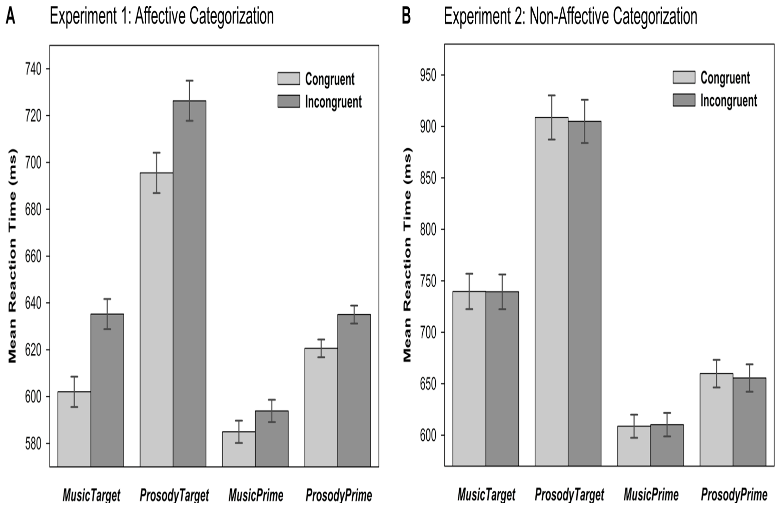

Figure 5. Behavioral results for experiment 1 (affective categorization) vs. experiment 2 (non-affective categorization). Shown are the reaction times for correct responses for congruent and incongruent conditions for prosody and music as targets and as primes.

3.2. ERP Results

First, a full RM-ANOVA was conducted on mean voltages within the N400 time-window including the factors *Condition* (*MusicTarget* vs *ProsodyTarget* vs *MusicPrime* vs *ProsodyPrime*), *Prime Valence* (positive vs negative), *Target Valence* (positive vs negative), *Region* (anterior vs central vs posterior), and *Hemisphere* (left vs right), with *Sex* (male vs female) and *Experiment* (1 vs 2) as between-subjects factors. Mean N400 amplitudes served as the dependent measure.

Results showed that the prime-target interactions significantly differed between Experiment 1 (see Figure 6) and experiment 2 (see Figure 7) at the N400 time-window, indicated by a significant 3-way interaction *Prime Valence* × *Target Valence* × *Experiment* [$F(1,77) = 11.39$, $p < 0.001$]. This interaction was further qualified by the factor *Condition*, as suggested by a significant 4-way interaction *Prime Valence* × *Target Valence* × *Experiment* × *Condition* [$F(3,231) = 3.38$, $p < 0.02$]. Follow-up ANOVAs were subsequently performed for each of the four conditions in experiment 1 (affective categorization) and experiment 2 (non-affective categorization).

3.2.1. Experiment 1: Affective Categorization

MusicTarget

Repeated measures ANOVA revealed a significant 2-way *Prime Valence* × *Target Valence* interaction for music targets at the N400 time-window [$F(1,30) = 4.8$, $p = 0.036$], indicating larger negativities for incongruent compared to congruent music targets. The data also showed a significant 3-way interaction *Prime Valence* × *Target Valence* × *Sex* for music targets [$F(1,30) = 7.6$, $p = 0.010$], suggesting that this effect was stronger in female than in male participants. However, sex as between-subject factor did not reach significance [$F(1,30) < 1$]. The *Prime Valence* × *Target Valence* interaction was not qualified by region or hemisphere, suggesting a global scalp distribution of the N400-like effect. See Figure 8 for a comparison between topographies of this N400-like effect between the four conditions of experiment 1.

ProsodyTarget

A significant 2-way *Prime Valence* × *Target Valence* interaction was observed for prosody targets at the N400 time-window [$F(1,30) = 4.8$, $p = 0.036$], indicating larger negativities for incongruent compared to congruent prosody targets. The *Prime Valence* × *Target Valence* interaction was not qualified by region or hemisphere, suggesting a global scalp distribution of the N400-like effect.

MusicPrime

A significant 3-way interaction *Prime Valence* × *Target Valence*× *Region* was found for visual word targets preceded by music primes at the N400 time-window [$F(1.1,33.1) = 8.3$, $p = 0.006$], accompanied by a main effect of region [$F(1.1,33.5) = 45.7$, $p < 0.001$], while the 2-way interaction *Prime Valence* × *Target Valence* did not reach significance [$F(1,30) < 1$]. This indicates significantly larger negativities for word targets following incongruent music primes compared to congruent primes only at anterior regions, as separate ANOVAs for each region revealed: at anterior regions, the 2-way interaction *Prime Valence* × *Target Valence* was significant [$F(1,30) = 10.3$, $p = 0.003$], but not at central [$F(1,30) < 1$] and posterior regions [$F(1,30) = 2.3$, $p = 0.137$].

ProsodyPrime

A significant 2-way *Prime Valence* × *Target Valence* interaction was observed for visual word targets preceded by prosody primes at the N400 time-window [$F(1,30) = 6.6$, $p = 0.015$], indicating larger negativities for incongruent compared to congruent targets. The *Prime Valence* × *Target Valence* interaction was not qualified by region or hemisphere, suggesting a global scalp distribution of the N400-like effect.

3.2.2. Experiment 2: Non-Affective Categorization

MusicTarget

RM-ANOVA revealed a significant 2-way *Prime Valence* × *Target Valence* interaction for music targets at the N400 time-window [$F(1,47) = 6.2$, $p < 0.001$]. In contrast to experiment 1, the effect was reversed: larger negativities were found for affectively congruent compared to affectively incongruent music targets during semantic classification. A significant 3-way interaction *Prime Valence* × *Target Valence* × *Sex* [$F(2,47) = 6.2$, $p = 0.004$] indicated that this effect was significantly larger in female than in male participants. No other interactions qualified the *Prime Valence* × *Target Valence* interaction.

ProsodyTarget

In contrast to experiment 1 (affective categorization), the 2-way *Prime Valence* × *Target Valence* interaction for prosody targets was not significant at the N400 time-window [$F(1,47) = 1.1$, $p = 0.300$], indicating that the same affectively incongruent prosody targets did not elicit larger negativities during semantic/phonological categorization.

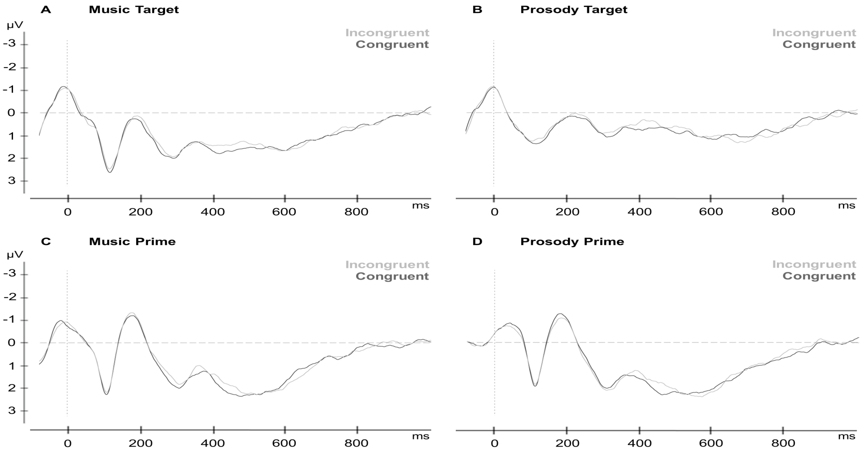

Figure 6. N400-like effect in response to affectively incongruent targets (gray) versus affectively congruent targets (black) during affective categorization in Experiment 1. Grand averages of 32 participants for a 1000 msec time-window post target onset at electrode site P3 are shown for (A) MusicTarget, (B) ProsodyTarget, (C) MusicPrime, and (D) ProsodyPrime.

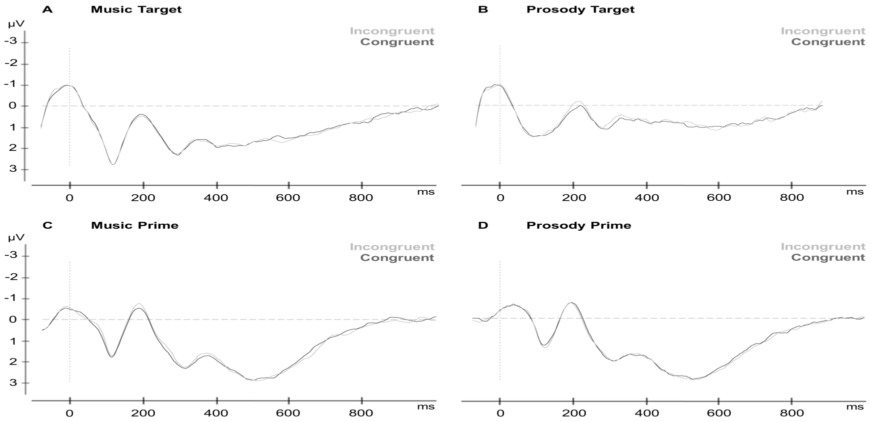

Figure 7. Lack of N400-like effect in response to affectively incongruent targets (gray) versus affectively congruent targets (black) during non-affective categorization in Experiment 2. Grand averages of 48 participants for a 1000 msec time-window post target onset at electrode site P3 are shown for (A) MusicTarget, (B) ProsodyTarget, (C) MusicPrime, and (D) ProsodyPrime.

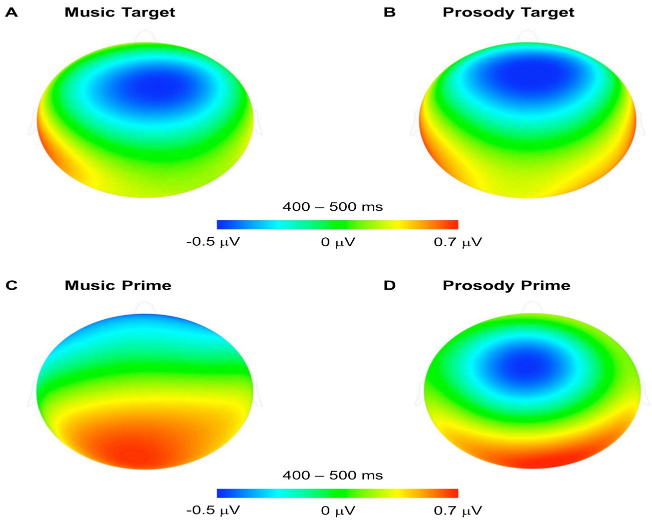

Figure 8. Topographic maps of N400-like effects during affective categorization in Experiment 1. Scalp distributions are shown for the difference waves of affectively incongruent conditions subtracted by affectively congruent conditions in the time-window 400–500 msec post target onset. N400-like effects showed a global scalp distribution for MusicTarget(A), ProsodyTarget (B), and ProsodyPrime (D), and an anterior locus of the N400-like effect for MusicPrime (C).

MusicPrime

The 2-way *Prime Valence* × *Target Valence* interaction was not significant at the N400 time-window during semantic/phonological categorization of word targets preceded by music primes [$F(1,47) < 1$].

ProsodyPrime

In contrast to experiment 1, the 2-way *Prime Valence* × *Target Valence* interaction was not significant at the N400 time-window [$F(1,47) = 1.4$, $p = 0.241$], indicating that the word targets preceded by affectively incongruent prosody primes did not elicit larger negativities during non-affective categorization.

4. Discussion

The present study was designed to systematically test the mechanisms underlying cross-modal affective priming between music, speech, and visually presented words. In experiment 1, participants categorized music, prosody, and word targets on the basis of their valence (affective categorization). In this experiment, both spreading of activation and response competition may underlie the affective priming effect. In experiment 2, participants categorized the targets based on non-affective characteristics. This design rendered affective prime characteristics irrelevant to the response dimension, excluding response competition as a possible mechanism of the affective priming effect, while leaving the possibility of spreading of activation to occur.

Our results revealed a consistent pattern at the behavioral and the electrophysiological level. During affective categorization (experiment 1), both music and prosody targets were evaluated faster when preceded by affectively congruent word primes, and vice versa. This affective priming effect was observed for prosodic primes and targets as well as for music targets. Significantly larger N400-like effects were observed for incongruent compared to congruent targets during affective categorization in each of the four conditions of experiment 1. In contrast, during non-affective categorization of the same stimuli in experiment 2, an affective priming effect was neither observed at the behavioral nor at the electrophysiological level in any of the four conditions.

The results of experiment 1 confirm previous findings of priming effects between musical and linguistic stimuli (Daltrozzo & Schoen, 2008, 2009; Sollberger et al., 2002; Steinbeis & Koelsch, 2008, 2009). Affective priming effects between music, prosody, and visually presented words at the behavioral level were accompanied by significantly larger negativities at the N400 time-window for affectively incongruent versus congruent targets during affective categorization.

The role of response conflict

Previous findings of affective priming effects accompanied by negativities at the N400 time-window have often been interpreted to reflect interference at the conceptual level, i.e. in terms of the spreading of activation account (Daltrozzo & Schoen, 2008, 2009; Steinbeis & Koelsch, 2008, 2009; Zhang, Lawson, Guo, & Jiang, 2006). Indeed, in a previous cross-modal affective priming study by Steinbeis and Koelsch (2008) using functional magnetic resonance imaging (fMRI), words preceded by affectively incongruent music chords elicited activity in the right medial temporal gyrus (MTG), and music chords preceded by affectively incongruent words were found to be related to activity in the right posterior superior temporal sulcus (STS), areas that are associated with semantic processing and which have been found in ERP studies using source localization to identify the locus of the N400 (Khateb, Pegna, Landis, Mouthon, & Annoni, 2010; Koelsch et al., 2004). In contrast, the N450 observed in ERP studies employing Stroop paradigms has been suggested to arise in the anterior cingulate cortex, an area related to error processing and conflict monitoring by studies using source localization techniques (Hanslmayr et al., 2008; Liotti et al., 2000; Szűcs and Soltész, 2010).

However, a body of empirical evidence is accumulating that argues against spreading of activation to be the sole cause of affective priming effects. The occurrence of affective priming seems to depend on the nature of the task employed: when participants are asked to evaluate targets based on their valence, affective priming effects are readily observed (e.g., De Houwer et al., 2002, Klauer & Musch, 2002, Klinger et al., 2000). When asked to evaluate targets based on non-affective features or to pronounce the targets, however, affective priming is less reliably found (e.g., De Houwer, Hermans, & Eelen, 1998, Klauer & Musch, 2001, Spruyt et al., 2004; but see De Houwer & Randell, 2004; Hermans et al., 2004; Spruyt et al., 2007a, for positive evidence of affective priming in such tasks). Task-induced modulation of affective priming was also reported by Daltrozzo and Schoen (2009) using musical and linguistic stimuli. These authors observed that affectively incongruent targets elicited a much smaller N400 effect during a more implicit lexical decision task than during affective categorization.

Taken together, the results of these studies suggest that spreading of activation may not be the only mechanism underlying affective priming, but that conflict at the response level may contribute to priming effects in the affective evaluation task. Bartholow and colleagues (2009) were the first to directly show with electrophysiological measurements that response competition contributes to affective priming between words in the evaluative categorization task by demonstrating the occurrence of preferred response activation after prime onset in motor cortex. Confirming a contribution of response competition during evaluative categorization, Eder and coworkers (2011) concluded that response priming as well as semantic priming contributes to affective priming effects between words and pictures, as indicated by an earlier occurring stimulus-locked LRP in affective congruent conditions and a larger N400 in affective incongruent conditions, respectively. Our findings of N400-like effects for music, prosody, and word targets in a task allowing for response conflict (experiment 1) but not in a task eliminating response conflict (experiment 2) are in line with a role of response conflict during affective priming.

The N400 and the N450

Affective priming tasks such as used in the present study are quite similar to stimulus-response compatibility tasks such as the Stroop task, which induces a high level of response competition (e.g., De Houwer, 2003; Klauer et al., 1997; for a review, see Klauer et al., 2011). In the classical Stroop paradigm, participants are asked to name the color a word is printed in. A mismatch between the color word and the color it is printed in slows down reaction times (Stroop effect). Interestingly, a number of ERP studies employing Stroop paradigms have also reported negativities at the N400 time-window (Rebai, Bernard, & Lannou, 1997; Liotti, Woldorff, Perez, & Mayberg, 2000; Hanslmayr, Pastötter, Bäuml, Gruber, Wimber, & Klimesch, 2008; Szűcs & Soltész, 2007, 2010). While those negativities have sometimes been interpreted as N400 effects (e.g., Rebai, Bernand, & Lannou, 1997), the authors of these studies tend to interpret them as a different effect, the N450, which has been proposed to reflect conflict at the response level (Qiu, Luo, Wang, Zhang, & Zhang, 2006; Szűcs & Soltész, 2007, 2010; West & Alain, 2000). ERP studies using source localization techniques corroborated this conclusion by showing that the source of the N450 is the anterior cingulate cortex, an area known to be involved in conflict monitoring and error processing (Hanslmayr et al., 2008; Liotti et al., 2000).

In sum, even though a contribution of conceptual level interaction cannot be excluded in those studies, the results of ERP studies employing Stroop paradigms provide indirect evidence for negativities at the N400 time-window being sensitive to conflict at the response level. This is in line with the present finding of N400-like effects in a task allowing for response conflict (experiment 1), but an absence of such effects when eliminating response conflict as a contributing factor to affective priming (experiment 2).

Given previous findings of conflict at the response level as a contributor to affective priming in the evaluative categorization task (Bartholow et al., 2009; Eder et al., 2011) on the one hand and negativities in the N400 time-window elicited by response conflict in Stroop tasks on the other hand, the question emerges whether negativities in the N400 range during affective priming reflect interference at the conceptual level (N400 interpretation), whether they are caused by conflict at the response level (N450 interpretation), or whether both mechanisms contribute to such negativities. The negativities in the N400 range observed in experiment 1 of the current study do not correspond to the centro-parietal (sometimes more right-hemispheric) topographic distribution of the classical N400 (e.g., Kutas & Federmaier, 2011) but show a global topography, indicated by the fact that neither the factor *Region* nor the factor *Hemisphere* qualified the *Prime Valence* × *Target Valence* interaction. For musical excerpts presented as primes, however, the N400-like effect had an anterior locus, as suggested by a significant 3-way interaction *Prime Valence* × *Target Valence* × *Region* (see Figure 8 for scalp topographies of the N400-like effects for the four conditions).

The current results do not allow for an unambiguous classification of the observed negativities in the N400 range (experiment 1) as corresponding to the classical N400 effect or to an N450. For this reason, we refer to the negativities observed here as an 'N400-like effect'.

Reversed priming effects

The only condition in experiment 2 that elicited significant differences at the N400 time-window was *MusicTarget*. In contrast to the negativities observed during affective categorization in experiment 1, this negativity pattern was reversed: larger negativities were found in response to affectively congruent conditions as compared to incongruent ones. This effect was found to be larger in female participants and did not surface at the behavioral level. Reversed N400-like priming effects such as observed in experiment 2 of the current study have been reported previously (Bermeitinger et al., 2008; Paulmann & Pell, 2010). These effects have been interpreted in the light of the center-surround inhibition theory (Carr & Dagenbach, 1990), arguing that briefly presented primes only weakly activate the concept associated with the prime; in order to increase activation of the prime concept surrounding concepts become inhibited, which leads to hampered access of the related targets, reversing the priming effect.

However, considering that the reversed N400-like effect in the present study occurred only during non-affective categorization (experiment 2) of music instruments (participants decided whether a music excerpt was played on the piano or with the guitar), and that the effect was significantly larger in female than in male participants, another explanation appears to be more plausible. Gender differences have been frequently observed in semantic tasks such as object decision or semantic fluency tasks (e.g., Barbarotto et al., 2002 for biological versus man-made objects; Capitani et al., 1999 for naming fruits versus tools; Laws, 2004 for tools and vehicles). Specifically, females seem to have a processing advantage for natural objects, whereas males show an advantage for artifactual objects (e.g., Laws, 1999; Laws & Hunter, 2006). Bermeitinger and colleagues (2008) tested this gender difference for natural versus artifactual objects in a semantic priming paradigm. In two experiments, they found that female participants showed positive priming effects for natural categories but reversed priming effects for artifactual categories. Males, however, showed positive priming for both natural and artifactual categories. A third experiment further showed that this priming pattern in females could be manipulated by focusing their attention on perceptual versus functional features. The authors interpreted these results as evidence for specific default processing modes that differ between females and males. Such a difference in processing modes could account for the reversed N400-like priming effects during the categorization of music instruments (= artifactual categories) observed in experiment 2 of the present study, which we found to be significantly larger in female compared to male participants. However, this interpretation remains speculative until future research replicates such reversed priming effects in the different genders during music processing.

Limitations

An important limitation of the current study is that the presence of response competition was confounded by attention directed to the affective dimension of the targets. Attentional factors have indeed been shown to influence affective priming effects in the evaluative categorization task (Gawronski et al., 2010). Therefore, we cannot exclude the possibility that in experiment 2, the absence of attention to the affective dimension prevented affective processing of the targets, undermining affective priming effects to occur (see, for example, Spruyt et al., 2007a, 2009). Future studies should attempt to control for the amount of attention devoted to the affective dimension to elucidate to what extent attention influences cross-modal affective priming between music and language and the accompanying N400-like effects.

Although our results quite consistently showed effects for experiment 1 but not for experiment 2, we inferred a contribution of response competition as a mechanism driving affective priming from the absence of an effect, rendering our evidence indirect. Though our findings are in line with recent studies that have found direct evidence for response competition during evaluative categorization, it may well be the case that both interference at the conceptual level and at the response level could contribute to affective priming effects. Future studies should investigate the relative contribution of each mechanism to affective priming.

Further, on the basis of the current data it remains unclear whether the observed N400-like effect for affective incongruence between music, speech, and words with emotional connotations resembles more closely the classical N400 effect or the N450 effect observed in ERP studies using Stroop paradigms. Future studies should use neuroimaging methods that allow localization of the brain regions mediating affective incongruency in music, prosody, and linguistic stimuli.

Conclusions

The results of the present study support the notion that affective music and speech prosody are capable of interfering with the processing of words with affective connotations, and vice versa. Affective incongruence seems to be associated with N400-like effects during evaluative categorization. Our findings further suggest a role of response competition during the affective categorization of music, prosody, and words with emotional connotations.

ACKNOWLEDGEMENTS

Niels O. Schiller was supported as a Fellow-in-Residence 2010/11 at the Netherlands Institute for Advanced Study (NIAS). We thank Jos Pacilly for designing computer scripts for this study, and Linda Geerligs for her support in ERP data analysis. This work was supported by a European Science Foundation EURYI grant (NWO no. 044035001) awarded to André Aleman.

5. References

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, DC: Hemisphere Press.

Anderson, J. R. (1983). Spreading activation. In J. R. Anderson & S. M. Kosslyn (Eds.), *Tutorials in learning and memory: Essays in honor of Gordon H. Bower* (pp. 61-90). San Francisco: Freeman

Arnold, M. B. (1960). *Emotion and Personality Vol. 1: Psychological Aspects*. New York: Columbia University Press.

Barbarotto, R., Laiacona, M., Macchi, V., & Capitani, E. (2002). Picture reality decision, semantic categories and gender: A new set of pictures, with norms and an experimental study. *Neuropsychologia, 40*, 1637-1653.

Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditionally automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology, 32*, 104-128.

Bartholow, B., Riordan, M., Saults, J., & Lust, S. (2009). Psychophysiological evidence of response conflict and strategic control of responses in affective priming. *Journal of Experimental Social Psychology*, *45*, 655-666.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.

Bermeitinger, C., Frings, C., & Wentura, D. (2008). Reversing the N400: Event-related potentials of a negative semantic priming effect. *NeuroReport, 19*, 1479-1482.

Bermeitinger, C., Wentura, D., & Frings, C. (2008). Nature and facts about natural and artifactual categories: Sex differences in the semantic priming paradigm. *Brain and Language, 106*, 153-163.

Boersma, P. & D. Weenink (1996). *Praat, a system for doing phonetics by computer* [Computer program].

Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology*, *41*, 259-268.

Capitani, E., Laiacona, M., & Barbarotto, R. (1999). Gender affects retrieval of certain categories in semantic fluency tasks. *Cortex, 35*, 273-278.

Carr, T. H., & Dagenbach, D. (1990). Semantic priming and repetition priming from masked words: Evidence for a center-surround attentional mechanism in perceptual recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16*, 341-350.

Daltrozzo, J., & Schön, D. (2008). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience, 21*, 1882-1892.

De Houwer, J., Hermans, D., & Eelen, P. (1998). Affective and identity priming with episodically associated stimuli. *Cognition and Emotion*, *12*, 145-169.

De Houwer, J., Hermans, D., & Spruyt, A. (2001). Affective priming of pronunciation responses: Effects of target degradation. *Journal of Experimental Social Psychology*, *37*, 85-91.

De Houwer, J., Hermans, D., Rothermund, K., & Wentura, D. (2002). Affective priming of semantic categorisation responses. *Cognition and Emotion*, *16*, 643-666.

De Houwer, J., & Randell, T. (2004). Robust affective priming effects in a conditional pronunciation task: Evidence for the semantic representation of evaluative information. *Cognition and Emotion, 18*, 251-264.

Eder, A. B., Leuthold, H., Rothermund, K., & Schweinberger, S. R. (2011). Automatic response activation in sequential affective priming: an ERP study. *Social Cognitive and Affective Neuroscience, 7,* 436-445.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, *15*, 115-141.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C, & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229-238.

Ferguson, M. J., & Bargh, J. A. (2003). The constructive nature of automatic evaluation. In: J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp.169-188). NJ: Erlbaum.

Gawronski, B, Cunningham, W. A., LeBel, E. P., & Deutsch, R. (2010). Attentional influences on affective priming: Does categorization influence spontaneous evaluations of multiply categorisable objects? *Cognition & Emotion, 24*, 1008-1025.

Hanslmayr, S., Pastötter, B., Bäuml, K., Gruber, S., Wimber, M., & Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. *Journal of Cognitive Neuroscience*, *20*, 215-225.

Hermans, D., Spruyt, A., De Houwer, J., & Eelen, P. (2003). Affective priming with subliminally presented pictures. *Canadian Journal of Experimental Psychology, 57*, 97-114.

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition and Emotion*, *15*, 143-165.

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception, 23*, 319-29.

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

Khateb, A., Pegna, A. J., Landis, T., Mouthon, M. S., & Annoni, J. M. (2010). On the origin of the N400 effects: An ERP waveform and source localization analysis in three matching tasks. *Brain Topography, 23*, 311-320.

Klauer, K.C., Roßnagel, C., & Musch, J. (1997). List-context effects in evaluative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 246-255.

Klauer, K., & Musch, J. (2001). Does sunshine prime loyal? Affective priming in the naming task. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *54A*, 727-751.

Klauer, K., & Musch, J. (2002). Goal-dependent and goal-independent effects of irrelevant evaluations. *Personality and Social Psychology Bulletin*, *28*, 802-814.

Klauer, K., & Musch, J. (2003). Affective priming: Findings and theories. In J Musch & K. C. Klauer (Eds.). *The psychology of evaluation: Affective processes in cognition and emotion* (7-49). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Klinger, M. R., Burton, P. C., & Pitts, G. S. (2000). Mechanisms of unconscious priming: I. Response competition, not spreading activation. *Journal of Experimental Psychology, 26,* 441-455.

Koelsch, S. (2011). Toward a neural basis of music perception - a review and updated model. *Frontiers in Psychology*, *2*, 1-20.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience, 7*, 302-307.

Kotz, S., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, *1151,* 107-118.

Kutas, M., & Federmeier K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology, 62*, 621-627.

Laws, K. R. (1999). Gender affects naming latencies for living and nonliving things: Implications for familiarity. *Cortex, 35*, 729-733.

Laws, K. R. (2004). Sex differences in lexical size across semantic categories. *Personality and Individual Differences, 36*, 23-32.

Laws, K. R., & Hunter, M. Z. (2006). The impact of colour, spatial resolution, and presentation speed on category naming. *Brain and Cognition, 62*, 89-97.

Liotti, M., Woldorff, M., Perez, R., & Mayberg, H. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia*, *38*, 701-711.

Paulmann S., & Pell, M. D. (2010). Contextual influences of emotional speech prosody on face processing: how much is enough? *Cognitive, Affective, & Behavioral Neuroscience, 10*, 230-242.

Qiu, J., Luo, Y., Wang, Q., Zhang, F., & Zhang, Q. (2006). Brain mechanism of Stroop interference effect in Chinese characters. *Brain Research, 1072*, 186-193.

Rebai, M., Bernard, C., & Lannou, J. (1997). The Stroop's test evokes a negative brain potential, the N400. *International Journal of Neuroscience, 91*, 85-94.

Schirmer, A., Kotz, S., & Friederici, A. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*, 228-233.

Schirmer, A., & Kotz, S. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *Journal of Cognitive Neuroscience*, *15*, 1135-1148.

Schirmer, A., Kotz, S., & Friederici, A. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research*, *24*, 442-452.

Sollberger, B., Reber, R., & Eckstein, D. (2003). Musical chords as affective priming context in a word-evaluation task. *Music Perception*, *20*, 263-282.

Spruyt, A., Hermans, D., Houwer, J., & Eelen, P. (2002). On the nature of the affective priming effect: Affective priming of naming responses. *Social Cognition*, *20*, 227-256.

Spruyt, A., Hermans, D., Pandelaere, M., De Houwer, J., & Eelen, P. (2004). On the replicability of the affective priming effect in the pronunciation task. *Experimental Psychology*, *51*, 109-115.

Spruyt, A., De Houwer, J., Hermans, D., & Eelen, P. (2007a). Affective priming of nonaffective semantic/phonological categorization responses. *Experimental Psychology*, *54*, 44-53.

Spruyt, A., Hermans, D., De Houwer, J., Vandromme, H., & Eelen, P. (2007b). On the nature of the affective priming effect: Effects of stimulus onset asynchrony and congruency proportion in naming and evaluative categorization. *Memory & Cognition*, *35*, 95-106.

Spruyt, A., De Houwer, J., & Hermans, D. (2009). Modulation of automatic semantic priming by feature-specific attention allocation. *Journal of Memory and Language*, *61*, 37-54.

Steinbeis, N., & Koelsch, S. (2008). Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. *PLoS ONE, 3,* e2226.

Steinbeis, N., & Koelsch, S. (2011). Affective priming effects of musical sounds on the processing of word meaning. *Journal of Cognitive Neuroscience, 23*, 604-621.

Szűcs, D., & Soltész, F. (2007). Event-related potentials dissociate facilitation and interference effects in the numerical Stroop paradigm. *Neuropsychologia, 45*, 3190-3202.

Szűcs, D., & Soltész, F. (2010). Stimulus and response conflict in the color-word stroop task: A combined electro-myography and event-related potential study. *Brain Research, 1325*, 63-76.

Wentura, D. (1999). Activation and inhibition of affective information: Evidence for negative priming in the evaluation task. *Cognition and Emotion*, *13*, 65-91.

Wentura, D. (2000). Dissociative affective and associative priming effects in the lexical decision task: Yes versus no responses to word targets reveal evaluative judgment tendencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 456-469.

Wundt, W. (1907). *Outlines of psychology*, trans. C. H. Judd. Leipzig: Wilhelm Engelmann.

Zajonc, R. B. (1980). Feelings and thinking: Preferences need no inferences. *American Psychologist, 35*, 151-175.

Zajonc, R. B. (1984). On the primacy of affect. *American Psychologist, 39*, 117-123.

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

SECOND EMPIRICAL PART: MODULATION BY ALEXITHYMIA

Chapter 3.1

Modulation of emotional prosody and music perception by alexithymia: ERP evidence

Contributions of first and second author:

Conceived the experiments: J. W. and K. G.
Designed the study materials and tasks: J. W.
Data collection: J. W. (pilot studies) and K .G (EEG)
Analyzed the data: J. W. (behavioral data) and K. G (EEG)
Wrote the first draft of the report: K. G.
Commented on the first draft: J. W.

Goerlich, K. S., Witteman, J., Aleman, A., & Martens, S. (2011). Hearing feelings: Affective categorization of music and speech in alexithymia, an ERP study. *PloS One, 6,* e19501.

Abstract

Alexithymia, a condition characterized by deficits in interpreting and regulating feelings, is a risk factor for a variety of psychiatric conditions. Little is known about how alexithymia influences the processing of emotions in music and speech. Appreciation of such emotional qualities in auditory material is fundamental to human experience and has profound consequences for functioning in daily life. We investigated the neural signature of such emotional processing in alexithymia by means of event-related potentials. Affective music and speech prosody were presented as targets following affectively congruent or incongruent visual word primes in two conditions. In two further conditions, affective music and speech prosody served as primes and visually presented words with affective connotations were presented as targets. Thirty-two participants (16 male) judged the affective valence of the targets. We tested the influence of alexithymia on cross-modal affective priming and on N400 amplitudes, indicative of individual sensitivity to an affective mismatch between words, prosody, and music. Our results indicate that the affective priming effect for prosody targets tended to be reduced with increasing scores on alexithymia, while no behavioral differences were observed for music and word targets. At the electrophysiological level, alexithymia was associated with significantly smaller N400 amplitudes in response to affectively incongruent music and speech targets, but not to incongruent word targets.

Our results suggest a reduced sensitivity for the emotional qualities of speech and music in alexithymia during affective categorization. This deficit becomes evident primarily in situations in which a verbalization of emotional information is required.

1. Introduction

Alexithymia (literally translated "no words for feelings") has been recognized as a major risk factor for a variety of psychopathological and medical conditions, including chronic pain, somatization, depression, and anxiety (Taylor, Bagby, & Parker, 1997). This condition is characterized by deficits in the identification and verbalization of one's feelings and the cognitive processing and regulation of emotions (Larsen, Brand, Bermond, & Hijman, 2003).

Previous research has shown that alexithymic individuals have difficulty identifying emotional facial expressions (Prkachin, Casey, & Prachin, 2008; Swart, Kortekaas, & Aleman, 2009), matching verbal and non-verbal emotional stimuli (Lane et al., 1996), and remembering words with emotional connotations (Luminet et al., 2006). Neuroimaging studies have provided additional evidence for an association of alexithymia with differences in brain activation for a variety of tasks that involve emotional processing, such as the processing of emotional pictures (Berthoz et al., 2002) and the processing of facial expressions of emotion (Mériau et al., 2006), the imagery of autobiographical emotional events (Mantani, Okamoto, Shirao, Okada, & Yamawaki, 2005), the observation of fearful body expressions (Pouga, Berthoz, De Gelde, & Grèzes, 2010), and during empathy for pain (Bird et al., 2010).

Since such impairment during the conscious processing of emotional information may be dependent upon dysfunctions at earlier processing stages, the investigation of automatic sensitivity to affective stimuli is of great importance to understanding the emotion processing deficits individuals with alexithymia exhibit. Recent studies have suggested impaired processing of emotions even at pre-attentive, automatic processing stages in this condition. When presented with emotionally aversive videos, for instance, individuals scoring high on alexithymia did not show an increase in electrodermal activity as low-scorers on alexithymia did, while no difference in self-reported arousal between high- and low-alexithymics was found (Franz, Schaefer, Schneider, Sitte, & Bachor, 2010). Smaller electrodermal responses were also found in a study using negative masked pictures, likewise suggesting a deficit in early emotional reactivity associated with alexithymia (Pollatos, Schubö, Herbert, Matthias, & Schandry, 2008). Other studies, however, report higher autonomic baseline levels in alexithymia (e.g., Gundel et al., 2004; see Lumley, Neely, & Burger, 2007 for a review). Aftanas and colleagues measured event-related synchronization (ERS) in participants watching emotional film clips. Results indicated greater emotional reactivity in the right hemisphere in high-scorers on alexithymia, suggesting enhanced negative affect and autonomous arousal associated with this condition (Aftanas & Varlamov, 2004).

Three recent studies using functional magnetic resonance imaging (Duan, Dai, Gong, & Chen, 2010; Kugel et al., 2008; Reker et al., 2010) further support the view that individuals with alexithymia show impairment during the subconscious processing of emotions. All of these studies assessed the influence of alexithymia on the automatic processing of masked facial expressions of emotions. Sad faces were found to be associated with lower responsiveness of the left (Reker et al., 2010) and right amygdala (Kugel et al., 2008) as a function of alexithymia. Additionally, Reker and colleagues found reduced activations of the insula, superior temporal gyrus, middle occipital and parahippocampal gyrus in response to sad and happy facial expressions with increasing scores on alexithymia (Reker et al., 2010). Masked surprised faces elicited decreased activation of the parahippocampal gyrus and fusiform gyrus as a function of alexithymia

(Duan et al., 2004). In sum, these studies providence evidence for a hypoactivation of brain areas related to the subconscious processing of facial emotions, suggesting that alexithymia is associated with reduced processing of automatic emotional information.

Affective Priming

A powerful technique to assess automatic processing of emotions is the affective priming paradigm. The affective priming effect refers to the observation that the affective connotation of a target stimulus, e.g., 'ugly' will be judged faster when preceded by an affectively related prime, e.g., 'hate' as compared to an affectively unrelated prime, e.g., 'love' (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). The effect is thought to be an early, fast-acting, automatic process that can occur outside of conscious awareness (Fazio, 2001; Klauer & Musch, 2003). It has been demonstrated for a variety of stimuli, such as pictures, prosody ("melody of speech"), music, and even odors.

Few studies have employed the affective priming paradigm in alexithymia. The first study to examine affective priming effects as a function of alexithymia was conducted by Suslow (Suslow, Junghanns, Donges, & Arolt, 2001). Positive and negative word targets (adjectives) primed by positive or negative words (nouns) were to be pronounced (pronunciation task) or evaluated as positive or negative as quickly as possible (evaluation task). Pearson's correlations revealed no influence of alexithymia on affective word priming during target pronunciation. During affective evaluation, however, alexithymia correlated positively with the affective priming effect for positive word targets, whereas the correlation with negative word targets failed to reach significance (Suslow, 1998).

In a follow-up study, the same word evaluation task as in the previous study (word – word prime – target pairs) was employed, and in addition a face evaluation task (face – face prime – target pairs) including happy and sad faces. The positive correlation of alexithymia with affective priming for positive word targets could not be replicated: In both the word evaluation task and the face evaluation task alexithymia did not correlate with the affective priming effect, neither for positive nor for negative targets (Suslow et al., 2001).

Suslow and Junghanns (2002) employed a lexical decision task on neutral or emotional target words and non-words primed by sentences with congruent or incongruent emotional content. High scorers on alexithymia showed a negative situation priming effect, indicated by faster lexical decisions for targets preceded by affectively incongruent primes.

Vermeulen and colleagues (Vermeulen, Luminet, & Corneille, 2006) used verbal (positive and negative words) and non-verbal (happy and sad schematic faces) as primes and targets to investigate affective priming effects in alexithymia. Regression analyses on the effect of prime type (happy and angry faces, positive and negative words) showed reduced affective priming with increasing alexithymia scores only for angry face primes, indicative of reduced emotion processing at an automatic level in alexithymia. Based on these findings, the authors suggest a specific impairment during the automatic processing of threatening stimuli (as represented by angry faces) associated with alexithymia (Vermeulen et al., 2006).

Taken together, previous studies using affective priming paradigms in alexithymia provide preliminary evidence for an impact of alexithymia on the automatic processing of visual emotional stimuli. However, no consistent picture has emerged with respect to the question of whether alexithymia is associated with reduced or increased affective priming effects during the automatic processing of emotions.

Music and Speech Prosody

An adequate processing of emotional qualities in auditory material such as emotional prosody ("melody of speech") and music is fundamental to human experience and has profound consequences for functioning in daily life. Both music and speech prosody have been shown to be capable of influencing the processing of visual emotional material (for prosody see Bostanov & Kotchoubey, 2004; Koelsch et al., 2004; Kotz & Paulman, 2007; Schirmer & Kotz, 2003; Schirmer, Kotz, & Friederici, 2002; Schirmer, Kotz, & Friederici, 2005; for music see Koelsch et al., 2004; Chen, Yuan, Huang, Chen, & Li, 2008; Daltrozzo & Schön, 2008; Logeswaran & Bhattacharya, 2009; Sollberger, Reber, & Eckstein, 2003; Steinbein & Koelsch, 2008; Steinbein & Koelsch, 2009; for a review see Marin & Bhattacharya, 2010). To the best of our knowledge, only two previous studies have addressed the impact of alexithymia on the processing of emotions conveyed by speech prosody and music. Swart and colleagues observed no behavioral differences for spoken sentences with incongruent semantics and affective prosody in individuals with alexithymia as compared to controls (Swart et al., 2009). Vermeulen and colleagues found that during the presentation of angry, but not happy background music, high scorers on alexithymia recognized fewer anger and joy words than low scorers, indicating hampered memory performance during angry music perception associated with alexithymia (Vermeulen, Toussaint, & Luminet, 2010).

In sum, the literature on the emotion processing deficit in alexithymic individuals demonstrates that alexithymia influences not only the conscious processing of emotion, but that aberrant emotion processing is evident already at very early, automatic processing stages. However, there is no consensus as to whether alexithymia is associated with decreased or increased automatic processing of emotions. In particular, little is known about the manifestation of this automatic processing deficit at the auditory processing level. We investigated the neural signature of automatic affective priming of words, speech prosody, and music as a function of alexithymia by means of event-related potentials (ERPs).

The present study is the first ERP study to investigate the neural signature of auditory emotional processing in alexithymia using affective speech prosody and music in a cross-modal priming paradigm. We hypothesized reduced affective priming effects with increasing scores on alexithymia at the behavioral level. Given the difficulty to identify emotions in alexithymia, we further hypothesized a reduced sensitivity to affective mismatches in alexithymia, reflected in diminished N400 amplitudes in response to affectively incongruent compared to congruent conditions as a function of alexithymia. The results of this study show that alexithymia was indeed associated with diminished N400 amplitudes for affective prosody and music.

2. Methods

2.1. Participants

Thirty-two students (16 male, mean age 23.8, SD 4.4) from the University of Groningen participated in the experiment. All participants were right-handed native speakers of Dutch, had normal or corrected-to-normal vision, no hearing impairment and no psychiatric condition in present or past. Participants received €20 for their participation in the two-hour EEG session. The Neuroimaging Center Institutional Review Board approved the experimental protocol and written informed consent was obtained from all participants prior to the study.

2.2. Toronto Alexithymia Scale (TAS-20)

The TAS-20 is the most widely used measure of alexithymia (Bagby, Parker, & Taylor, 1994a; Bagby, Parker, & Taylor, 1994b) with a demonstrated validity, reliability, and stability (see Taylor, 2000). A validated Dutch translation of the scale (Kooiman, Spinhoven, & Trijsburg, 2002) was used for the present study. The scale consists of 20 self-report items rated on a 5-point Likert scale (1: strongly disagree, 5: strongly agree), with five negatively keyed items.

The TAS-20 comprises the subscales: (1) difficulty identifying feelings (e.g., "I often don't know why I'm angry"), (2) difficulty describing feelings (e.g., "I find it hard to describe how I feel about people"), and (3) externally oriented thinking (e.g., "I prefer talking to people about their daily activities rather than their feelings"). Possible scores range from 20 to 100, higher scores indicate higher degrees of alexithymia.

It has been suggested that alexithymia comprises two related, but distinct types (Vorst & Bermond, 2001; but see Bagby et al., 2009), which can be assessed with another self-report questionnaire, the BVAQ (Bermond-Vorst Alexithymia Questionnaire, 45). Type I alexithymia is thought to be characterized by a general lack of responsiveness to emotion at any level, whereas in type II alexithymia, basic responses to affective stimuli are assumed to be intact, whereas the ability to cognitively access and verbalize them is impaired (Lumley et al., 2007). Note that the TAS-20 assesses only type II alexithymia. Thus, the findings presented here allow conclusions with regard to type II alexithymia but might not be applicable to type I alexithymia.

Individuals with TAS-20 scores lower or equal to 51 are considered non-alexithymic, a score from 52 to 60 indicates moderate alexithymia. The clinical threshold for alexithymia is a score of 61 (Taylor et al., 1997). Alexithymia scores of our study sample ranged from 31 to 68 (mean: 43.25, SD: 9.89, median: 41.5, skewness: 0.98).

2.3. Materials

The stimulus set consisted of 48 words for visual presentation (24 positive, 24 negative), 48 pseudo-words spoken in happy (24) and sad (24) prosody, and 48 music segments expressing happy (24) or sad (24) emotion. All stimuli were validated in three separate pilot studies prior to the experiment.

In the visual word pilot, ten independent raters of Leiden University judged the words with emotional connotations on a 9-point Likert scale (-4 = very negative, 0 = neutral, 4 = very positive). Only words rated 3 or higher by 9 out of 10 raters were included as positive word stimuli, only words rated -3 or lower by 9 out of 10 raters were included as negative word stimuli (see Table 1).

Table 1. Affective word stimuli (Dutch originals and English translation) used in experiment (identical to Table 1 of Chapter 2.5 but repeated here for ease of reference).

	Positive Words	**English Translation**	**Negative Words**	**English Translation**
1	Bloem	*Flower*	Beul	*Hangman*
2	Bonbon	*Candy*	Bom	*Bomb*
3	Cake	*Cake*	Braaksel	*Vomit*
4	Echtpaar	*Married Couple*	Dief	*Thief*
5	Expert	*Expert*	Galg	*Gallows*
6	Genie	*Genius*	Graf	*Grave*
7	Geschenk	*Present*	Hoer	*Whore*
8	Goedzak	*Good Soul*	Junk	*Junk*
9	Held	*Hero*	Klootzak	*Asshole*
10	Honing	*Honey*	Monster	*Monster*
11	Ijsje	*Ice Cream*	Pijnbank	*Rack*
12	Kanjer	*Hunk*	Pis	*Piss*
13	Lieverd	*Darling*	Pus	*Pus*
14	Maatje	*Buddy*	Racist	*Racist*
15	Paleis	*Palace*	Sadist	*Sadist*
16	Parel	*Perl*	Schijt	*Shit*
17	Ross	*Rose*	Slet	*Slut*
18	Satijn	*Satin*	Sloerie	*Slut*
19	Schatje	*Baby*	Tiran	*Tyrant*
20	Snoep	*Candy*	Tumor	*Tumor*
21	Vriend	*Friend*	Vandaal	*Vandal*
22	Vriendin	*Girlfriend*	Vergif	*Poison*
23	Winnaar	*Winner*	Vetkwab	*Fat Roll*
24	Zon	*Sun*	Viespeuk	*Dirt Bag*

For the prosody pilot, bisyllabic pseudo-words that obeyed Dutch phonotactics were recorded with the help of an actress, cut to a length of approximately 600 ms and amplitude normalized using the Praat speech processing software (Boersma & Weenink, 1996). The normalization procedure amplified every stimulus item such that the digitalized sample with the maximum amplitude was set at the maximum positive or negative value of the converter range, and all other samples were scaled proportionally. As a result, all stimuli had about equal intensity. Ten independent raters at Leiden University judged the pseudo-words on a 9-point Likert scale (−4 = very sad, 0 = neutral, 4 = very happy). Only pseudo-words rated 3 or higher for happy prosody and −3 or lower for sad prosody by 9 out of 10 raters were included in the study.

Music segments were created from a number of piano pieces by composers of Western classical music (e.g., Bach, Beethoven, Chopin). Segments with a length of 600 ms were excerpted in Praat (cut at zero-crossings), amplitude normalized and judged by 13 independent raters at the University of Groningen on a 9-point Likert scale (−4 = very sad, 0 = neutral, 4 = very happy). Only music segments rated 3 or higher for happy music and −3 or lower for sad music by 11 out of 13 raters were included in the study.

2.4. Procedure

The cross-modal affective priming paradigm included four main conditions (see Figure 1): *MusicTarget* (music targets preceded by visual word prime), *ProsodyTarget* (prosody target preceded by visual word prime), *MusicPrime* (visual word target preceded by music prime), and *ProsodyPrime* (visual word target preceded by prosody prime). Each main condition comprised two congruent and two incongruent sub-conditions (congruent: positive prime – positive target, negative prime, negative target, incongruent: positive prime – negative target, negative prime – positive target).

Each of the four main conditions (*MusicTarget, ProsodyTarget, MusicPrime, ProsodyPrime*) consisted of 96 trials. Overall, each word, prosody and music stimulus was presented twice, once congruent and once incongruent, eliminating stimulus characteristics as an explanation of priming effects. All stimuli (primes as well as targets) were presented for 600 ms. Prime – target pairs were created and presented in a randomized fashion. The four main conditions were presented in four separate blocks, the order of which was presented counterbalanced according to a Latin square.

Stimulus presentation was controlled using E-Prime version 1.2 (Schneider, Eschman, & Zuccolotto, 2002). Each trial started with a black fixation cross in the middle of the screen (1500 ms), followed by a red fixation cross (500 ms) signaling the occurrence of the prime. When the red fixation cross disappeared, the prime was presented. Two hundred ms after prime onset, the target was presented. An SOA of 200 ms was chosen based on findings that the affective priming effect dissipates after 300 ms (Kutas & Federmeier, 2011). Reaction time was recorded from the onset of the target. To reduce blink artifacts, participants were instructed to blink when the fixation cross was black, and not to blink anymore when it turned red.

The task of the participants was to judge the valence of the word targets (positive or negative) and music and prosody targets (happy or sad) as fast and accurately as possible (affective categorization). Directly after the EEG session, participants completed the TAS-20 questionnaire.

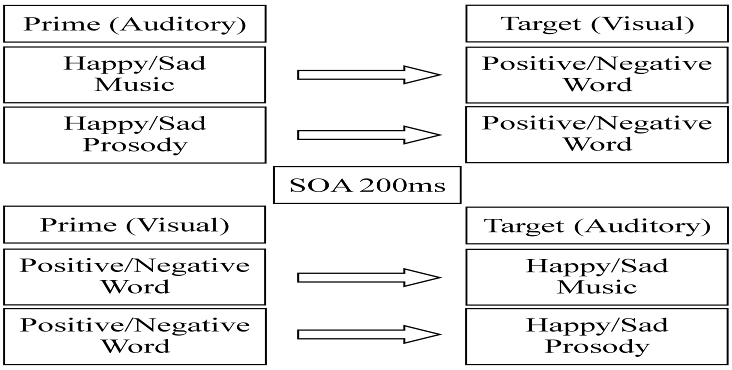

Figure 1. Design of the cross-modal affective priming paradigm.

2.5. ERP Recordings

Electroencephalogram (EEG) was recorded from 64 tin electrodes mounted in an elastic electro cap organized according to the international 10/20 system. EEG data were recorded with a linked mastoid physical reference and were re-referenced by using an average reference. Bipolar vertical and horizontal electrooculograms (EOGs) were recorded for artifact rejection purposes.

The ground electrode was applied to the sternum. Impedance of all electrodes was kept below 5 kΩ for each participant. EEG was continuously recorded with a sampling rate of 500 Hz, amplified, and off-line digitally low-pass filtered with a cut-off frequency of 30 Hz. Participants were seated in front of a monitor at a distance of approximately 50 cm in a dimly lit, electrically shielded and sound-attenuated cabin. Music and speech stimuli were presented via loudspeakers placed at the left and right side of the participant at approximately 70dB.

2.6. Behavioral Data Analysis

First, we aimed to establish the occurrence of cross-modal affective priming in each of the four experimental conditions. To this end, behavioral data were analyzed in a 2 (congruence: congruent vs. incongruent) by 2 (valence: positive vs. negative) repeated-measures analysis of variance (RM-MANOVA) with sex as a between-subjects factor. The analysis of accuracy showed that performance was higher than 90 percent in all conditions, indicating ceiling effects. Therefore, only the results of the reaction time (RT) analyses on correctly identified targets are reported (see Figure 3).

Secondly, the impact of alexithymia on affective priming was tested in a 2 (congruence: congruent vs. incongruent) by 2 (valence: positive vs. negative) repeated-measures analysis of covariance (RM-MANCOVA) with alexithymia as a covariate and sex as a between-subjects factor.

Lastly, Pearson's correlations were conducted to test the impact of alexithymia on differences in reaction time between affectively congruent and incongruent targets. In order to test for an effect of valence, alexithymia scores were further correlated with differences in reaction time for positive and negative targets separately.

2.7. ERP Data Analysis

The EEG data were analysed with Brain Vision Analyzer (version 1.05). Prior to averaging, trials with eye-movement and blink artifacts were excluded from analysis. Criteria for artifact rejection within an epoch were a maximal voltage step of 50 µV, a maximal difference between two values in a segment of 100 µV, and a minimal and maximal amplitude of -100 µV and 100 µV, respectively. A total mean number of 360.1 trials (SD 21.8) was recorded for each of the 32 participants (mean 89.2, SD 5.2 per experimental condition). Artefact rejection excluded a mean percentage of 3.4 percent of all trials (ranging from 0.3 percent to 23.3 percent across participants), leaving a total mean number of 343.7 trials (SD 27.1) for analysis, with a mean number of 85.7 trials (SD 6.7) per experimental condition.

ERP epochs for each subject were computed in a 1000 ms time-window following the onset of the targets, which were aligned to a 100 ms pre-target baseline. Visual inspection of the data revealed negativities in response to affectively incongruent compared to congruent targets between 400 and 500 ms following the onset of the targets. These negativities were found consistently between 400 and 500 ms for music and prosody targets as well as for visual word targets, indicating that regardless of modality, affectively incongruent targets elicited N400 effects in a time-window of 400 – 500 ms following target onset. Based on this observation and previous N400 literature, the time-window 400 – 500 ms post-target onset was chosen for statistical analysis. Mean amplitudes for positive and negative music, speech, and word targets were computed at the N400 time-window (400 – 500 ms after target-onset) in each participant, beginning at the onset of the targets. RM-MANOVA was conducted in SPSS (17.0) using a total of 30 electrodes in six topographic regions (anterior, central, posterior) in the left and right hemisphere (see Figure 2). The left anterior region included electrodes F3, F5, F7, FC3, and FC5, the right anterior region electrodes F4, F6, F8, FC4, and FC6. The left central region included electrodes C3, C5, CP3, CP5, and T7, the right central region electrodes C4, C6, CP4, CP6, and T8. The left posterior region included electrodes P3, P5, P7, PO3, and PO7, the right posterior region electrodes P4, P6, P8, PO4, and PO8.To test for an effect of affective congruence (i.e., affective priming) between primes and targets as well as for effects of valence, congruence and valence were entered into the analysis as separate factors. Topographic region and hemisphere were additionally included as within-subject factors. Based on previous findings of sex differences in emotional prosody processing (Schirmer & Kotz, 2003; Schirmer et al., 2002; Schirmer et al., 2005) and the processing of emotions conveyed by music (Altemüler, Schürmann, Lim, & Parlitz, 2002; Flores-Gutiérrez et al., 2009; Nater, Abbruzzese, Krebs, & Ehlert, 2006), sex was included as a between-subjects factor. In case of sphericity violations, Greenhouse-Geisser corrected p-values are reported. A Sidak correction of p-values was used in pairwise comparisons between the levels of factors.

Secondly, in order to test for the impact of alexithymia on affective priming and valence of primes and targets, RM-MANCOVA was carried out using the same factors as above and additionally including scores on the alexithymia questionnaire TAS-20 as a covariate.

Lastly, as in previous studies on affective priming in alexithymia (Suslow, 1998; Suslow et al., 2001; Vermeulen et al., 2006), correlation analyses were conducted to test the impact of alexithymia on N400 amplitudes in response to an affective mismatch between primes and targets. Given that the N400 reflects the processing and integration of meaning (for a recent review, see Kutas & Federmeier, 2011), its amplitude can be used as an indicator of individual sensitivity to mismatches in affective meaning between stimuli such as music, prosody, and words. To obtain an index of the relative increase in N400 amplitudes in affectively incongruent compared to congruent conditions, N400 mean voltages following congruent target onsets were subtracted from N400 mean voltages following incongruent target onsets at each electrode site.

In order to reduce the number of comparisons and thereby the probability of false positives, correlations were not carried out at the 30 electrode sites separately, but N400 means of the five electrodes contained in each of the six topographic regions (left anterior, central, posterior; right anterior, central, posterior) were collapsed. The resulting N400 amplitude means for the six regions were used in subsequent correlation analyses.

Standardized alexithymia scores (31 - 68, mean: 43.25, SD: 9.89) were then correlated with the absolute difference in N400 amplitude means between affectively incongruent and congruent conditions. In order to test for effects of valence, Pearson's correlations with alexithymia were further conducted separately for positive and negative targets. For this purpose, separate N400 means for positive and negative targets (indexes of valence effects) were obtained by subtracting N400 means for positive (negative) targets in congruent conditions from N400 means for positive (negative) targets in incongruent conditions.

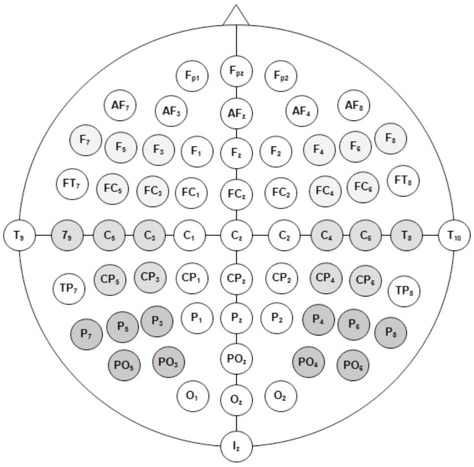

Figure 2. Map of electrode sites used for analysis with left and right anterior, central, and posterior regions identified

3. Results

3.1. Behavioral Results

MusicTarget

RM-MANOVA revealed a significant affective priming effect for music targets primed by words with affective connotations. Participants evaluated music segments preceded by affectively congruent visual word primes significantly faster than music segments preceded by affectively incongruent word primes, as indicated by a main effect of congruence [$F(1,30) = 27$, $p < 0.01$]. A main effect of valence [$F(1,30) = 11.1$, $p < 0.05$] revealed faster categorization of happy music targets compared to sad music targets (608 ms vs. 630 ms). Further, a main effect of sex was found [$F(1,30) = 4.2$, $p = 0.05$], indicating that female participants categorized affective music targets significantly faster than male participants (590 ms vs. 647 ms).

After controlling for alexithymia in RM-MANCOVA, the effect of congruence [$F(1,29) = 6.6$, $p < 0.05$] and sex [$F(1,29) = 4.2$, $p = 0.05$] remained significant. However, the effect of valence failed to reach significance [$F(1,29) = 2.2$, $p = 0.15$]. Alexithymia as a between-subjects effect was not significant [$F(1,29) < 1$], and no interactions with alexithymia were observed.

Pearson's correlations revealed no significant impact of alexithymia on the behavioral affective priming effect for music targets preceded by visual word primes ($r = -.24$, $p = 0.19$). There was no effect of valence ($r = -.15$, $p = 0.42$).

ProsodyTarget

RM-MANOVA demonstrated a significant affective priming effect for prosody targets primed by words with affective connotations. Prosody targets were evaluated significantly faster when preceded by affectively congruent as opposed to affectively incongruent visual word primes, as indicated by a main effect of congruence for reaction time [$F(1,30) = 13.1$, $p < 0.01$]. No main effect of valence was observed [$F(1,30) < 1$]. A significant two-way interaction between congruence and valence showed that sad prosody was categorized significantly faster than happy prosody in affectively incongruent, but not congruent conditions (716 ms vs 737 ms, $p < 0.01$).

Including alexithymia as a covariate in RM-MANCOVA showed that the effect of congruence remained significant [$F(1,29) = 6.7$, $p < 0.05$]. A trend toward an alexithymia × congruence interaction [$F(1,29) = 3.3$, $p = 0.08$] suggested that this affective priming effect tended to be reduced in individuals with higher alexithymia scores. As between-subjects effect, alexithymia was not significant [$F(1,29) < 1$]. No main effect of valence was observed [$F(1,29) = 1.6$, $p = 0.21$], and the alexithymia × valence interaction failed to reach significance [$F(1,29) = 1.3$, $p = 0.27$].

Correlation analyses confirmed a trend toward a negative correlation between alexithymia and reaction times for prosody targets preceded by visual word primes ($r = -.30$, $p = 0.09$), suggesting a trend toward reduced affective priming with increasing alexithymia scores. No correlation between alexithymia and the valence of prosodic targets was found ($r = -.22$, $p = 0.23$).

MusicPrime

RM-MANOVA showed a trend to categorize affective word targets faster when preceded by affectively congruent vs. incongruent music primes (i.e., affective priming effect) as suggested by a marginally significant effect of congruence [$F(1,30) = 3.4$, $p = .07$]. A main effect of valence was not observed [$F(1,30) < 1$]; however, a significant interaction between valence and congruence [$F(1,30) = 12.9$, $p < 0.01$] suggested that affective priming by music on words was stronger for positive word targets. The effect of sex was not significant [$F(1,30) < 1$].

RM-MANCOVA including alexithymia as a covariate showed no significant main effects or interactions for word targets preceded by music primes.

Correlation analyses confirmed the absence of an effect of alexithymia on affective priming in this condition: no significant correlations were observed between alexithymia and affective congruence ($r = -.08$, $p = 0.66$) and the valence of affective words ($r = -.23$, $p = 0.21$).

ProsodyPrime

RM-MANOVA revealed a significant affective priming effect for word targets primed by emotional prosody. Words with emotional connotations were evaluated significantly faster when preceded by affectively congruent as opposed to affectively incongruent prosody primes, as indicated by a main effect of congruence [$F(1,30) = 14.6$, $p < 0.01$]. There was no main effect of valence [$F(1,30) < 1$] and sex [$F(1,30) = 1.5$, $p = 0.23$]. A significant congruence × valence interaction $F(1,30) = 12.4$, $p < 0.01$] showed that the affective priming effect was stronger for positive than for negative words. This effect tended to be qualified by sex differences: a marginally significant three-way congruence × valence × sex interaction [$F(1,30) = 3.8$, $p = 0.06$] suggested that in female participants, affective priming of words was evident regardless of valence, whereas male participants showed affective priming only for positive word targets.

When including alexithymia as a covariate (RM-MANCOVA), the affective priming effect was only marginally significant [congruence: $F(1,30) = 3.8$, $p = 0.06$]. No further main effects were observed in this condition. Alexithymia did not reach significance as between-subjects effect [$F(1,30) < 1$] and did not interact with congruence [$F(1,30) = 1.3$, $p = 0.27$] or valence $F(1,30) < 1$]. The three-way interaction congruence × valence × sex remained marginally significant [$F(1,30) = 3.8$, $p = 0.06$].

Correlation analyses confirmed the absence of an effect of alexithymia on words primed by emotional prosody: no significant correlations were observed between alexithymia and affective congruence ($r = -.22$, $p = 0.24$) and the valence of affective words ($r = .04$, $p = 0.84$).

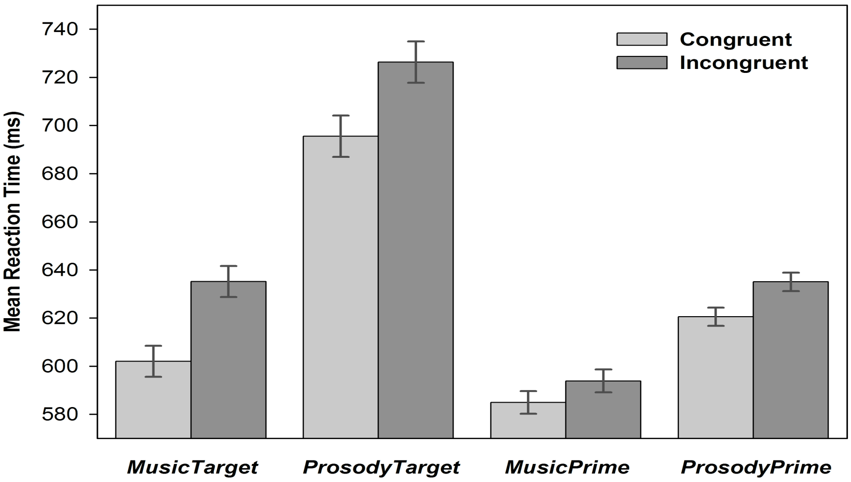

Figure 3. Behavioral affective priming effects. Behavioral affective priming effects during affective categorization of the targets. *MusicTarget*, p < 0.01, *ProsodyTarget*, p < 0.01, *MusicPrime*, p = 0.07, *ProsodyPrime*, p < 0.01. Error bars indicate 95% confidence intervals.

3.2. ERP Results

MusicTarget

RM-MANOVA revealed a main effect of congruence at the N400 time-window [$F(1,30) = 4.8$, $p = 0.04$], indicating a larger N400 for incongruent compared to congruent music targets (see Figure 4 for all conditions). There was no main effect of valence [$F(1,30) < 1$]. A significant congruence × sex interaction [$F(1,30) = 7.5$, $p = 0.01$] indicated that N400 amplitudes were larger in female than in male participants. Further, a significant three-way interaction of congruence × valence × sex [$F(1,30) = 7.6$, $p = 0.01$] suggested that in female participants, affectively incongruent music targets elicited larger N400 amplitudes regardless of valence, whereas in male participants the N400 occurred only for happy music targets. Sex as a between-subject factor did not reach significance [$F(1,30) < 1$].

Further, a main effect of region [$F(1,60) = 96.6$, $p < 0.01$] showed that negativities were largest at anterior regions. A significant interaction between region and hemisphere [$F(1,60) = 14.8$, $p < 0.01$] further suggested more negative voltages at left anterior and central regions (compared to their right counterparts) and more negative voltages at the right posterior region (compared to its left counterpart).

After controlling for alexithymia, RM-MANCOVA yielded no main effect of congruence [$F(1,29) < 1$]. However, a significant congruence × sex interaction [$F(1,29) = 7.4$, $p = 0.01$] revealed that the N400 for emotional music occurred only in female participants. The main effect of region remained [$F(1,58) = 12.1$, $p < 0.01$]. No further main effects were observed. The two-way interaction region × hemisphere [$F(1,60) = 14.8$, $p < 0.01$] remained significant. Alexithymia showed a marginally significant interaction with hemisphere [$F(1,29) = 4.0$, $p = 0.06$], indicating a trend toward larger negativities in the left hemisphere in individuals with higher alexithymia scores.

Pearson's correlations confirmed an association of alexithymia with N400 amplitudes for affectively incongruent compared to congruent music targets (Figure 5). This effect was found to be left-lateralized. For affectively incongruent music irrespective of valence, alexithymia correlated negatively with N400 amplitudes at the left central region ($r = -.36$, $p = 0.04$), and tended to correlate negatively with N400 amplitudes at the left posterior region ($r = -.33$, $p = 0.07$). For happy music targets only, alexithymia also correlated negatively with N400 amplitudes in the left anterior region ($r = -.40$, $p = 0.02$).

ProsodyTarget

A main effect of congruence was observed for prosody targets at the N400 time-window [$F(1,30) = 7.1$, $p = 0.01$], indicating a larger N400 for affectively incongruent compared to congruent targets. No main effect of valence was found [$F(1,30) < 1$]. A main effect of region [$F(1,30) = 7.9$, $p < 0.01$] suggested that most negative voltages occurred at anterior regions. No further main effects or interactions reached significance.

After controlling for alexithymia, the effect of congruence for prosody targets failed to reach significance [$F(1,29) < 1$]. RM-MANCOVA revealed no significant main effects. There was a trend toward a congruence × hemisphere interaction [$F(1,29) = 3.1$, $p < 0.09$], indicating that the N400 for affectively incongruent prosody tended to be larger in the left hemisphere compared to the right hemisphere. This lateralization effect tended to be stronger in individuals with high scores on alexithymia, as indicated by a marginally significant three-way congruence × hemisphere × alexithymia interaction [$F(1,29) = 3.5$, $p < 0.07$].

Pearson's correlations confirmed an association of alexithymia with N400 amplitudes in response to affectively incongruent prosody (see Figure 5). This effect was found to be located at the posterior region in the left hemisphere. Alexithmia showed a significant negative correlation with N400 amplitudes in response to happy prosody ($r = -.35$, $p = 0.05$), and tended to correlate negatively with N400 amplitudes to sad prosody ($r = -.30$, $p = 0.09$).

MusicPrime

For words primed by affective music, RM-MANOVA revealed no significant effect of congruence [$F(1,30) < 1$] or of valence [$F(1,30) < 1$]. No significant interactions with congruence were found. Thus, the occurrence of an N400 effect to affectively incongruent words primed by music could not be established.

The effect of alexithymia when included as a covariate was not significant [$F(1,29) < 1$]. Congruence [$F(1,29) = 1.8$, $p = 0.18$] and valence [$F(1,29) < 1$] remained insignificant.

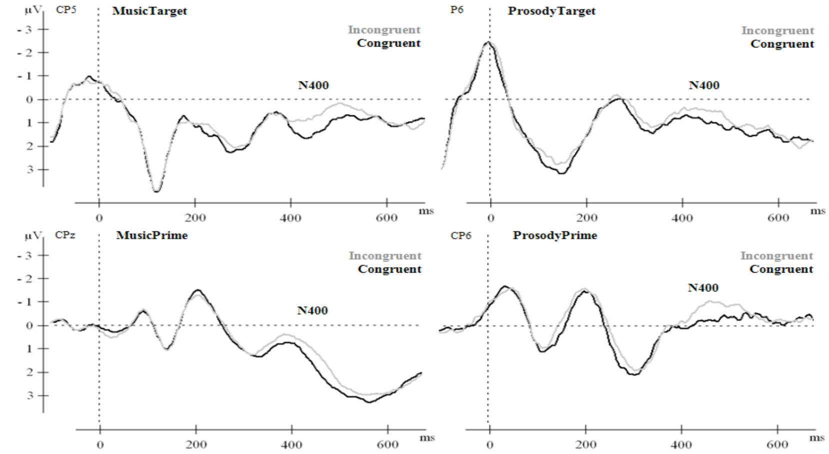

Figure 4. The N400 to Affective Incongruency. N400 in response to affectively incongruent targets (gray) versus affectively congruent targets (black) during affective categorization. Grand averages of 32 subjects at electrode site P3 are shown for A: *MusicTarget,* B: *ProsodyTarget*, C: *MusicPrime*, D: *ProsodyPrime*.

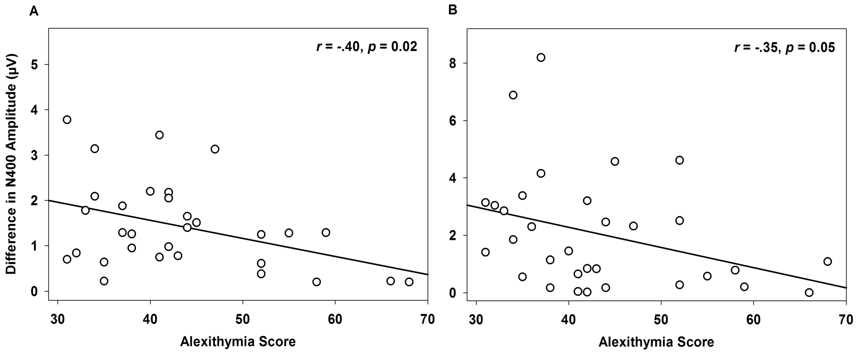

Figure 5. Correlation of alexithymia with the N400. Panel A: Negative correlation of alexithymia with amplitudes of the N400 in response to happy music at the left anterior region. Panel B: Negative correlation of alexithymia with amplitudes of the N400 in response to happy prosody at the left posterior region.

As RM-MANOVA and RM-MANCOVA showed that there was no N400 in response to affectively incongruent words primed by music, Pearson's correlations between alexithymia and N400 amplitudes could not be conducted in this condition.

ProsodyPrime

A main effect of congruence was observed for visual word targets preceded by prosody primes at the N400 time-window [$F(1,30) = 6.6$, $p = 0.02$], indicating a larger N400 for affectively incongruent compared to congruent target words. No main effect of valence was found [$F(1,30) < 1$]. The congruence effect was qualified by a three-way congruence × valence × hemisphere interaction [$F(1,30) = 4.6$, $p = 0.04$], which indicated that in the left hemisphere, the N400 was elicited by both positive and negative words, while in the right hemisphere this was only true for negative words.

A main effect of region [$F(1,60) = 50.7$, $p < 0.01$] suggested more negative voltages at anterior (mean: -1.9 μV) and central regions (mean: -1.0 μV) compared to posterior regions (mean: 2.5 μV). The effect of sex reached significance [$F(1,30) = 4.8$, $p = 0.04$], indicating more negative voltages in females as compared to male participants. A significant region × hemisphere interaction [$F(1,60) = 14.3$, $p < 0.01$] showed more negative voltages in the left anterior region, but more negative voltages at central and posterior regions in the right hemisphere. This interaction was further qualified by a three-way interaction region × hemisphere × sex [$F(1,60) = 6.7$, $p < 0.01$], suggesting more negative voltages at right central and posterior regions in both genders, and more negative voltages in the left anterior region in females but no anterior lateralization in male participants.

Including alexithymia as a covariate in RM-MANCOVA, the congruence effect became insignificant [$F(1,29) < 1$], instead a main effect of valence was observed [$F(1,29) = 4.5$, $p = 0.04$], indicative of more negative voltages in response to negative compared to positive target words. Alexithymia interacted with this valence effect [$F(1,29) = 4.7$, $p = 0.04$]: for negative targets the difference in voltages did not vary as a function of alexithymia, whereas positive words elicited more negative voltages with increasing alexithymia scores. However, the interaction congruence × valence × alexithymia was insignificant [$F(1,29) < 1$], indicating that the interaction between alexithymia and valence did not qualify the N400 effect to incongruent vs congruent conditions. As a between-subjects effect, alexithymia was not significant [$F(1,29) < 1$]. The effect of region remained significant [$F(1,58) = 8.4$, $p < 0.01$], and so did the effect of sex [$F(1,29) = 4.4$, $p = 0.04$]. Lastly, the three-way interaction region × hemisphere × sex was still significant [$F(1,58) = 6.3$, $p < 0.01$].

Pearson's correlations confirmed that there was no significant impact of alexithymia on N400 amplitudes in response to affective words primed by prosody ($p > 0.1$). No effect of alexithymia on the valence of the word targets was observed ($p > 0.1$).

4. Discussion

The results of the present study indicate that alexithymia is associated with impairment in the automatic processing of emotion conveyed by music and speech prosody during affective categorization. At the electrophysiological level, alexithymia scores correlated negatively with amplitudes of the N400, an ERP component indicative of individual sensitivity to affective incongruence. This correlation was observed in the left hemisphere in response to affectively incongruent music and speech prosody. For prosodic targets, the effect had a posterior locus; for music targets, differences in N400 amplitudes were more broadly distributed over central and posterior regions for happy and sad music, and additionally included the left anterior region for happy music targets only. No difference was found during affective categorization of word targets. At the behavioral level, we observed a trend toward a reduced affective priming effect with increasing alexithymia scores for prosodic targets (irrespective of valence), and no impact of alexithymia on affective priming for music and word targets.

The results of the present study replicate previous findings of cross-modal affective priming effects between speech prosody and visually presented words (Bostanov & Kotchoubey, 2004; Koelsch et al., 2004; Kotz & Paulman, 2007; Schirmer & Kotz, 2003; Schirmer, Kotz, & Friederici, 2002; Schirmer, Kotz, & Friederici, 2005) and between music and linguistic stimuli (Koelsch et al., 2004; Chen et al., 2008; Daltrozzo & Schön, 2008; Logeswaran & Bhattacharya, 2009; Sollberger et al., 2003; Steinbeis & Koelsch, 2008; Steinbeis & Koelsch, 2009). Our results further confirm the occurrence of an N400 effect in response to affectively incongruent music, prosody, and linguistic stimuli in a cross-modal priming paradigm (Daltrozzo & Schön, 2008; Steinbeis & Koelsch, 2008; Steinbeis & Koelsch, 2009). However, we failed to replicate the N400 effect for words primed by affective music. In contrast to the studies by Steinbeis and colleagues (Steinbeis & Koelsch, 2008; Steinbeis & Koelsch, 2009) we used short natural music excerpts rated as happy or sad instead of music chords. Dissonant chords such as used in those previous studies are perceived as unpleasant, which was presumably (though not formally tested) not the case for our sad piano music excerpts. This difference in pleasantness of the stimuli used could account for the higher potency of chords in priming visual words compared to natural music excerpts such as used in the present study. Our finding of an only marginally significant affective priming effect for music primes at the behavioral level supports this hypothesis.

Furthermore, our findings support the notion of sex differences in the perception of emotion in music and prosody. Female participants categorized both happy and sad music targets faster than men in the present study. This behavioral difference was accompanied by significantly larger N400 amplitudes in response to affectively incongruent music in women compared to men. In addition, women showed an N400 effect irrespective of music valence, whereas in men this effect was only observed for happy music. These differences, indicative of a higher sensitivity to musical emotions in women are in line with previous reports of larger brain activation and greater positive attribution to affective music in women compared to men (Altenmüller et al., 2002), larger networks of coherent brain oscillations in response to pleasant music in female participants (Flores-Gutiérrez et al., 2009), and greater psychophysiological reactivity reflected by elevated finger temperature and skin conductance level in women compared to men (Nater et al., 2006).

Sex differences in emotional prosody perception have likewise been reported repeatedly: women recognized emotional prosody faster than men (Schirmer et al., 2002), and showed an N400 in an emotional prosody Stroop task while no such effect was observed in men (Schirmer & Kotz, 2003). Even at pre-attentive processing levels sex differences seem to exist: deviants in emotional prosody elicited larger amplitudes of the Mismatch Negativity (MMN) in response to prosodic deviants in women, but not in men (Schirmer et al., 2005). We did not find significant effects of sex for emotional prosody targets, however, when prosody served as a prime women tended to show an affective priming effect for word targets irrespective of target valence, whereas this effect occurred in men only for positive word targets. This behavioral difference was accompanied by generally larger negativities to words primed by emotional prosody in women compared to men (an effect not found when the same words were primed by emotional music). The absence of a sex difference for prosody as a target in our study could be due to the low task difficulty, indicated by a ceiling effect in performance. Low task difficulty may have masked possible differences in emotional prosody categorization in our study, and could explain why sex differences did occur when emotional prosody functioned as a prime, but not as a target.

The present study is the first to provide electrophysiological evidence for an emotional categorization deficit in alexithymia for music and speech prosody. Our finding of decreased N400 amplitudes both for emotions conveyed by music and by speech prosody seems conceivable given that music and speech prosody have been shown to use the same acoustic features to convey emotions (Ilie & Thompson, 2006; Juslin & Laukka, 2003; Zatorre, Belin, & Penhune, 2002). In light of this similarity, it seems reasonable to assume that individuals with alexithymia, a condition characterized by a difficulty to identify emotions, will exhibit comparable differences in brain responses to emotions conveyed by both music and speech prosody. Future studies should employ auditory affective material to provide further evidence for the generality of the emotion processing deficit in alexithymia, which so far has been investigated using almost exclusively visual emotional information.

Our results of reduced N400 amplitudes during affective priming in response to emotional music and speech as a function of alexithymia confirm and extend previous findings of studies employing electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) that have indicated impaired subconscious processing of visual affective material in alexithymia. Corresponding to our finding of reduced left-hemispheric N400 amplitudes for affective music and speech, decreased early theta synchronization (brain oscillations related to the cortico-hippocampal-limbic interaction during cognitive-emotional processing) was observed in the left hemisphere of alexithymic individuals during the processing of emotional pictures. This was interpreted as a disruption in automatic affective processing and as an analytical, categorical decoding difficulty of emotional stimuli in alexithymia (Aftanas & Varlamov, 2004). Our findings are further in line with fMRI studies on the automatic processing of facial expressions of emotions in alexithymia. These studies found that alexithymia was associated with reduced activation of several brain areas during the automatic processing of masked happy, sad, and surprised faces (Duan et al., 2010, Kugel et al., 2008, Reker et al., 2010), suggesting that alexithymia is associated with reduced emotional processing at an automatic processing level.

The observed trend toward a negative correlation between alexithymia and affective categorization of prosodic targets at the behavioral level is partially in line with a previous study (Swart et al., 2009). Although individuals with alexithymia showed lower accuracy and longer reaction times during prosody identification of sentences with mismatching prosody and semantics in that previous study, these differences did not reach significance. Possibly, this was due to the long duration of the sentences presented (20 s), while the current study employed very short prosodic targets with a length of 600 ms, thereby increasing task demands.

We did not find a relationship between alexithymia levels and N400 amplitudes during affective categorization of visual word targets following music and prosody primes. This finding reveals an asymmetry regarding the effect of alexithymia on the processing of emotional music and prosody: During the categorization of word targets primed by emotional music and prosody, alexithymia had no significant impact on N400 amplitudes. In contrast, when affective music and prosody targets were to be categorized, alexithymia scores were associated with significantly decreased N400 amplitudes. This asymmetry for affective categorization of music and prosody targets versus word targets could be due to the fact that in order to categorize the auditory targets (music and prosody), they had to be (internally) verbalized before they could be categorized as happy or sad. In contrast, in conditions in which music and prosody served as primes, such verbalization of auditory affective information was not necessary as the decision was to be made on the visual word targets.

Difficulties to identify and verbalize emotions are diagnostic criteria of alexithymia (Nemiah, Freyberger, & Sifneos, 1976). Hitherto, the question as to whether alexithymia is associated with impairment in verbalizing emotions conveyed by prosody and music has not been addressed. However, alexithymia exhibits a high comorbidity with Autism Spectrum Disorder (ASD) (Hill, Berthoz, & Frith, 2004; Bethoz & Hill, 2005). In a recent study on the experience of music in ASD, Allen and colleagues report that individuals with diagnoses on the autism spectrum showed conscious awareness of the emotional arousal induced by music, but exhibited limitations in the terms used to describe the emotional effect of music (Allen, Hill, & Heaton, 2009). Such difficulty to verbalize emotions conveyed by music could underlie our observation of diminished N400 amplitudes with increasing alexithymia levels during the affective categorization of emotional music and prosody targets, but not word targets.

Our finding of unaffected emotional word processing is seemingly at odds with previous studies suggesting an impact of alexithymia on affective priming for word targets (Suslow, 1998; Vermeulen et al., 2006). However, it should be born in mind that the larger affective priming effect for positive word targets with increasing scores on alexithymia (Suslow, 1998) could not be replicated in a follow-up study (Suslow et al., 2001), despite the fact that the same word evaluation task was employed in a larger sample of participants. Instead, the follow-up study revealed no correlations between alexithymia and the affective priming effect, neither for positive word targets nor for negative word targets. Our findings confirm these results. Moreover, Suslow and colleagues used an SOA of 300 ms, although it has been shown that strategic components can come into effect below an SOA of 300 ms (Klauer, Rossnagel, & Musch, 1997) and that the affective priming effect dissipates already at 300 ms (Hermans et al., 2001). Therefore, an involvement of non-automatic, strategic processes cannot be ruled out in these studies.

Vermeulen and colleagues showed that alexithymia correlated negatively with the affective priming effect for word targets when these were primed by angry faces (Vermeulen et al., 2006). This reduced affective priming effect was not found for happy face primes and neither for positive and negative word primes. The authors interpreted this finding as an anger/threat-related automatic processing deficit associated with alexithymia. This interpretation found further support in findings of hampered memory performance during the perception of angry, but not happy background music in alexithymia (Vermeulen et al., 2010). The present study used happy and sad prosody and music as primes for affective word targets and found that alexithymia did not correlate with the affective priming effect for word targets. These results do not contradict the findings of Vermeulen and colleagues (Vermeuen et al., 2006), nor can they confirm these results as angry emotion was not included in our paradigm. It would be interesting to test in future studies whether the hypothesis of an anger-specific processing deficit in alexithymia holds when angry speech prosody and anger/threat evoking music are used to prime word targets during affective categorization.

Limitations

It should be kept in mind that the alexithymia construct may comprise two related, but distinct types, type I and type II alexithymia (Vorst & Bermond, 2001), though a recent study failed to find empirical support for this distinction (Bagby et al., 2009). The TAS-20 questionnaire, used here in agreement with previous studies on affective priming in alexithymia (Suslow, 1998; Suslow et al., 2001, Suslow et al., 2002; Vermeulen et al., 2006; Vermeulen et al., 2010), covers only type II alexithymia. This type is characterized by deficits to cognitively access and verbalize emotions, while the general emotional responsiveness is thought to be intact. Type I alexithymia, characterized by a general lack of emotional responsiveness could not be controlled for in the present study. Varying scores on type I alexithymia might have confounded the present results and could contribute to the fact that the present findings are not in line with previous studies on affective priming in alexithymia, which likewise used the TAS-20 and thus did not control for this possible confound. Future studies should additionally use the BVAQ questionnaire, which distinguishes between the two types of alexithymia and would thus makes it possible to control for this possible confound.

A further limitation of the present study is the lack of a correction for multiple comparisons during product-moment correlations between alexithymia and N400 amplitudes in response to affective words, music, and prosody. The number of comparisons conducted in the present study was relatively small, however, the present findings should be treated as preliminary for this reason. Future studies should attempt to overcome these limitations.

Conclusions

In sum, the results of this study suggest a reduced sensitivity to emotional qualities of speech and music in alexithymia at a neurophysiological level. Our findings of differential brain responses to affective categorization of music and speech prosody as compared to visual words with emotional connotations indicates that alexithymia impairs the categorization of affective stimuli primarily in situations in which a verbalization of the emotional information is required. However, this interpretation remains speculative until future research provides further insight into the nature of the emotional processing deficit in alexithymia.

ACKNOWLEDGEMENTS

We thank Jos Pacilly for computer scripting assistance.

5. References

Aftanas, L., & Varlamov, A. (2004). Associations of alexithymia with anterior and posterior activation asymmetries during evoked emotions: EEG evidence of right hemisphere 'electrocortical effort'. *International Journal of Neuroscience, 114,* 1443-1462.

Allen, R., Hill, E., & Heaton, P. (2009). The subjective experience of music in autism spectrum disorder. *Annals of the New York Academy of Sciences, 1169,* 326-331.

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,* 2242-2256.

Bagby, R., Parker, J., Taylor, & G. J. (1994a). The twenty-item Toronto Alexithymia Scale: I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38,* 23-32.

Bagby, R., Taylor, G. J., & Parker, J. (1994b). The twenty-item Toronto Alexithymia Scale: II. Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research, 38,* 33-40.

Bagby, R. M., Quilty, L. C., Taylor, G. J., Grabe, H. J., Luminet, O., Verissimo, R., De Grootte, I, & Vanheule S. (2009). Are there subtypes of alexithymia? *Personality and Individual Differences, 47,* 413-418.

Berthoz, S., Artiges, E., Van de Moortele, P. F., Poline, J. B., Rouquette, S., Consoli., SM., & Martinot, J. L. (2002). Effect of impaired recognition and expression of emotions on frontocingulate cortices: An fMRI study of men with alexithymia. *American Journal of Psychiatry, 159,* 961-967.

Berthoz, S., & Hill, E. (2005). Reliability of the Bermond-Vorst Alexithymia Questionnaire: data from adults with autism spectrum disorder, their relatives and normal controls. *European Journal of Psychiatry, 20, 281*-298.

Bird, G., Silani, G., Brindley, R., White, S., Frith, & U., Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain, 133,* 1515-1525.

Boersma, P., & Weenink, D. (1996). *Praat, a system for doing phonetics by computer: version 3.4* [computer program].

Bostanov, V., & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology, 41,* 259-268.

Chen, J., Yuan, J., Huang, H., Chen, C., & Li, H. (2008). Music-induced mood modulates the strength of emotional negativity bias: An ERP study. *Neuroscience Letters, 445,* 135-139.

Daltrozzo, J., & Schön, D. (2008). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience, 21,* 1882-1892.

Duan, X., Dai, Q., Gong, Q., & Chen, H. (2010). Neural mechanism of unconscious perception of surprised facial expression. *NeuroImage, 52,* 401-407.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion, 15,* 115-141.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50,* 229-238.

Flores-Gutiérrez, E. O., Díaz, J. L., Barrios, F. A., Guevara., Del Río-Portilla, Y., Corsi-Cabrera, M., & Del Flores-Gutiérrez, E. O. (2009). Differential alpha coherence hemispheric patterns in men and women during pleasant and unpleasant musical emotions. *International Journal of Psychophysiology, 71,* 43-49.

Franz, M., Schaefer, R., Schneider, C., Sitte, W., & Bachor, J. (2004). Visual event-related potentials in subjects with alexithymia: Modified processing of emotional aversive information? *The American Journal of Psychiatry, 161,* 728-735.

Gundel, H., Greiner, A., Ceballos-Baumann, A. O., Ladwig, K. H., Von Rad, M., Förstl, H., & Jahn, T. (2004). Alexithymia is no risk factor for exacerbation in spasmodic torticollis patients. *Journal of Psychosomatic Research, 56*, 699-705.

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition and Emotion, 15*, 143-165.

Hill, E., Berthoz, S., & Frith, U. (2004). Brief Report: Cognitive processing of own emotions in individuals with autistic spectrum disorder and in their relatives. *Journal of Autism and Developmental Disorders, 34,* 229-235.

Ilie, G. T., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception, 23,* 319-29.

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

Klauer, K. C., & Musch J. (2003). Affective priming: Findings and theories. In: K. Klauer & J. Musch (Eds.). *The psychology of evaluation: Affective processes in cognition and emotion* (7-50). Mahwah, NJ: Erlbaum.

Klauer, K. C., Rossnagel, C., & Musch, J. (1997). List-context effects in evaluative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 246-255.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience, 7,* 302-307.

Kooiman, C., Spinhoven, P., & Trijsburg, R. (2002). The assessment of alexithymia: A critical review of the literature and a psychometric study of the Toronto Alexithymia Scale-20. *Journal of Psychosomatic Research, 53,* 1083-1090.

Kotz, S., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research, 1151,* 107-118.

Kugel, H., Eichmann, M., Dannlowski, U., Ohrmann, P., Bauer, J., Arolt, V, Heindel, W... Suslow, T. (2008). Alexithymic features and automatic amygdala reactivity to facial emotion. *Neuroscience Letters, 435,* 40-4.

Kutas, M., & Federmeier, KD. (2011). Thirty years and counting: finding meaning in the n400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62,* 621-627.

Lane, R. D., Sechrest, L., Reidel R., Weldon, V., Kaszniak, A., & Schwartz, G. E. (1996). Impaired verbal and nonverbal emotion recognition in alexithymia. *Psychosomatic Medicine, 58,* 203-210.

Larsen, J., Brand., N., Bermond., B., & Hijman., R. (2003). Cognitive and emotional characteristics of alexithymia: A review of neurobiological studies. *Journal of Psychosomatic Research, 54,* 533-541.

Logeswaran, N., & Bhattacharya, J. (2009). Crossmodal transfer of emotion by music. *Neuroscience Letters, 455,* 129-133.

Luminet, O., Vermeulen, N., Demaret, C., Taylor, G. J., & Bagby, R. M. (2006). Alexithymia and levels of processing: Evidence for an overall deficit in remembering emotion words. *Journal of Research in Personality, 40,* 713-733.

Lumley, M. A., Neely, L. C., & Burger, A. J. (2007). The assessment of alexithymia in medical settings: Implications for understanding and treating health problems. *Journal of Personality Assessment, 89,* 230-246.

Mantani, T., Okamoto, Y., Shirao, N., Okada, G., & Yamawaki, S. (2005). Reduced activation of posterior cingulate cortex during imagery in subjects with high degrees of alexithymia: A functional magnetic resonance imaging study. *Biological Psychiatry, 57,* 982-990.

Marin, M. M., & Bhattacharya, J. (2010). Music induced emotions: some current issues and cross-modal comparisons. In: J. Hermida & M. Ferreo (Eds.). *Music Education* (1-38). Hauppauge, NY: Nova Science Publishers.

Mériau, K., Wartenburger, I., Kazzer, P., Prehn, K., Lammers, C. H., Van der Meer, E., Villringer, A., & Heekeren, HR. (2006). A neural network reflecting individual differences in cognitive processing of emotions during perceptual decision making. *NeuroImage, 33,* 1016-1027.

Nater, M. U., Abbruzzese, E., Krebs, M., & Ehlert, U. (2006). Sex differences in emotional and psychophysiological responses to musical stimuli. *International Journal of Psychophysiology, 62,* 300-308.

Nemiah, J., Freyberger, H., & Sifneos, PE. (1976). Alexithymia: A view of the psychosomatic process. In O. W. Hill (Ed.). *Modern Tends in Psychosomatic Medicine, 3*, (430-439). London: Butterworths.

Pollatos, O., Schubö A., Herbert, B. M., Matthias, E., & Schandry, R. (2008). Deficits in early emotional reactivity in alexithymia. *Psychophysiology, 45,* 839-846.

Pouga, L., Berthoz, S., De Gelder, B., & Grèzes, J. (2010). Individual differences in socioaffective skills influence the neural bases of fear processing: The case of alexithymia. *Human Brain Mapping, 31,* 1469-1481.

Prkachin, G. C., Casey, C., & Prkachin., K. M. (2008). Alexithymia and perception of facial expressions of emotion. *Personality and Individual Differences, 46,* 412-417.

Reker, M., Ohrmann, P., Rauch, A. V., Kugel, H., Bauer, J., Dannlowski, U Arolt, V... Suslow, T. (2010). Individual differences in alexithymia and brain response to masked emotion faces. *Cortex, 46,* 658-667.

Schirmer, A., & Kotz, S. (2003). ERP evidence for a sex-specific Stroop effect in emotional speech. *Journal of Cognitive Neuroscience, 15,* 1135-1148.

Schirmer, A., Kotz, S., & Friederici, A. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research, 14,* 228-233.

Schirmer, A., Kotz S., & Friederici, A. (2005). On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cognitive Brain Research, 24,* 442-452.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide.* Pittsburgh: Psychology Software Tools Inc.

Sollberger, B., Reber, R., & Eckstein, D. (2003). Musical chords as affective priming context in a word-evaluation task. *Music Perception, 20,* 263-282.

Steinbeis, N., & Koelsch, S. (2008). Comparing the processing of music and language meaning using eeg and fmri provides evidence for similar and distinct neural representations. *PLoS ONE, 3,* e2226.

Steinbeis, N., & Koelsch, S. (2009). Affective priming effects of musical sounds on the processing of word meaning. *Journal of Cognitive Neuroscience, 23,* 604-621.

Suslow, T. (1998). Alexithymia and automatic affective processing. *European Journal of Personality, 12,* 433-443.

Suslow, T., & Junghanns, K. (2002). Impairments of emotion situation priming in alexithymia. *Personality and Individual Differences, 32,* 541-550.

Suslow, T., Junghanns, K., Donges, U. S., & Arolt, V. (2001). Alexithymia and automatic processing of verbal and facial affect stimuli. *Cahiers de Psychologie Cognitive* [Current Psychology of Cognition]*, 20,* 297-324.

Swart, M., Kortekaas., R., & Aleman., A. (2009). Dealing with feelings: Characterization of trait alexithymia on emotion regulation strategies and cognitive-emotional processing. *PLoS ONE, 4,* e5751.

Taylor, GJ. (2000). Recent developments in alexithymia theory and research. [The Canadian Journal of Psychiatry] *La Revue canadienne de psychiatrie, 45,* 134-142.

Taylor, G. J., Bagby., R.M., & Parke., J. D. A. (1997). *Disorders of Affect Regulation: Alexithymia in Medical and Psychiatric Illness.* Cambridge: Cambridge University Press.

Vermeulen, N., Luminet, O., & Corneille, O. (2006). Alexithymia and the automatic processing of affective information: evidence from the affective priming paradigm. *Cognition and Emotion, 20,* 64-91.

Vermeulen., N., Toussaint, J., & Luminet, O. (2010). The influence of alexithymia and music on the incidental memory for emotion words. *European Journal of Personality, 24,* 551-568.

Vorst, H. C. M., & Bermond, B. (2001). Validity and reliability of the Bermond-Vorst Alexithymia Questionnaire. *Personality and Individual Differences, 30,* 413-434.

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

Chapter 3.2

Modulation of emotional prosody perception by alexithymia: fMRI evidence

Contributions of first and second author:

Conceived the experiments: J. W.
Designed the study materials and tasks: J. W.
Data collection: J. W.
Analyzed the data: J. W. and K. G
Wrote the first draft of the report: J. W. (methods) and K. G (introduction, results, conclusion).
Commented on the first draft: J. W.

Goerlich, K. S., Witteman, J., Schiller, N. O., Van Heuven, V. J., Aleman, A., & Martens, S. (2013). Blunted feelings: Alexithymia is associated with a diminished neural response to speech prosody. *Social Affective And Cognitive Neuroscience, 9,* 1108-1117.

Abstract

How we perceive emotional signals from our environment depends on our personality. Alexithymia, a personality trait characterized by difficulties in emotion regulation has been linked to aberrant brain activity for visual emotional processing. Whether alexithymia also affects the brain's perception of emotional speech prosody is currently unknown. We used functional magnetic resonance imaging to investigate the impact of alexithymia on hemodynamic activity of three a-priori regions of the prosody network: the superior temporal gyrus, the inferior frontal gyrus, and the amygdala. Twenty-two subjects performed an explicit task (emotional prosody categorization) and an implicit task (metrical stress evaluation) on the same prosodic stimuli. Irrespective of task, alexithymia was associated with a blunted response of the right superior temporal gyrus and the bilateral amygdalae to angry, surprised, and neutral prosody. Individuals with difficulty describing feelings deactivated the left superior temporal gyrus and the bilateral amygdalae to a lesser extent in response to angry compared to neutral prosody, suggesting that they perceived angry prosody as relatively more salient than neutral prosody. In conclusion, alexithymia seems to be associated with a generally blunted neural response to speech prosody. Such restricted prosodic processing may contribute to problems in social communication associated with this personality trait.

1. Introduction

The perception and interpretation of emotional signals is an important part of social communication. Body gestures and posture, facial expressions as well as the tone of voice provide crucial insight into another person's mind (Van Kleef, 2009). However, how emotional signals are perceived and interpreted may differ considerably across individuals, and the same emotional signal can evoke a different response in different people (Hamann & Canli, 2004; Ormel et al., 2012). Indeed, recent neuroimaging studies indicate that personality modulates the brain's response to emotional signals in our environment (e.g., Brück, Kreifelts, Kaza, Lotze, & Wildgruber., 2011); Brühl et al., 2011; Canli, 2004; Frühholz, Ceravolo, & Grandjean, 2011; Hooker, Verosky, Miyakawa, Knight, & D'Esposito, 2008).

Difficulties interpreting emotions lie at the core of alexithymia ('no words for feelings'), a personality construct referring to a specific deficit in emotional processing (Sifneos, 1973). Individuals scoring high on alexithymia have difficulty identifying, analyzing, and verbalizing their feelings, reading emotions from faces (e.g., Parker, Prkachin, & Prkachin, 2005; Prkachin, Casey, & Prkachin., 2009; Swart, Kortekaas, & Aleman, 2009) and describing other's emotional experiences (Bydlowski et al., 2005). These difficulties in the cognitive processing of emotions, which constitute the cognitive alexithymia dimension, may be accompanied by reduced capacities to experience emotional arousal (affective alexithymia dimension). Flattened affect paired with diminished empathy for the feelings of others (Grynberg, Luminet, Corneille, Grezes, & Berhoz, 2010; Guttman & Laporte, 2002) may lead to a perception of alexithymic individuals as cold and distant (Spitzer, Barnow, Grabe, & Freyberger, 2005) and interpersonally indifferent (Vanheule, Desmet, Meganck, & Bogaerts, 2007), resulting in problems in social life.

In the past decade, neuroimaging studies have begun to reveal the neural basis of emotion processing deficits associated with alexithymia. By means of functional magnetic resonance imaging (fMRI), these studies demonstrated aberrant brain activity in individuals scoring high on alexithymia for a variety of emotional processing tasks, such as the viewing of facial and bodily expressions of emotions (Berthoz et al., 2002; Kugel et al., 2008; Mériau et al., 2006; Pouga, Berthoz, De Gelder, & Grezes, 2010), during empathy for pain (Bird et al., 2010), and during the imagery of autobiographic emotional events (Mantani, Okamoto, Shirao, Okada, & Yamawaki, 2005). Such differences were not only found during explicit processing (i.e. when participants were asked to explicitly evaluate the emotional dimension of the stimuli), but also during implicit processing (i.e., when participants were asked to direct their attention toward another dimension than emotion or were unaware of the emotional dimension; see De Houwer, 2006), such as during the brief presentation of masked emotional faces (Duan, Dai, Gong, & Chen, 2010; Kugel et al., 2008; Leweke et al., 2004; Reker et al., 2010; see Grynberg et al., 2012, for a review).

Previous research into the neural basis of alexithymia has mostly focused on the processing of visual emotional stimuli such as facial or bodily expressions of emotions or emotional pictures and videos. Surprisingly, the impact of alexithymia on the perception of emotional prosody (the melody of speech) has received little attention despite its importance in conveying emotion through the voice in daily conversation. In an electroencephalography (EEG) study using a cross-modal affective priming paradigm, we tested the impact of alexithymia on the N400 component, an indicator of the perception of mismatches in affective meaning, in response to music and speech

(Goerlich, Witteman, Aleman, & Martens, 2011). Alexithymia correlated negatively with N400 amplitudes for mismatching music and speech, suggesting that people scoring high on this personality trait may be less sensitive to aurally perceived emotions. The suggestion of a reduced sensitivity to emotional speech prosody in alexithymia was confirmed in a further EEG study, in which we additionally observed that alexithymia did not only affect the explicit, but also the implicit perception of emotional prosody qualities (Goerlich, Aleman, & Martens, 2012).

Taken together, these findings suggest that alexithymia is linked to differences in the way the brain processes emotions conveyed through the voice. However, the brain regions underlying such differences remain unknown as neuroimaging studies with the necessary higher spatial resolution to identify them are currently lacking. Therefore, the present fMRI study aimed to investigate how alexithymia affects the neural processing of emotional prosody. As the impact of alexithymia may vary depending on whether attention is directed toward the emotional prosodic dimension or not, we investigated both explicit processing (participants categorized emotional prosody, i.e. attention was directed toward the emotional dimension) and implicit processing of the same prosodic stimuli (participants evaluated the metrical stress position, i.e. attention was directed toward a different dimension than emotional prosody).

The neural network underlying emotional prosody perception has been investigated by numerous studies over the past decades (for a recent review, see e.g. Kotz & Paulmann, 2011). A recent meta-analysis of the lesion literature concluded that the right hemisphere is relatively more involved in prosodic processing than the left (Witteman, Van Ijzendoorn, Van de Velde, Van Heuven, & Schiller, 2011). Regarding a specific network for prosodic processing, converging evidence suggest the involvement of fronto-temporal regions and subcortical structures (for a recent meta-analysis, see Witteman, Van Heuven, & Schiller, 2012). The processing of emotional prosody has been proposed to involve three phases (Schirmer & Kotz, 2006). In the initial phase, basic acoustic properties are extracted, a process presumably mediated by the middle superior temporal gyrus (Wildgruber, Ethofer, grandjean, & Kreifelts, 2009). In the second phase, the extracted acoustic information is integrated into an emotional percept or 'gestalt'. This process takes place in the superior temporal cortex, with the laterality and anterior-posterior distribution of activity within this structure being sensitive to stimulus-specific features (medium, valence) as well as to task conditions such as attentional focus (for a meta-analysis on the functional segregation of the superior temporal cortex, see Frühholz & Grandjean, 2013). In the third and final phase, emotional prosody is explicitly evaluated and integrated with other cognitive processes in the inferior frontal gyrus (Brück et al., 2011a). With respect to the involvement of subcortical structures, the amygdala has been implicated by several studies in emotional prosody processing (Brück et al., 2011a; Ethofer et al., 2009; Frühholz & Grandjean, in press; Jacob et al., 2012; Leitman et al., 2010; Mothes-Lasch, Mentzel, Miltner, & Straube, 2011; Sander et al., 2005; Wiethoff, Wildgruber Grodd, & Ethofer, 2009; Wildgruber, Ethofer, Kreifelts, & grandjean, 2008, but see Adolphs & Tranel, 1999; Adolphs, Tranel, & Damasio, 2001; Wildgruber et al., 2005; Witteman et al., 2012), with the role ascribed to the amygdala in this context being the initial detection of emotional salience and relevance (Kotz & Paulmann, 2011). The amygdala is a crucial structure for the processing of emotions in general (see Armony, 2013, for a recent review), and has further been suggested to act as a general detector of personal significance (Sander et al., 2003). Amygdala responsiveness seems to depend on personality characteristics, as has recently been demonstrated for the case of neuroticism (Brück et al., 2011a), and several studies reported a reduced responsiveness of the amygdala during the implicit and explicit processing of visual emotional information in alexithymia (Leweke et al., 2004; Kugel et al., 2008; Pouga et al., 2010).

Taken together, previous evidence suggests that the superior temporal gyrus (STG), the inferior frontal gyrus (IFG), and the amygdala are critically involved in the neural processing of emotional prosody. Therefore, these three regions were chosen as a-priori regions of interest (ROIs) for present study. The aim of the study was to investigate the impact of alexithymia on the neural processing of emotional prosody under explicit as well as under implicit task conditions. Given previous evidence of a reduced sensitivity to emotional prosody and a reduced responsiveness of the amygdala to visual emotional information in alexithymia, we hypothesized diminished ROI activity during the implicit and explicit processing of emotional prosody with higher levels of alexithymia.

2. Methods

2.1. Participants

A total of 22 subjects (9 males; mean age 24.8 ± 5.3) participated in the study. Three subjects performed at chance level on the implicit task while performing normally on the explicit task. Therefore, analyses of the implicit task did not consider the data of these subjects. All participants were right-handed as determined by the Edinburgh Handedness Inventory (mean 88.11 ± 11.32, minimum 67). Participants were native speakers of Dutch, had normal or corrected-to-normal vision, no hearing problems, and no psychiatric or neurological disorder in present or past. All participants gave informed consent prior to the experiment and received € 20 compensation for participation. The study was approved by the local ethics committee and conducted in accordance with the Declaration of Helsinki.

2.2. Materials

Pseudowords (see appendix A) with a bisyllabic structure were generated. All pseudowords obeyed Dutch phonotactics and were verified for absence of semantic content. All pseudowords were expressed with neutral, (pleasantly) surprised and angry prosody, with stress on the first and second syllable by two professional actors (one male, one female). Stimuli were recorded at 16 bit resolution and a 44.1 kHz bit sample rate in a sound proofed cabin. Surprised and angry in addition to neutral prosody was chosen in order to sample positive and negative emotions that are both considered to be approach emotions (thus, there was no confounding role of the approach-withdrawal dimension). Items were intensity normalized (i.e. did not differ in mean intensity between categories) and had a mean duration of 756 ± 65 ms. In line with previous literature (Scherer, 2003), the emotional categories differed from neutral prosody in mean F0, and F0 variability and anger additionally in intensity variability (Table 1). The validity of the intended prosodic contrasts was verified by a panel of five healthy volunteers who classified each stimulus in a forced choice task (categories: angry, surprised, happy, sad, neutral, other). Only pseudowords whose prosodic category was classified correctly by at least 80% of the participants for both actors were chosen.

Two categorization tasks were created presenting identical stimuli in an implicit and an explicit emotional task. For the implicit task, the subjects' attention was directed toward a non-emotional dimension as they were asked to decide whether the metrical stress lay on the first or the second syllable of the pseudoword. For the explicit task, the subjects' attention was explicitly directed toward the emotional prosody dimension as subjects decided whether the pseudowords were spoken with neutral, angry, or surprised intonation. Tasks and instructions were identical except for the words that instructed participants to either respond to non-emotional (implicit task) or emotional (explicit task) prosodic characteristics.

From the pool of validated stimuli, 32 items of each emotional category were selected, half of which had metrical stress on the first syllable, the other half on the second syllable. Speaker gender was balanced across items.

2.3. Procedure

Each subject completed both tasks (12 minutes each). The implicit task was always presented first in order to prevent subjects from devoting attention explicitly to the emotional dimension of the stimuli in this task. Subjects were instructed that they would hear a nonsense word and to categorize the task-relevant category (emotion or metrical stress) as fast and accurately as possible with a right hand button press. Assignment of individual categories to response-buttons was counterbalanced across subjects. Participants were instructed that they could respond while the stimulus was still playing (i.e., RT was recorded from the onset of the stimulus).

Subjects first practiced the tasks in the scanner with simulated scanner noise until reaching at least 75 % performance accuracy. Then, the experimental trials started which encompassed a total of 96 trials (32 items per emotional category). Throughout the experiment a black fixation cross was presented in the center of a grey background. Auditory stimuli were presented binaurally through MR-compatible headphones and a trial ended 2 sec after stimulus onset. Stimuli were presented in an event-related fashion with a jittered inter-stimulus interval (between 4 and 8 sec). The order of stimulus presentation was pseudo-random with the restriction of no more than two consecutive presentations of the same stimulus category.

Subjects were instructed to fix their gaze on the fixation cross throughout the experiment. Stimulus presentation was controlled using E-prime 1.2 and stimulus material was presented at 16 bit resolution and a 44.1 kHz sampling frequency at a comfortable intensity level. All subjects reported that the stimuli could be perceived clearly despite the scanner noise.

Table 1. Acoustic properties of the prosodic stimuli for each emotional category.

	Neutral	**Anger**	**Surprise**
Mean intensity (dB)	79.45	79.27	80.64
Mean variation (SD) intensity	8.84	10.74	8.83
Mean F_0 (Hz)	180.73	281.35	282.46
Mean variation (SD) F_0	44.78	78.56	101.19
Mean total duration (s)	0.79	0.76	0.72

2.3.1. Twenty-Item Toronto Alexithymia Scale (TAS-20)

The TAS-20 is the most widely used measure of alexithymia with a demonstrated validity, reliability, and stability (Bagby et al., 1994 a,b). A validated Dutch translation of the scale was used (Kooiman, Spinhoven, & Trijsburg, 2002). The TAS-20 consists of 20 self-report items rated on a 5-point Likert scale (1: strongly disagree, 5: strongly agree), with five negatively keyed items. It comprises three subscales: (1) difficulty identifying feelings (e.g., "I often don't know why I'm angry"), (2) difficulty describing feelings (e.g., "I find it hard to describe how I feel about people"), and (3) externally oriented thinking (e.g., "I prefer talking to people about their daily activities rather than their feelings"). Possible scores range from 20 to 100, higher scores indicate higher degrees of alexithymia.

2.3.2. State Trait Anxiety Inventory (STAI)

The STAI (Spielberger, Gorsuch, & Lushene, 1970) differentiates between the temporary condition of "state anxiety" and the more general and long-standing quality of "trait anxiety". For the present study, the Trait Anxiety version of the STAI (T-STAI) was used in order to control for trait anxiety, which has been reported to be closely linked to alexithymia (Berthoz et al., 1999). The T-STAI evaluates relatively stable aspects of anxiety proneness (general states of calmness, confidence, and security), and thus refers to a general tendency to respond with anxiety to perceived threats in the environment. The scale consists of 20 items rated from 1 (almost never) to 4 (almost always); higher scores indicate more trait anxiety.

2.3.3. fMRI data acquisition

Imaging data were acquired on a Philips 3.0-T Achieva MRI scanner with an eight-channel SENSE head coil for radiofrequency transmission and reception (Philips Medical Systems, Best, The Netherlands). For each task, whole-brain fMRI data were acquired using T*2 - weighted gradient-echo echo-planar imaging (EPI) with the following scan parameters: 355 volumes (the first five were discarded to reach signal equilibrium); 38 axial slices scanned in ascending order; repetition time (TR) = 2200 ms; echo time (TE) = 30 ms; flip angle = 80°; FOV = 220 × 220 mm; 2.75 mm isotropic voxels with a 25 mm slice gap. A high-resolution anatomical image (T1-weighted ultra-fast gradient-echo acquisition; TR = 9.75 ms; TE = 4.59 ms; flip angle = 8°; 140 axial slices; FOV = 224×224 mm; in-plane resolution .875×.875 mm; slice thickness = 1.2 mm), and a high-resolution T*2- weighted gradient echo EPI scan (TR = 2.2 s; TE = 30 ms; flip angle = 80°; 84 axial slices; FOV = 220 × 220 mm; in-plane resolution 1.96 × 1.96 mm, slice thickness = 2 mm) were additionally acquired for registration to standard space.

2.3.4. fMRI data preprocessing

Prior to analysis, all fMRI data sets were submitted to a visual quality control check to ensure that no gross artifacts were present in the data. Data were analyzed using FSL Version 4.1.3 (Smith et al., 2004). The following preprocessing steps were applied to the EPI data sets: motion correction, removal of non-brain tissue, spatial smoothing using a Gaussian kernel of 8 mm full width at half maximum (FWHM), grand-mean intensity normalization of the entire 4D dataset by a single multiplicative factor, and a high pass temporal filter of 70s (i.e., ≥ 0.07 Hz). The dataset was registered to the high resolution EPI image, the high resolution EPI image to the T1-weighted image, and the

T1-weighted image to the 2 mm isotropic MNI-152 standard space image (T1-weighted standard brain averaged over 152 subjects; Montreal Neurological Institute, Montreal, QC, Canada). The resulting transformation matrices were then combined to obtain a native to MNI space transformation matrix and its inverse (MNI to native space).

2.4. Data analysis

2.4.1 Behavioral data analysis

Accuracy (ACC, proportion correct responses) and reaction times (RT for correct responses) were entered as dependent variables in analyses of covariance (ANCOVA) with the within-subjects factors Task (Explicit vs. Implicit), Emotion (Neutral vs. Angry vs. Surprised), Sex as a between-subjects factor, and scores on the TAS-20 alexithymia questionnaire and the T-STAI anxiety scale included as covariates.

2.4.2. fMRI data analysis

Data analysis was performed using FEAT (FMRI Expert Analysis Tool) Version 5.98, part of FSL (FMRIBs Software Library, http://www.FMRIb.ox.ac.uk/fsl). In native space, the fMRI time-series was analyzed in an event-related fashion using the General Linear Model with local autocorrelation correction applied (Woolrich et al., 2004).

For both runs, onset of each of the stimulus categories (neutral, angry, surprised) were modeled separately as an event with an 800 ms duration at the first level. Each effect was modeled on a trial-by-trial basis using a square wave function convolved with a canonical hemodynamic response function and its temporal derivative. At second level a whole-brain analysis was performed across both tasks to examine the main effect of emotional (angry and surprised) versus neutral prosody, using clusters determined by $p < 0.01$ ($z > 2.3$), and a cluster-corrected significance threshold $p_{FDR} < 0.05$. If errors were present, these trials were included in the model but not in the contrasts of interest.

2.4.3. ROI analyses

Region of interest analyses were performed on three a-priori regions: the STG (mid and posterior part), the IFG (pars opercularis), and the amygdala (all bilaterally). Anatomical ROIs for these regions were created as defined by the Harvard-Oxford atlas (http://www.fmrib.ox.ac.uk/fsl/data/atlas-descriptions.html#ho). Mean z-scores of each ROI were calculated for each stimulus category (anger, surprise, neutral) against baseline for each task and extracted for each participant using Featquery (http://www.fmrib.ox.ac.uk/fsl/ feat5/featquery.html). These z-transformed parameter estimates indicate how well the mean signal of each ROI is explained by the model. In order to identify emotion-specific ROI activity, mean z-scores were further extracted for the contrasts angry versus neutral and surprised versus neutral prosody.

2.4.4. Statistical analyses of ROI data

The first analysis used the z-scores against baseline as a dependent variable in a repeated-measures (RM) ANCOVA with Task (Explicit vs. Implicit), ROI (STG vs. IFG vs. Amygdala), Hemisphere (Left vs. Right), and Emotion (Angry vs. Surprised vs. Neutral) as within-subject factors, Sex as between-subject factor, and TAS-20 alexithymia and T-STAI anxiety scores as covariates. This analysis served to identify the impact of alexithymia on ROI activity in response to angry, surprised, and neutral prosody compared to baseline (scanner noise).

In the second analysis, emotional prosody (angry, surprised) was contrasted to neutral prosody in an RM-ANCOVA using the mean z-scores of the contrasts angry versus neutral and surprised versus neutral prosody as dependent variable, with all other factors being identical to the first analysis. In keeping with the common procedure of contrasting emotional to neutral stimuli employed in the alexithymia literature, this analysis served to identify the impact of alexithymia on ROI activity specifically in response to emotional (angry / surprised) relative to neutral prosody.

For both analyses, follow-up tests and Pearson's correlations were conducted to identify the sources of the observed effects. As the aim of the current study was to investigate the influence of alexithymia on prosodic processing, results will be reported with a focus on main effects of and interactions with the TAS-20 alexithymia scale (the effects of Task and Emotion are reported elsewhere).

3. Results

3.1. Behavioral data

TAS-20 alexithymia scores of the 22 (9 male) participants ranged from 26 to 60 (mean: 42.18 ± 7.45), i.e. none of the participants had clinical alexithymia (cut-off score ≥ 61, see Taylor et al., 1997). Alexithymia scores were unrelated to age ($r = -.143$, $p = 0.526$) and did not differ between male and female participants ($t = 1.19$, $p = 0.245$). TAS-20 scores were significantly correlated with T-STAI anxiety scores ($r = .498$, $p = .018$), indicating more trait anxiety in individuals with higher alexithymia scores. T-STAI scores were included in all analyses as covariates of no interest in order to control for the impact of trait anxiety.

Analysis of the accuracy data revealed no significant effect of Task ($F(1,15) = 2.695$, $p = 0.121$). Neither TAS-20 nor T-STAI scores showed a significant effect on accuracy (TAS-20: $F(1,15) = 1.290$, $p = 0.274$, STAI: $F < 1$), suggesting that alexithymia did not impair the identification of neutral, angry, and surprised prosody (explicit task) and of metrical stress (implicit task). Analysis of RT data likewise revealed no significant main effects or interactions.

3.2. fMRI Data

3.2.1. Whole-brain analysis

The whole brain cluster-corrected ($p_{FDR} < 0.05$) analysis for emotional > neutral prosody across both tasks revealed a large cluster in the right STG (Table 2). Figure 1 shows the global brain activity for prosodic stimuli compared to baseline (scanner noise) across tasks, Figure 2 the specific brain activity for emotional compared to neutral prosody across tasks.

Table 2. Results of the whole brain cluster-corrected analysis for emotional versus neutral prosody across tasks.

Area	Cluster size (voxels)	P_{FDR} value	Z score	Cluster peaks x	y	z
R STG	673	0.05	4.48	60	-10	-6
			3.95	62	0	-10
			3.95	64	4	-10
			3.71	60	-30	2
			3.47	72	-28	4
			3.42	52	-42	8

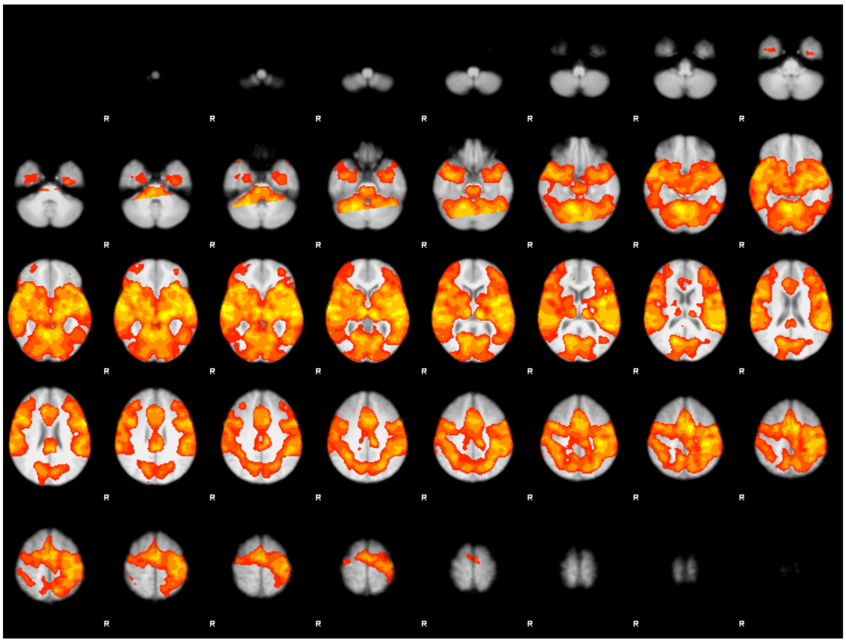

Figure 1. Global brain activity during prosodic processing across tasks at the threshold $P_{uncorr} < 0.001$.

3.2.2. ROI analyses

3.2.2.1. Prosody versus baseline

Figure 3 shows the mean ROI activity (STG, IFG, amygdala) in the left and right hemisphere in response to angry, surprised, and neutral prosody versus baseline in the implicit (top panel) and the explicit (bottom panel) prosody task. RM-ANCOVA with z-scores of angry, surprised, and neutral prosody versus baseline as dependent variable revealed a significant main effect of TAS-20 alexithymia scores as a between-subjects factor [$F(1,15) = 7.119$, $p = 0.018$]. T-STAI scores and Sex as between-subjects factors were not significant [$F < 1$]. There was no main effect of Emotion and no interaction of Emotion with alexithymia, suggesting that effects applied to neutral prosody as well as to angry and surprised prosody. There was a significant main effect of ROI [$F(2,30) = 3.663$, $p = 0.038$] and of Hemisphere [$F(1,15) = 4.785$, $p = 0.045$] as well as a three-way interaction ROI × Hemisphere × TAS-20 scores [$F(2,30) = 3.698$, $p = 0.037$].

Follow-up tests on the factor ROI showed a significant main effect of TAS-20 scores as a between-subjects factor for the amygdala [$F(1,15) = 8.302$, $p = 0.011$], suggesting reduced activity of the left and right amygdala for neutral, angry, and surprised prosody in individuals with high alexithymia scores. For the STG, there was a trend for TAS-20 alexithymia scores as a between-subjects factor [$F(1,15) = 3.883$, $p = 0.068$]. Further, a main effect of Hemisphere [$F(1,15) = 8.353$, $p = 0.011$] and a Hemisphere × TAS-20 scores interaction [$F(1,15) = 8.740$, $p = 0.010$] was found, indicative of reduced activity of the right STG for neutral, angry, and surprised prosody with increasing scores on alexithymia. Pearson's correlations confirmed these effects.

Figure 4 visualizes the correlations of alexithymia with activity of the left and right amygdala and the right STG for neutral, angry, and surprised prosody versus baseline. Additional Pearson's correlations on the three alexithymia facets showed that none of these correlations was driven by a particular facet but rather by the entire personality construct. For the IFG, there was no significant main effect or interaction with TAS-20 alexithymia scores.

3.2.2.2. Emotional versus neutral prosody

A second RM-ANCOVA with z-scores of angry versus neutral and surprised versus neutral prosody as dependent variable was conducted in order to identify emotion-specific effects of alexithymia. The results revealed a trend toward a main effect of TAS-20 alexithymia scores as a between-subjects factor [$F(1,15) = 4.027$, $p = 0.063$] and no main effect of T-STAI and Sex ($F < 1$). There was a significant Task × ROI × Hemisphere × Emotion × TAS-20 interaction [$F(2,30) = 7.989$, $p = 0.002$]. Follow-up Pearson's correlations indicated positive correlations of alexithymia with activity of the left amygdala during the explicit evaluation of angry (> neutral) prosody ($r = 0.547$, $p = 0.015$). During the implicit perception of angry (> neutral) prosody, positive correlations of alexithymia with activity of the left STG ($r = 0.481$, $p = 0.037$) and the left ($r = 0.582$, $p = 0.009$) and right amygdala ($r = 0.515$, $p = 0.024$) were observed as shown in Figure 5.

However, rather than indicating higher ROI activity with increasing alexithymia scores, these seemingly positive correlations were caused by a relatively stronger ROI deactivation in response to neutral prosody (> baseline) than to angry prosody (> baseline) with increasing alexithymia scores, as visualized in Figure 6. Thus, individuals with higher alexithymia scores deactivated the left STG and the bilateral amygdalae less in response to angry prosody than in response to neutral prosody, suggesting that angry prosody was perceived as relatively more salient than neutral prosody.

Additional Pearson's correlations with the three alexithymia facets showed that the correlation with the left amygdala in the explicit and implicit task and the correlation with the left STG in the implicit task were mainly driven by the facet 'difficulty describing feelings'.

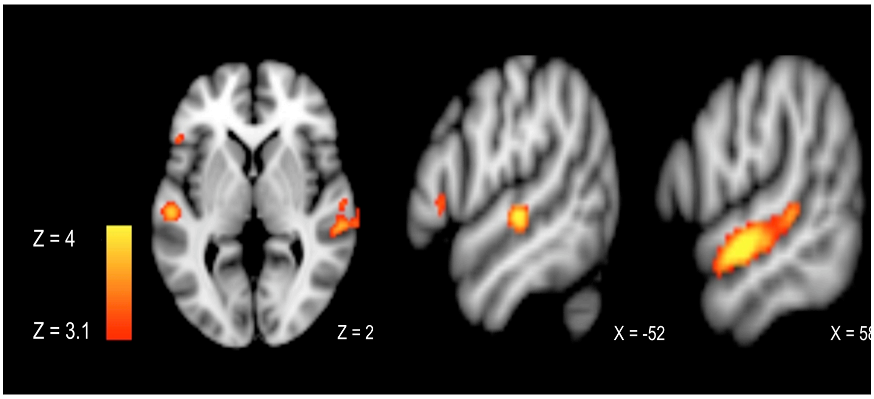

Figure 2. ROI activity for emotional > neutral prosody across tasks at the threshold $P_{uncorr} < 0.001$.

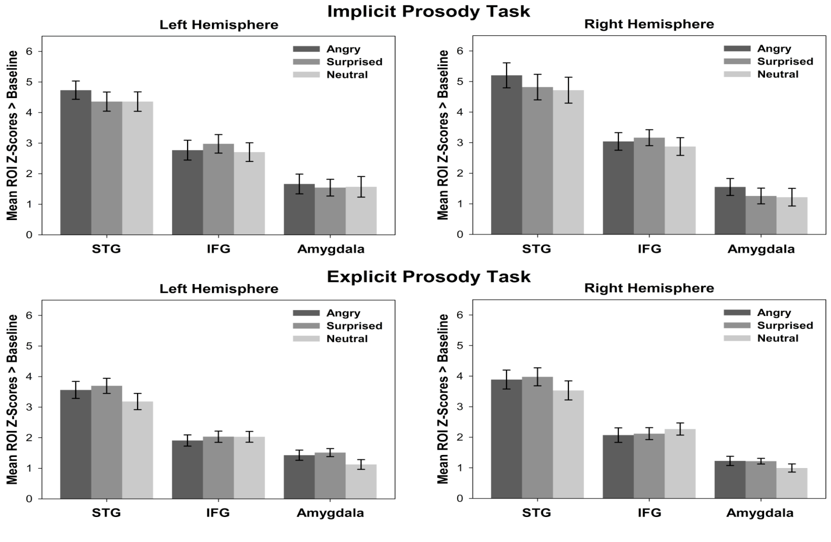

Figure 3. Left-hemispheric (left) and right-hemispheric (right) ROI activity (mean z-scores) in response to angry, surprised and neutral prosody compared with baseline, in the implicit (top) and the explicit prosody task (bottom). Error bars indicate the standard error of the mean.

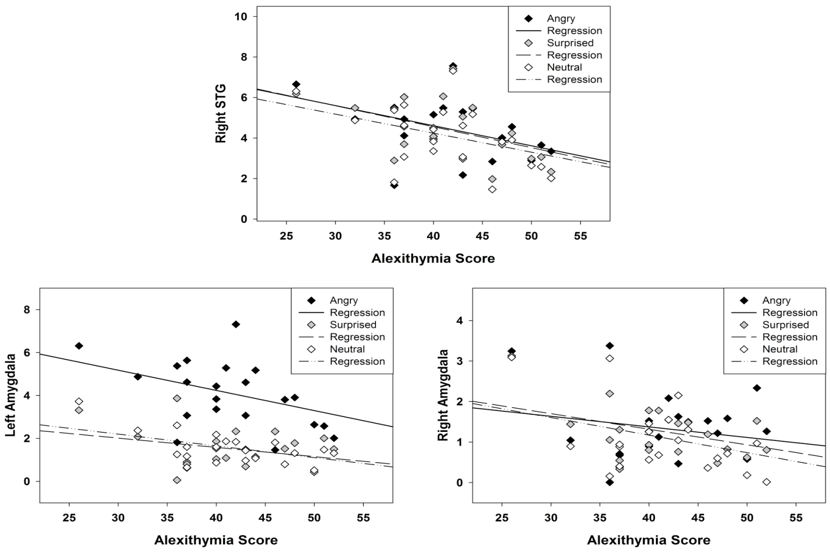

Figure 4. Alexithymia correlations with ROI activity in response to angry, surprised and neutral prosody vs baseline across tasks. Top panel: Negative correlation of alexithymia with activity of the right STG. Bottom panel: Negative correlation of alexithymia with the left and right amygdala.

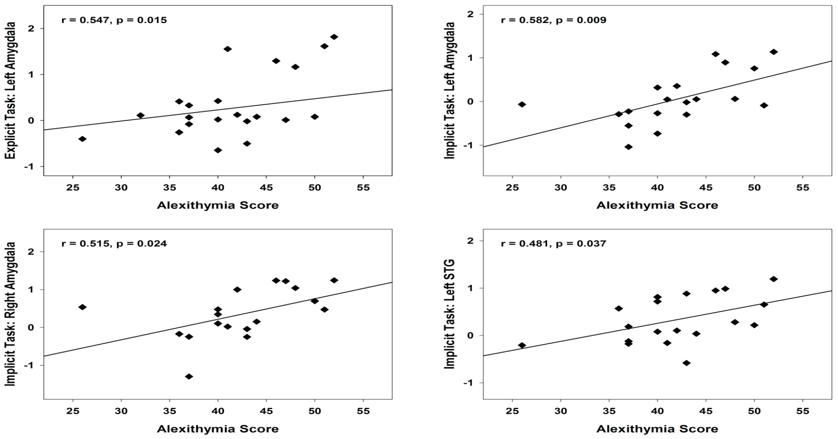

Figure 5. Alexithymia correlations with ROI activity for angry prosody directly contrasted to neutral prosody. Top panel: Positive correlation of alexithymia with activity of the left amygdala in the explicit (left) and the implicit task (right). Bottom panel: Positive correlations of alexithymia with activity of the right amygdala (left) and the left STG (right) in the implicit task.

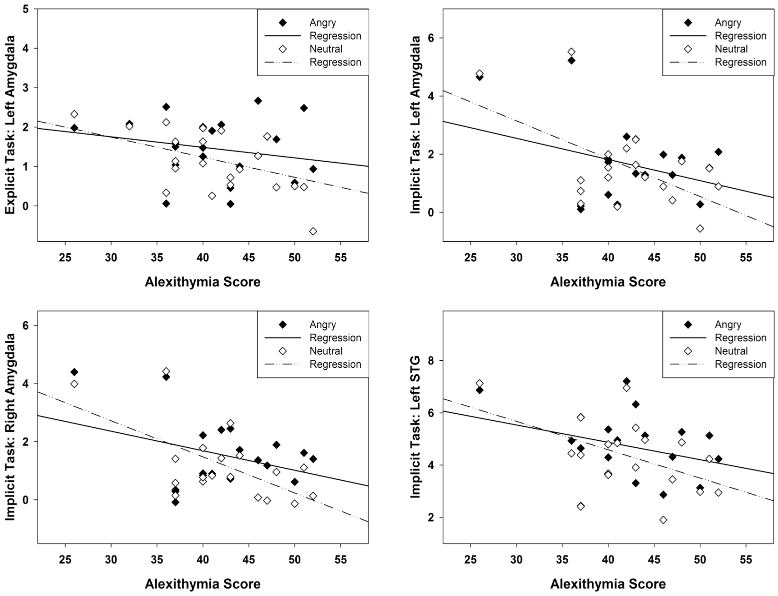

Figure 6. Comparison between alexithymia correlations with ROI activity for neutral prosody > baseline (dashed lines) and angry prosody > baseline (solid lines), demonstrating stronger ROI deactivation for neutral than for angry prosody. Top panel: Negative correlation of alexithymia with activity of the left amygdala in the explicit (left) and the implicit task (right). Bottom panel: Negative correlations of alexithymia with activity of the right amygdala (left) and the left STG (right) in the implicit task.

4. Discussion

The aim of this study was to investigate the impact of alexithymia on the neural processing of emotions conveyed by speech prosody. Our initial analysis contrasting prosody to baseline showed that activity of the right STG and the bilateral amygdalae for angry, surprised, and neutral prosody was reduced with increasing scores on alexithymia, both during implicit and explicit emotional processing. When specifically contrasting emotional versus neutral prosody, we observed a relatively stronger deactivation of the left STG and the bilateral amygdalae for neutral compared to angry prosody, particularly in individuals with difficulty describing feelings.

The present study is the first to investigate the neural basis of emotional prosody processing in alexithymia. Our finding of modulated brain activity in response to emotional prosody confirm previous findings of alexithymia-related deficits in visual emotional processing (e.g., Berthoz et al., 2002; Karlsson et al., 2008; Pouga et al., 2010) and suggest that such aberrant emotional processing extends to the auditory level. Our results further confirm previous EEG findings indicating a reduced sensitivity to the emotional qualities of speech prosody in alexithymia (Goerlich et al., 2011; 2012), and extend these findings by suggesting that such blunted processing may be localized to the STG and the amygdala. In line with our result of a deactivation in these regions, the same regions have previously been observed to be less responsive to facial expressions of emotion in alexithymia (for a review, see Grynberg et al., 2012). Reduced STG activity has been found during the implicit (masked) processing of surprised (Duan et al., 2010) and happy and sad faces (Reker et al., 2010). Deactivation of the amygdala has been observed in response to fearful bodily expressions (Pouga et al., 2010) and during the implicit and explicit processing of emotional faces (Leweke et al., 2004; Kugel et al., 2008). The present findings extend these results by suggesting that the amygdala and the STG also show a blunted response to emotions conveyed through the voice with increasing alexithymia scores. In addition, they confirm that also in the auditory domain, alexithymia affects emotional processing not only when attention is explicitly directed toward the emotional dimension but also when emotion is implicitly processed.

In line with a previous behavioral study on prosodic processing in alexithymia (Swart et al., 2009), we observed no alexithymia-related differences at the behavioral level. This pattern of differences at the neural or electrophysiological level in the absence of behavioral differences has been observed repeatedly in alexithymia during the processing of visual emotional information (Franz et al., 2004; Mériau et al., 2006; Vermeulen et al., 2008). In our previous EEG studies (Goerlich et al., 2011, 2012), we did not observe significant alexithymia-related differences in behavioral performance for angry and sad prosody either. In addition, a recent study investigating prosodic processing as a function of neuroticism found a similar pattern of differences at the neural level in the absence of behavioral differences (Brück et al., 2011a). Thus, it appears that deficits in emotional prosody processing in a non-clinical alexithymia sample such as the current one might be of a rather subtle nature, and can thus be detected at the neural level even though they do not tend to surface at the behavioral level. A subtle deficit in emotional prosody processing seems not surprising considering that alexithymic individuals are generally high-functioning and socially adapted people. The pursuit of social conformity is a characteristic feature of alexithymia (Taylor et al., 1997) and implies learning to interpret emotional signals encountered in social life to the best of one's ability.

In favor of this interpretation, alexithymia was associated with reduced activity in the STG and the amygdala, regions that are involved in earlier phases of emotional prosody processing (Brück et al., 2011b; Kotz & Paulmann, 2011; Wildgruber et al., 2009), but not in the IFG. The IFG is thought to mediate the final phase of emotional prosody processing, in which emotional prosodic information is explicitly evaluated and integrated with other cognitive processes (Brück et al., 2011b). The lack of a modulation of IFG activity in relation to alexithymia could be due to the good performance of participants and the absence of behavioral differences as a function of alexithymia. Taken together, our findings of alexithymia-dependent modulations of neural activity in earlier phases of emotional prosody processing in the absence of such modulations at the final stage and of behavioral differences suggests that alexithymia might predominantly affect earlier stages of prosodic processing, while these processing differences may be compensated in the final (pre-response execution) stage of this process, resulting in adequate behavioral performance.

Interestingly, we found that individuals with high alexithymia scores showed reduced ROI activity not only for emotional (angry and surprised) but also for neutral prosody, irrespective of whether they focused on emotional aspects of the stimuli or not. In fact, direct comparisons between angry and neutral prosody revealed that alexithymia was associated with a significantly higher deactivation of the left STG and the bilateral amygdalae for neutral compared to angry prosody. This association was driven particularly by the alexithymia facet difficulty describing feelings. On the one hand, this suggests that also individuals with difficulty describing feelings perceive an angry tone of voice as relatively more salient and attention-capturing than a neutral one, in line with the well-known phenomenon of emotional stimuli being more attention-capturing than neutral ones (for a review on emotional attention, see Vuilleumier, 2005). On the other hand, this finding indicates that individuals with difficulty describing their feelings may assign less personal significance to human voices conveying emotions, and even less to those using neutral prosody. This would indicate an opposite pattern than in psychopathological conditions such as borderline personality disorder, in which neutral faces can be perceived as threatening (Wagner & Linehan, 1999) and elicit the same degree of amygdala hyperactivity as emotional faces (Donegan et al., 2003), and in schizophrenia, in which amygdala deactivations for fearful compared to neutral faces in fact resulted from increased amygdala activity for neutral faces (Hall et al., 2008). Recent advances in neuroscience suggest that the role of the amygdala does not seem to be restricted to emotional stimuli but that this structure may represent a more general relevance detector for salient, personally and socially relevant, or novel stimuli (for reviews, see Armony, 2013; Sander, Grafan, & Zalla, 2003). For instance, amygdala activity was found to be higher for neutral stimuli if these were socially relevant compared to neutral, non-social stimuli (Vrtička, Sander, & Vuilleumier, 2012). In this line of reasoning, our findings of diminished neural responses to human voices regardless of emotionality could hint to the existence of a more general deficit in the processing of not only emotional but also socially relevant information including speech prosody in alexithymia. However, this hypothesis should be considered speculative and remains to be tested in future studies on the neural basis of alexithymia.

Limitations

While the use of the TAS-20 scale facilitates comparability of our results to previous findings, it should be noted that this scale assesses only the cognitive dimension of alexithymia. Recent evidence suggests that the affective dimension of alexithymia, the dimension of emotional experience, may differentially affect emotional processing (Bermond et al., 2010; Moormann et al., 2008). In addition, while the present study controlled for the influence of trait anxiety, alexithymia may also be associated with depression (Picardi et al., 2011) not assessed here, and levels of depression may alter the perception of emotional prosody (Naranjo et al., 2011; for a review, see Garrido-Vásquez, Jensen, & Kotz, 2011). Thus, it may be worthwhile to take both alexithymia dimensions into account and to additionally control for levels of depression in future studies on prosodic perception in alexithymia.

Conclusions

Alexithymia seems to be associated with a blunted response of the superior temporal gyrus and the amygdala to speech prosody. This diminished response does not seem to be specific for emotional prosody but occurs also for neutral prosody, hinting to the possibility of a more general deficit in the processing of socially relevant information in alexithymia. Neural alterations in the processing of speech prosody may contribute to problems in social communication associated with this personality trait.

5. References

Adolphs, R., & Tranel, D. (1999). Intact recognition of emotional prosody following amygdala damage. *Neuropsychologia, 37,* 1285-1292.

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

Armony, J. L. (2013). Current emotion research in behavioral neuroscience: the role(s) of the amygdala. *Emotion Review, 5,* 104-115.

Bagby, R., Parker, J., & Taylor, G. (1994a). The twenty-item Toronto Alexithymia Scale: I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38,* 23-32.

Bagby, R., Taylor, G., & Parker, J. (1994b). The twenty-item Toronto Alexithymia Scale: II. Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research, 38,* 33-40.

Bermond, B., Bierman, D., Cladder, M., Moormann, P., & Vorst, H. (2010). The cognitive and affective alexithymia dimensions in the regulation of sympathetic responses. *International Journal of Psychophysiology, 75,* 227-333.

Bermond, B., Righart, R., Ridderinkhof, K., & Moormann, P. (2008). Alexithymia and the brain potential P300. *Nederlands Tijdschrift voor de Pyshcologie [Netherlands Journal of Psychology], 64,* 65-77.

Berthoz, S., Artiges, E., Van de Moortele, P. F., Poline, J. B., Rouquette, S., Consoli, S. M., et al. (2002). Effect of impaired recognition and expression of emotions on frontocingulate cortices: An fMRI study of men with alexithymia. *The American Journal of Psychiatry, 159,* 961-967.

Berthoz, S., Consoli, S., Perez-Diaz, F., & Jouvent, R. (1999). Alexithymia and anxiety: Compounded relationships? A psychometric study. *European Psychiatry, 14,* 372-378.

Bird, G., Silani, G., Brindley, R., White, S., Frith, U., & Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain, 133,* 1515-1525.

Brück, C., Kreifelts, B., Kaza, E., Lotze, M., & Wildgruber, D. (2011a). Impact of personality on the cerebral processing of emotional prosody. *NeuroImage, 58,* 259-268.

Brück, C., Kreifelts, B., & Wildgruber, D. (2011b). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Phyics of Life Reviews,* 8, 383-403.

Brühl, A., Viebke, M., Baumgartner, T., Kaffenberger, T., & Herwig, U. (2011). Neural correlates of personality dimensions and affective measures during the anticipation of emotional stimuli. *Brain Imaging and Behavior, 5,* 86-96.

Bydlowski, S., Corcos, M., Jeammet, P., Paterniti, S., Berthoz, S., Laurier, C., et al. (2005). Emotion-processing deficits in eating disorders. *International Journal of Eating Disorders, 37,* 321-329.

Canli, T., Zhao, Z., Desmond, J. E., Kang, E., Gross, J., & Gabrieli, J. D. E. (2001). An fMRI study of personality influences on brain reactivity to emotional stimuli. *Behavioral Neuroscience, 15,* 33-42.

De Houwer, J. (2006). What are implicit measures and why are we using them. In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (219-244). Mahwah, NJ: Erlbaum.

Donegan, H. N., Sanislow, C. A., Blumberg, H. P., Fulbright, R. K., Lacadie, C., Skudlarski, P., . . . Wexler, B. E. (2003). Amygdala hyperreactivity in borderline personality disorder: implications for emotional dysregulation. *Biological Psychiatry, 54,* 1284-1293.

Duan, X., Dai, Q., Gong, Q., & Chen, H. (2010). Neural mechanism of unconscious perception of surprised facial expression. *NeuroImage, 52,* 401-407.

Ethofer, T., Kreifelts, B., Wiethoff , S., Wolf, J., Grodd, W., Vuilleumier , P., & Wildgruber, D. (2009). Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody. *Journal of Cognitive Neuroscience, 21,* 1255-1268.

Frühholz, S., Ceravolo, L., & Grandjean, D. (2011). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex, 22,* 1107-1117.

Frühholz, S., & Grandjean, D. (2013). Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: A quantitative meta-analysis. *Neuroscience and Biobehavioral Reviews, 37,* 24-35.

Frühholz, S., & Grandjean, D. (2013). Amygdala subregions differentially respond and rapidly adapt to threatening voices. *Cortex, 49,* 1394-1401.

Garrido-Vásquez, P., Jessen, S., & Kotz, S. A. (2011). Perception of emotion in psychiatric disorders: on the possible role of task, dynamics, and multimodality. *Social Neuroscience, 6,* 515-536.

Goerlich, K. S., Aleman, A., & Martens, S. (2012). The sound of feelings: Electrophysiological responses to emotional speech in alexithymia. *PloS ONE, 7, e36951.*

Goerlich, K. S., Witteman, J., Aleman, A., & Martens, S. (2011). Hearing feelings: Affective categorization of music and speech in alexithymia, an ERP study. *PloS ONE, 6,* e19501.

Grynberg, D., Luminet, O., Corneille, O., Grezes, J., & Berthoz, S. (2010). Alexithymia in the interpersonal domain: A general deficit of empathy? *Personality and Individual Differences, 49,* 845-850.

Grynberg, D., Chang, B., Corneille, O., Maurage, P., Vermeulen, N., Berthoz, S., & Luminet, O. (2012). Alexithymia and the processing of emotional facial expressions (EFEs): Systematic review, unanswered questions and further perspectives. *PloS ONE, 7,* e42429.

Guttman, H., & Laporte, L. (2002). Alexithymia, empathy, and psychological symptoms in a family context. *Comprehensive Psychiatry, 43,* 448-455.

Hall, J., Whalley, H. C., McKirdy, J. W., Romaniuk, L., McGonigle, D., McIntosh, A. M., . . . Lawrie, S. M. (2008). Overactivation of fear systems to neutral faces in schizophrenia. *Biological Psychiatry, 64,* 70-73.

Hamann, S., & Canli, T. (2004). Individual differences in emotion processing. *Current Opinion In Neurobiology, 14,* 233-238.

Hooker, C., Verosky, S., Miyakawa, A., Knight, R., & D'Esposito, M. (2008). The influence of personality on neural mechanisms of observational fear and reward learning. *Neuropsychologia, 46,* 2709-2724.

Jacob, H., Kreifelts, B., Brück, C., Erb, M., Hösl, F., & Wildgruber, D. (2012). Cerebral integration of verbal and nonverbal emotional cues: Impact of individual nonverbal dominance. *NeuroImage, 61,* 738-747.

Karlsson, H., Näätänen, P., & Stenman, H. (2008). Cortical activation in alexithymia as a response to emotional stimuli. *The British Journal of Psychiatry, 192,* 32-38.

Kooiman, C., Spinhoven, P., & Trijsburg, R. (2002). The assessment of alexithymia: A critical review of the literature and a psychometric study of the Toronto Alexithymia Scale-20. *Journal of Psychosomatic Research, 53,* 1083-1090.

Kotz, S., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass, 5,* 108-125.

Kugel, H., Eichmann, M., Dannlowski, U., Ohrmann, P., Bauer, J., Arolt, V., . . . Suslow, T. (2008). Alexithymic features and automatic amygdala reactivity to facial emotion. *Neuroscience Letters, 435,* 40-44.

Leitman, D. I., Wolf., D. H., Ragland, J. D., Laukka, P., Loughead, J., Valdez, J. N., . . . Gur, R. C. (2010). "It's not what you say but how you say it": A reciprocal temporo-frontal network for affective prosody. *Frontiers in Human Neuroscience, 4,* 19-31.

Leweke, F., Stark, R., Milch, W., Kurth, R., Schienle, A., Kirsch, P., . . Vaitl, D. (2004). Patterns of neuronal activity related to emotional stimulation in alexithymia. *Psychotherapy and Psychosomatic Medicine, 54,* 437-444.

Mantani, T., Okamoto, Y., Shirao, N., Okada, G., & Yamawaki, S. (2005). Reduced activation of posterior cingulate cortex during imagery in subjects with high degrees of alexithymia: A functional magnetic resonance imaging study. *Biological Psychiatry, 57,* 982-990.

Mériau, K., Wartenburger, I., Kazzer, P., Prehn, K., Lammers, C. H., Van der Meer, E…Heekeren, H. R. (2006). A neural network reflecting individual differences in cognitive processing of emotions during perceptual decision making. *NeuroImage, 33,* 1016-1027.

Moormann, P., Bermond, B., Vorst, H., Bloemendaal, A., Teijn, S., & Rood, L. (2008). New avenues in alexithymia research: The creation of alexithymia types. *Emotion regulation: Conceptual and clinical issues* [e-book]. New York, NY US: Springer Science + Business Media; 27-42.

Mothes-Lasch, M., Mentzel, H., Miltner, W., & Straube, T. (2011). Visual attention modulates brain activation to angry voices. *The Journal of Neuroscience, 31, 9594*-9598.

Naranjo, C., Kornreich, C., Campanella, S., Noël, X., Vandriette, Y., Gillain, B., . . Constant, E. (2011). Major depression is associated with impaired processing of emotion in music as well as in facial and vocal stimuli. *Journal of Affective Disorders, 128,* 243-51.

Ormel, J., Bastiaansen, J. A., Riese, H., Bosa, E. H., Servaas, M., Ellenbogen, M., . . . Aleman, A. (2013). The biological and psychological basis of neuroticism: Current status and future directions. *Neuroscience and Biobehavioral Reviews, 37,* 59-72

Parker, P. D., Prkachin, K. M., & Prkachin, G. C. (2005). Processing of facial expressions of negative emotion in alexithymia: the influence of temporal constraint. *Journal of Personality, 73,* 1087-1107.

Picardi, A., Fagnani, C., Gigantesco, A., Toccaceli, V., Lega, I., & Stazi, M. (2011). Genetic influences on alexithymia and their relationship with depressive symptoms. *Journal of Psychosomatic Research, 71,* 256-263.

Pouga, L., Berthoz, S., De Gelder, B., & Grèzes, J. (2010). Individual differences in socioaffective skills influence the neural bases of fear processing: The case of alexithymia. *Human Brain Mapping, 31,* 1469-1481.

Prkachin, G., Casey, C., & Prkachin, K. (2009). Alexithymia and perception of facial expressions of emotion. *Personality and Individual Differences, 46,* 412-417.

Reker, M., Ohrmann, P., Rauch, A. V., Kugel, H., Bauer, J., Dannlowski, U., Arolt, V… Suslow, T. (2010). Individual differences in alexithymia and brain response to masked emotion faces. *Cortex, 46,* 658-667.

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40,* 227-256.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: An evolved system for relevance detection. *Reviews in the Neurosciences, 14,* 303-316.

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *NeuroImage, 28,* 848-858.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Science, 10,* 24-30.

Sifneos, P. E. (1973). The prevalence of 'alexithymic' characteristics in psychosomatic patients. *Psychotherapy & Psychosomatics, 22,* 255-262.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *STAI Manual for the State-Trait Anxiety Inventory.* Palo Alto: Consulting Psychologists Press.

Spitzer, C., Siebel-Jurges, U., Barnow, S., Grabe, H. J., & Freyberger, H. J. (2005). Alexithymia and interpersonal problems. *Psychotherapy & Psychosomatics, 74,* 240-246.

Swart, M., Kortekaas, R., & Aleman, A. (2009). Dealing with feelings: Characterization of trait alexithymia on emotion regulation strategies and cognitive-emotional processing. *PLoS ONE, 4,* e5751.

Taylor, G. J., Bagby, R. M., & Parker, J. D. A. (1997). *Disorders of Affect Regulation: Alexithymia in Medical and Psychiatric Illness.* Cambridge: Cambridge University Press.

Vanheule, S., Desmet, M., Meganck, R., & Bogaerts, S. (2007). Alexithymia and interpersonal problems. *Journal of Clinical Psychology, 63,* 109-117.

Van Kleef, G. (2009). How emotions regulate social life: The emotions as social information (EASI). model. *Current Directions in Psychological Science, 18,* 184-188.

Vrtička, P., Sander, D., & Vuilleumier, P. (2012). Lateralized interactive social content and valence processing within the human amygdala. *Frontiers in Human Neuroscience, 6*: 358.

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences, 9,* 585-594.

Wagner, A. W., & Linehan, M. M. (1999). Facial expression recognition ability among women with borderline personality disorder: Implications for emotion regulation? *Journal of Personality Disorders, 13,* 329-344.

Wiethoff, S., Wildgruber, D., Grodd, W., & Ethofer, T. (2009). Response and habituation of the amygdala during processing of emotional prosody. *Neuroreport, 20,* 1356-1360.

Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology, 11,* 277-281.

Wildgruber, D., Ethofer, T., Kreifelts, B., & Grandjean, D. (2008). Cerebral processing of emotional processing: A network model based on fMRI studies. In SP-2008, pp 217-22. http://isca-seech.org/archive/sp2008/papers/2p08_217.pdf.

Wildgruber, D., Riecker, A, Hertrich, I., Erb, M., Grodd, W., Ethofer, T., et al. (2005). Identification of emotional intonation evaluated by fMRI. *NeuroImage, 24,* 1233-1241.

Witteman, J., Van Heuven, V. J. P., & Schiller, N. O. (2012). Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia, 50,* 2752-2763.

Witteman, J., Van IJzendoorn, M., Van de Velde, D., Van Heuven, V. J. P., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. *Neuropsychologia, 49,* 3722-3738.

Woolrich, M., Jenkinson, M., Brady, J., & Smith, S. (2004). Fully Bayesian spatio-temporal modelling of FMRI data. *IEEE Transactions on Medical Imaging, 23,* 213-231.

Appendix A. Pseudowords used in the experiments

konpon
dinpil
duldin
kondon
duldun
daldan
paldan
dalpal

Chapter 4

Summary and conclusion

4.1. Recapitulation of research questions

Imagine again that you are sitting in a train, absorbed by a good book, and suddenly hearing a mother shouting angrily to her child: "Tim, you are going to annoy the other passengers!" As we have observed in Chapter 1 ('Introduction'), many people would be familiar with the involuntary tendency to turn one's head into the direction of Tim's mother, revealing that the utterance apparently constitutes a powerful social signal that our brain seems to process even we when we do not intend to.

In Chapter 1, I have explained that the communicative layer of speech that Tim's mother is using to convey anger is called 'prosody' in linguistics. Further, we have elucidated that prosody cannot only be used to convey emotional meaning (which we have called 'emotional prosody') but also part of the linguistic structure of an utterance (which we have called 'linguistic prosody'). The first empirical section of the present thesis was concerned with *what* network in the brain supports these two communicative functions of prosody, *how* the network does so (i.e., using what series of operations), *whether* this network indeed sometimes processes this information 'automatically' and, if so, *why* this might be the case. Concerning the last question, we have hypothesized that the existence of a 'hard-wired' system in the brain that prioritizes the processing of social signals that indicate potential harm to the organism (such as an angry mother, in the example above) can explain such automaticity of processing (Öhman & Mineka, 2001). To test this hypothesis, we have directly contrasted processing of emotional prosody with an artificial acoustic signal that has probably emerged more recently in evolutionary history (and is therefore probably not supported by a hard-wired dedicated system) that can also powerfully convey emotion – *music*.

Although the basic architecture of the prosody perception network would be expected to be relatively stable across individuals (otherwise, it would be hard to explain why communication through the prosodic layer of speech is relatively stable in its efficiency across individuals), there are also subtler differences in affective processing style among individuals (Canli, 2004). One personality trait that is associated with affective processing style is called *alexithymia* ('no words for feelings'), a trait that is associated with difficulties in recognizing, identifying and verbalizing emotions (Sifneos, 1973). As variation along the alexithymia continuum is associated with emotional processing differences, it is plausible that such differences are reflected in the emotional prosody perception network. Therefore, in the second empirical section of the present thesis, it was investigated whether (non-clinical) variation in alexithymia indeed modulates activity within the emotional prosody perception network, and if so, whether alexithymia is primarily reflected at a relatively early (and hence potentially 'automatic') or a relatively late stage of emotional processing (or both).

4.2. Towards a cognitive neuroscience of prosody perception and its modulation by alexithymia

4.2.1. Interhemispheric models of prosody perception

In Figure 1 (identical to Figure 1 of Chapter 1, repeated here for ease of reference), a schematic visualization is presented of the prosody perception pathway, as hypothesized by the two major models of (emotional) prosody perception (Schirmer & Kotz, 2006; Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009) that existed when the work for the present thesis was initiated. As has been pointed out in Chapter 1 and will be explained in more detail in the next section, these models suggest that prosody perception is a multi-stage process, with elementary acoustic processing first taking place in the primary auditory cortex (Heschl's gyrus, or HG) and the middle part of the superior temporal gyrus (m-STG). Subsequently, more abstract and complex auditory processing would take place in either the anterior STG (a-STG) or posterior STG (p-STG). Finally, abstract evaluation of prosody is hypothesized to be sub-served by the inferior frontal gyrus (IFG).

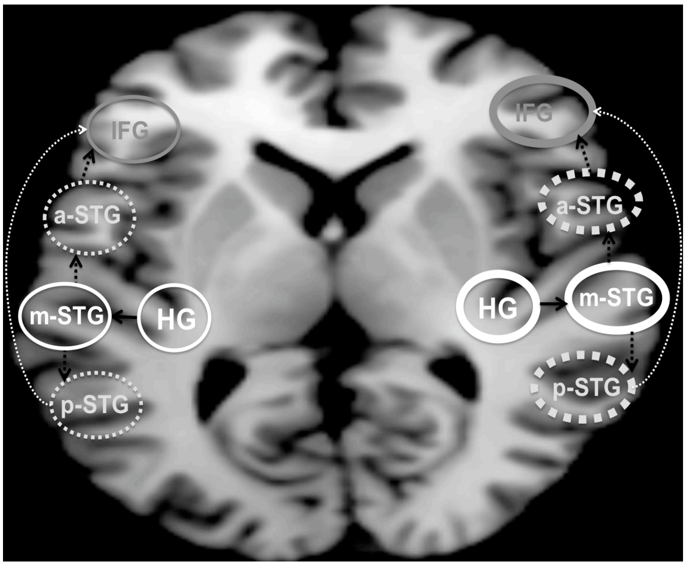

Figure 1 (repetition of Figure 1 of Chapter 1). Schematic visualization of the prosody perception pathway as hypothesized by two dominant models of (emotional) prosody perception, superimposed on an axial slice of an MRI scan. Solid circles and arrows indicate regions and connections that are hypothesized to be important in (emotional) prosody perception by both models. Dashed circles and arrows indicate areas and connections hypothesized to be involved in prosody perception by only one of the two models. White circles indicate areas hypothesized to be involved in the first stage, light-grey circles in the second stage and dark-grey circles in the final stage of prosody perception. Bold circles in the right hemisphere as compared to the left indicate hypothesized right hemispheric superiority for emotional prosody perception.
Abbreviations: HG = Heschl's Gyrus, m-STG = mid superior temporal gyrus, p-STG = posterior superior temporal gyrus, a-STG = anterior superior temporal gyrus, IFG = inferior frontal gyrus.

Early work in the cognitive neuroscience of prosody perception used the lesion-deficit approach in an attempt to answer the question of whether one cerebral hemisphere might be superior at (or 'specialized' in) processing of emotional prosody. As I have argued in Chapter 1, lesion-deficit data indeed provide the most powerful test of hemispheric specialization hypotheses since such data inform us about whether a hemisphere is *necessary* for a hypothesized mental function such as prosody perception. Three hypotheses concerning hemispheric specialization for prosody perception had been put forward:

(1) The right-hemisphere hypothesis (Ross, 1981) proposes that the right hemisphere is specialized in emotional prosody perception;
(2) The *functional* lateralization hypothesis (Van Lancker, 1980) adds that the hemispheres are specialized in the processing of different *functional categories* with the left and right hemisphere being specialized at the processing of linguistic and emotional categories, respectively;
(3) *Acoustic* lateralization hypotheses, in contrast, propose that hemispheric specialization for prosody perception can be traced back to specialization of the hemispheres for different acoustic dimensions of the speech signal. A prominent acoustic lateralization hypothesis states that the left hemisphere is better at processing of temporal information while the right is better at the processing of spectral information (Van Lancker & Sidtis, 1992). Because spectral information is important for the perception of emotional prosody, a right-hemisphere specialization for emotional prosody perception could then be explained by a low-level specialization for spectral processing.

Note that the functional lateralization hypothesis can be seen as an elaboration of the right-hemisphere hypothesis. Further note that the acoustic and functional lateralization hypotheses are not mutually exclusive – they could both be true and represent different stages of the prosody perception process (symbolized by the circles with different shades of grey in Figure 1).

When the work for the present thesis was initiated, the lesion-deficit literature had not provided consensus about whether there is hemispheric specialization for prosody perception and, if so, which of the above hypotheses can best explain it. Therefore, in Chapter 2.1, a test of these hypotheses was performed using quantitative meta-analysis of the lesion-deficit literature. Using the power of the combined sample size of all lesion-deficit studies on emotional and linguistic prosody perception published to date, it was found that *both* hemispheres are necessary for emotional and linguistic prosody perception (i.e., damage to each hemisphere significantly degraded linguistic and emotional prosodic perception performance). However, damage to the right hemisphere was more detrimental to emotional prosody perception performance than equivalent left-hemispheric damage. Thus, we found meta-analytic evidence for *relative* right-hemispheric specialization for emotional prosody perception and no evidence in favor of hemispheric specialization for linguistic prosody perception. In conclusion, based on the lesion-deficit studies to date, there is no support for the *functional* lateralization hypothesis (because no left-hemispheric specialization for linguistic prosody perception was found) and only evidence in favor of a weak (relative) version of Ross' (1981) right hemisphere hypothesis. However, the meta-analysis did not allow for a direct test of the acoustic lateralization hypothesis versus the functional lateralization hypothesis.

Therefore, in Chapter 2.2, a systematic test of the functional versus acoustic lateralization hypotheses of prosody perception was performed. This was achieved by presenting participants with bi-dimensional pseudowords that had both either angry or sad prosody and either linguistic stress on the first or second syllable. Exactly the same pseudowords were presented dichotically to participants while event related potentials (ERPs) were recorded. However, about half of the participants were instructed to categorize the *emotional* prosody dimension of the stimuli, while the other half categorized the *linguistic* prosody dimension. If the functional lateralization hypothesis were correct, we would expect to find a shift from a right to left hemispheric advantage in activation over the scalp for the emotional versus the linguistic task. Note that, since the acoustic material that was presented to both groups of participants was identical, such a shift would be evidence for the hypothesis that the functional processing mode (emotional versus linguistic mode of prosody perception) *pre se* can indeed drive hemispheric asymmetry. However, no such shift in hemispheric advantage was found, neither as indicated by the ear advantage nor on the electrophysiological level, even though the statistical power of our study was relatively high. Thus, in Chapter 2.2 we did not find evidence in favor of the functional lateralization hypothesis of prosody perception, which is in keeping with the results of the quantitative meta-analysis performed in the preceding chapter.

Can acoustic lateralization hypotheses, then, better account for the relative right-hemispheric specialization for emotional prosody found using meta-analysis of the lesion-deficit literature in Chapter 2.1? Note that, if the relative right-hemispheric specialization would indeed be driven by superiority of the right hemisphere in processing of acoustic cues relevant for emotional prosody perception, we would expect a right-hemispheric advantage to emerge somewhere relatively early (at the level of the auditory processing centers) in the prosody perception pathway visualized in Figure 1. In Chapter 2.3 of this thesis, a quantitative meta-analysis was performed on the neuroimaging literature of emotional prosody perception, which will be discussed in more detail in the next section. This meta-analysis revealed a bilateral temporo-frontal network involved in emotional prosody perception, again illustrating that *both* hemispheres are involved in emotional prosody perception. Formal meta-analyses of hemispheric asymmetry, however, revealed higher activation probability in the right transverse temporal gyrus (HG) and p-STG than in their left-hemispheric homotopes, albeit only at a liberal statistical threshold. Thus, meta-analytic evidence again points to *relative* hemispheric specialization for emotional prosody perception in the context of *bi*hemispheric involvement. Further, there is initial evidence that relative right-hemispheric specialization for emotional prosody perception is driven by specialization for elementary (HG) and more abstract (p-STG) acoustic processing within the prosody perception pathway, in line with *acoustic* lateralization hypotheses of emotional prosody perception. This result is also in line with our primary neuroimaging study of emotional prosody perception reported in Chapter 2.4, where we found a relative rightward hemispheric asymmetry for emotional prosody perception across tasks in the STG.

To summarize, regarding interhemispheric models of prosody perception, the present thesis finds evidence in support of *relative* right-hemispheric specialization for emotional prosody perception and no hemispheric specialization for linguistic prosody perception, based on meta-analysis of the lesion-deficit literature. Further, in a primary study using electroencephalography to systematically test the *functional* lateralization hypothesis, we do not find evidence for the functional lateralization hypothesis of prosody perception. Quantitative meta-analysis of the neuroimaging literature further points to the involvement of a *bi*hemispheric network in emotional prosody perception. Formal meta-analytic analyses of hemispheric asymmetry within this network, suggest that relative right-hemispheric specialization for emotional prosody perception is driven by superiority of the right-hemisphere auditory processing centers, in line with *acoustic* lateralization hypotheses of emotional prosody perception. These results are in keeping with a mounting evidence base in support of hemispheric specialization for basic dimensions of the speech signal (Boemio, Fromm, Braun, & Poeppel, 2005) and hence bottom-up explanations of hemispheric specialization for more complex auditory signals, such as prosody.

In Figure 2, the insights gained from this thesis regarding the interhemispheric models of prosody perception are visualized by the bold circles that indicate a hemispheric advantage in the right prosody perception pathway as compared to the left, based on initial meta-analytic evidence.

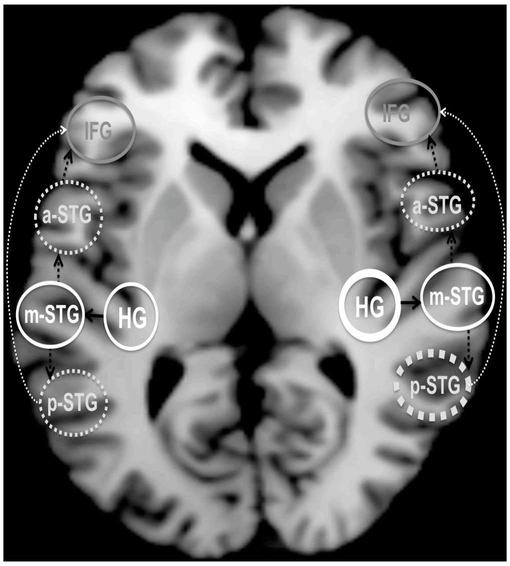

Figure 2. Adaptation of the model presented in Figure 1 based on the insights gained from the present thesis regarding interhemispheric models of prosody perception. The bold circles indicate right hemispheric processing centers in the emotional prosody perception pathway where initial evidence for a relative right hemispheric advantage has been found in this thesis (for further details, see Figure 1).

4.2.2. Intrahemispheric models of prosody perception

Functional magnetic resonance imaging (fMRI) allows us to further probe the brain to investigate which areas within each cerebral hemisphere are involved in prosody perception. When the work for this thesis was initiated, there already was a substantial body of fMRI literature available on emotional prosody perception. These neuroimaging studies had inspired two very similar three-stage models (Schirmer & Kotz, 2006; Wildgruber, Ethofer, Grandjean, & Kreifelts, 2009) of emotional prosody perception. In Figure 2, a schematic visualization can be found of the neural network of prosody perception these models propose (and the disagreement between the two models). Briefly, both models propose that in an initial stage (white circles in Figure 2) basic acoustic processing of (emotional) prosody takes place in the transverse temporal gyrus or Heschl's gyrus (HG) and the middle superior temporal gyrus (m-STG). Subsequently, in a second stage (light-grey circles in Figure 2), both models propose that more complex acoustic processing takes place to integrate the auditory information from stage one into more abstract units of representation, but the models disagree regarding the location of this processing stage. While the Wildgruber model places stage-two prosody perception in the posterior STG (p-STG), the Kotz model positions it further down the auditory object recognition pathway (the 'what' stream) in the anterior STG (a-STG). Finally, both models propose that in a third stage (dark-grey circles in Figure 2), abstract evaluation of prosody takes place in the inferior frontal gyrus (IFG).

In Chapter 2.3, a quantitative meta-analysis of functional neuroimaging studies was performed to test with high statistical power which of these two models can best account for the neuroimaging literature to date. Two functional contrasts were meta-analyzed, (i) a low-level contrast capturing all hypothesized phases of the emotional perception process, and (ii) a high-level contrast that captures more abstract (stage two and three) emotional prosody perception. The low-level contrast revealed above-chance convergence of activation likelihood in the bilateral m-STG continuing medially into HG, p-STG and the IFG. However, no significant convergence of activation likelihood was found in the a-STG. In the high-level contrast, significant convergence of activation likelihood was found in the right p-STG and IFG. Importantly, no significant convergence was found in either contrast for the amygdala, a structure hypothesized to be important for emotional perception (Schirmer & Kotz, 2006) and to be the evolved neural structure supporting automatic processing (Öhman & Mineka, 2001), an issue we will return to in the next paragraph. On the one hand, the amygdala might just not be crucial for emotional prosody perception, as suggested by lesion studies (Adolphs & Tranel, 1999; Bach, Hurlemann, & Dolan, 2013). However, as has been pointed out before, it is possible that the amygdala quickly habituates to emotional prosody, preventing its detection with fMRI paradigms that typically use sustained stimulation (Wiethoff, Wildgruber, Grodd, & Ethofer, 2009; but see Scheuerecker et al., 2007). Thus, the quantitative meta-analyses reported in Chapter 2.3 confirmed that HG and m-STG are likely to be involved in stage one and the IFG in stage three prosody perception, respectively, as proposed by both intrahemispheric models of prosody perception. Concerning stage-two prosody perception, however, our meta-analyses suggest that this more abstract acoustic processing stage is more likely to be supported by the p-STG as predicted by the Wildgruber model than the a-STG as predicted by the Kotz model. In Figure 3, a final modified schematic illustration of the prosody perception pathway is provided based on our meta-analyses of the neuroimaging literature on emotional prosody perception.

Next, we asked whether automaticity of emotional prosody perception can be demonstrated in the emotional prosody perception network illustrated in Figure 3. As I have explained in Chapter 1, automaticity of prosody perception can be demonstrated at the neural level, by diverting attention away from the emotional prosody dimension of an utterance and observing whether there is continuation of above-threshold neural processing of emotional prosody (Anderson, Christoff, Panitz, De Rosa, & Gabrieli, 2003). If we find such sustained neural processing of emotional prosody when participants do not intend to analyze the emotional prosody, this is evidence for the 'unintentionality feature' of automaticity of processing, as proposed by Moors and De Houwer (2001). Note that, based on Darwin's theory of evolution, it has been proposed that a hard-wired system dedicated to the perception of emotional prosody could explain automaticity for prosody perception, and that such a dedicated system would be particularly plausible for emotional information that signals potential harm to the organism (Öhman & Mineka, 2001; Anderson, Christoff, Panitz, De Rosa, & Gabrieli, 2003; Schupp et al., 2004). If this hypothesis were true, we would particularly expect automatic perception of emotional prosody that signals *threat*, i.e., anger prosody. Thus in Chapter 2.4, using fMRI, we investigated whether continued supra-threshold activation could be found for anger prosody and, as a control, a non-threatening emotional category (surprise) when participants do not intend to analyze emotional prosody. It was found that for surprise no significant supra-threshold activation could be demonstrated when attention was diverted from the emotional prosody. For anger processing, however, continued supra-threshold activation was found in the right STG when subjects did not attend to the emotional prosody. Thus, in Chapter 2.4, we provide evidence for automatic processing of emotional prosody, but only when emotional prosody signals potential harm (i.e. anger), as predicted by models that assume an evolved dedicated system for the detection of social information that signals potential harm to the organism (Öhman & Mineka, 2001). However, our results suggest that the neural substrate does not reside in the amygdala but in the right superior temporal gyrus. Thus, also based on our meta-analyses of the lesion and neuroimaging literature above, it might be the specialized acoustic processing centers in the right hemisphere instead of the amygdala that have evolved to detect danger in the auditory modality. Alternatively, as has been pointed out above, it might be the case that these auditory processing centers are engaged by the amygdala first, but that activation of the amygdala itself is missed by fMRI because the amygdala quickly habituates to emotional prosody stimulation.

As I have elucidated in Chapter 1, if automaticity of emotional prosody perception can be explained by an evolved hard-wired system dedicated to the detection of emotional states of conspecifics, we would not expect automatic processing of a probably more recently invented artificial signal that can powerfully convey emotion – *music*. Alternatively, cross-cultural recognition of music by culturally disparate groups suggests that there might be a dedicated biological system for the recognition of emotion in music, too (Fritz et al., 2009) pointing to a potential fitness value of emotional music perception. Further, the 'super-expressive voices' hypothesis (Juslin & Västfjäll, 2008) proposes that emotional music obtains its powerful emotional expression capability by imitating and subsequently exaggerating the acoustic properties of emotional prosody. Thus, by imitating emotional prosody, music could engage a system dedicated to the processing of emotional prosody and as such be processed automatically, too. If any of these two alternative hypotheses were true, we would expect automaticity of emotional music perception as well.

Recall that, in addition to the unintentionality feature of automaticity discussed above, rapid processing has been proposed to be a feature of automaticity (Moors & De Houwer, 2001). As was explained in Chapter 1, the affective priming paradigm (Fazio, 2001) takes advantage of the hypothesized fast processing of emotional information by first presenting an emotional stimulus (the affective prime) that is either positive or negative and almost immediately (typically 200 ms) thereafter presenting a second emotional stimulus (the target). Participants are required to categorize targets as positive or negative. The affective prime hence can be either congruent or incongruent with the valence of the affective target. If the affective prime is indeed processed very rapidly (i.e. in less than 200 ms), it should be able to facilitate responses to congruent targets but inhibit responses to incongruent targets. In Chapter 2.5 we presented participants with very short (800 ms) segments of emotional prosody and emotional music that were either happy or sad. Shortly (200 ms) after the onset of these affective primes, positive and negative affective target words were presented, which participants were required to categorize with respect to valence. While participants were engaged in the affective priming task, ERPs were recorded. We found a significant affective priming effect (APE) for prosody but not for music. Further, the so-called N400 effect was observed for incongruent vs. congruent trials for prosody and music, which had previously been proposed to be associated with automatic spreading of activation. However, in a second study, participants performed exactly the same tasks but with attention diverted from the affective dimension of the emotional targets. This time no APEs or N400's were found for prosody (nor for music). Thus, although mainly emotional prosody (but not emotional music) seems to fulfill the rapid processing feature of automaticity as evidenced by a significant APE in the first experiment, it did not fulfill the unintentionality criterion as measured in the second experiment. Moreover, the presence of the N400 in the first study combined with its absence in the second, suggested that such rapid APE effects are caused by response level interference.

Thus, to summarize the results regarding automaticity of processing, the present thesis does find evidence for the unintentionality criterion of automaticity for emotional prosody perception, but only when prosody signals threat, as predicted by phylogenetically inspired models that propose a hard-wired neural system dedicated to the detection of social information that signals potential harm to the organism. Further, we find evidence for the rapid processing criterion of automaticity for happy and sad emotional prosody but not for affective music. Further, happy and sad prosody do not fulfill the unintentionality criterion of automaticity as measured with the affective priming paradigm. Thus, we do not find strong evidence for a dedicated system for the perception of emotional music. This (though somewhat tenuously) could be understood within the theory of evolution, if we assume that emotional music perception either had no fitness value or emerged too late in evolutionary history to prompt the evolution of a dedicated system. Also, we do not find evidence in favor of the 'super-expressive voices' hypothesis, as on the basis of this hypothesis we would have expected stronger affective priming effects for music than prosody. Further, the result that APEs for prosody do not persist when attention is diverted away from the affective dimension of the stimuli, suggests that automaticity for emotional prosody perception is *relative* (i.e. processes can be more or less automatic, but even relatively strong automatic processes will cease to operate when attentional resources are sufficiently depleted), as has been suggested previously (Pessoa, McKenna, Gutierrez, & Ungerleider, 2002; Mothes-Lasch, Miltner, & Straube, 2012).

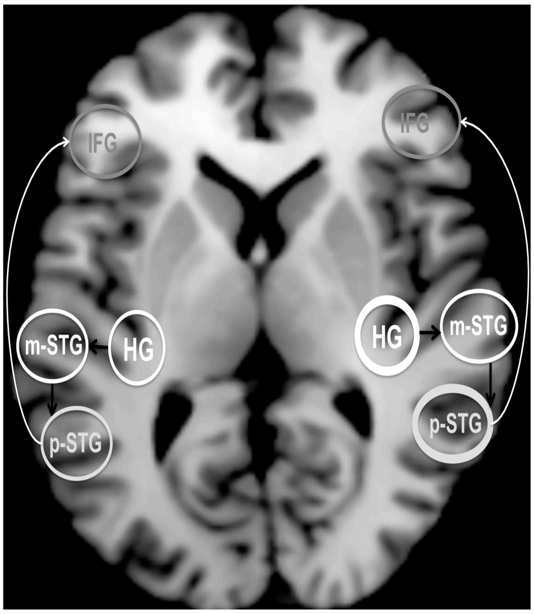

Figure 3. Final adaptation of the model presented in Figure 1 based on the insights gained from the present thesis regarding inter- and intrahemispheric models of prosody perception. White circles indicate the most likely neural substrate for stage-one, light grey circles for stage-two, and dark-grey circles for stage-three emotional prosody perception. Bold circles indicate right-hemispheric processing centers in the emotional prosody perception pathway where initial evidence has been found for greater activation likelihood as compared to their left hemispheric homotopes. For further details, see Figure 1.

4.2.3. Modulation of activity in the emotional prosody perception network by alexithymia

In the second empirical part of the present thesis, we asked whether the personality trait *alexithymia* might modulate processing within the emotional prosody perception network, and if so, whether relatively early (and possibly automatic) or relatively late emotional perception processes are associated with (normal variation in-) alexithymia.

In Chapter 3.1 we investigated whether alexithymia affects automatic processing of emotional prosody and music, as measured with the affective priming paradigm and concurrently recorded ERPs. Alexithymia did not significantly affect the behavioral APEs found during affective categorization. However, at the electrophysiological level alexithymia was associated with a reduced N400 component for affectively incongruent primes and targets. These results point to modulation by alexithymia of relatively automatic (and hence 'early') affective processing, without effects being evident at the behavioral level, possibly due to a higher level compensation mechanism.

In Chapter 3.2, we examined using fMRI whether alexithymia modulated the neural response to emotional prosody when attention was directed at emotional prosody and when attention was not directed at emotional prosody. Alexithymia did again not affect behavioral performance. However, on the neural level alexithymia was associated with a reduced response of the amygdala and the STG to emotional prosody, both when attention was directed to emotional prosody and when attention was diverted from the emotional prosody. Activation of the IFG, however, was not affected by alexithymia. Thus, alexithymia seemed to be associated with a relatively early stage of emotional prosody perception (stage-one and two processing in the STG) but not with a relatively late stage (stage-three processing in the IFG). Further, in keeping with the ERP results presented in Chapter 3.1, while modulation of automatic affective processing by alexithymia is evident at the neural level, it does not translate into effects at the behavioral level.

Together, then, these two studies suggest that (non-clinical) alexithymia primarily modulates relatively early emotional prosody perception stages. In the only other study on modulation of neural processing of emotional prosody by alexithymia, however, both modulation of relatively early and late ERP components was found during attended and unattended emotional prosody perception (Goerlich, Aleman, & Martens, 2012), without translating into effects at the behavioral level. In a recent meta-analysis of the neuroimaging literature on modulation of visual emotional processing by alexithymia (Van der Velde et al., 2013), it was found that alexithymia is associated with a *de*creased response of subcortical structures (such as the amygdala and the insula) but with an *in*creased response of the anterior cingulate cortex (ACC). The authors suggested that the decreased subcortical response might reflect disturbance of early emotional processing in individuals scoring high on alexithymia, while the increase in ACC activation may reflect a compensatory effort to allocate more attention to the emotional stimuli. Although admittedly speculative, this hypothesized primary early emotional deficit in combination with a higher order compensation mechanism may indeed explain the early emotional processing effects observed in this thesis, and why these effects do not translate into behavioral effects (due to higher level compensation). However, it should be noted that a limitation of the present research is that only modulation by normal (non-clinical) variation in alexithymia was studied and hence the conclusions may not generalize to clinical levels of alexithymia.

4.3. Summary of conclusions

We have started this concluding chapter by asking *what* network in the brain supports perception of prosody, *how* it does so, *whether* this network indeed sometimes processes prosody automatically, and if so, *why* this might be the case. Further, we asked whether alexithymia modulates processing in the emotional prosody network, and if so, whether early or late emotional processing is affected (or both).

The series of studies reported in the first empirical section of this thesis suggest that a bilateral temporo-frontal network comprising the HG, m-STG, p-STG and IFG analyzes prosody in at least three processing steps. Within this network, there is relative acoustic specialization of the right HG and p-STG for emotional prosody perception but no hemispheric specialization for linguistic prosody perception. Further, automatic processing of emotional prosody can indeed be demonstrated (but not for emotional music) and is particularly evident for anger prosody, suggesting that such automatic processing may be supported by a hard-wired neural substrate that has evolved to detect (social) threat in order to avoid harm to the organism.

Last, the studies reported in the second empirical section of this thesis confirmed that non-clinical variation in alexithymia modulates activity within the emotional prosody perception network. More specifically, alexithymia seems to primarily modulate early emotional prosody perception stages without translating into behavioral effects, pointing to a potential higher-level neural compensation mechanism.

4.4. Future directions

There are ample new avenues towards advancing our understanding of the cognitive neuroscience of prosody perception and its modulation by alexithymia. For instance, most of the neuroimaging literature to date has focused on emotional prosody perception. Future neuroimaging work could investigate to what extent the network identified in Figure 3 is also involved in various linguistic prosody functions. Regarding emotional prosody perception, it has hardly been directly tested yet whether there are indeed substantial structural and functional connections between the areas in Figure 3, and whether the information flow is indeed in the hypothesized directions. However, initial steps in this direction have been taken (Ethofer et al., 2013). Further, by combining the temporal resolution of ERPs and the spatial resolution of fMRI, it could be investigated whether the areas identified in Figure 3 are indeed active in early versus late stages of the prosody perception process as hypothesized. To gain knowledge regarding the *necessity* of the areas identified in Figure 3 for emotional and linguistic prosody perception, Transcranial Magnetic Stimulation (TMS) could be employed to investigate whether inhibition of activity in these areas does indeed degrade performance as hypothesized. Indeed, initial steps in this direction have been taken as well (Hoekert, Vingerhoets, & Aleman, 2010). Regarding automaticity of emotional prosody perception, future studies could investigate, with more emotional categories in addition to anger and with further reduced levels of attention to emotional prosody, how robust automaticity of emotional perception is and whether it is indeed specific for anger.

Last, concerning modulation within the prosody network by alexithymia, functional and structural connectivity studies may investigate whether alexithymia might be associated with altered structural connections between the ACC and early emotional prosody processing centers, and, using functional connectivity analyses, whether there is altered control from the ACC over early emotional prosody processing centers along these pathways in alexithymia.

Such new directions will undoubtedly further advance our understanding of the cognitive neuroscience of prosody perception and its modulation by alexithymia.

References

Adolphs, R., & Tranel, D. (1999). Intact recognition of emotional prosody following amygdala damage. *Neuropsychologia*, *37,* 1285-1292.

Anderson, A. K., Christoff, K., Panitz, D., De Rosa, E., & Gabrieli, J. D. E. (2003). Neural correlates of the automatic processing of threat facial signals. *The Journal of Neuroscience, 23*, 5627-5633.

Bach, D. R., Hurlemann, R., & Dolan, R. J. (2013). Unimpaired discrimination of fearful prosody after amygdala lesion. *Neuropsychologia, 51,* 2070-2074.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience, 8*, 389-395.

Canli, T. (2004). Functional brain mapping of extraversion and neuroticism: Learning from individual differences in emotion processing. *Journal of Personality, 72*, 1105-1132.

Ethofer, T., Bretscher, J., Bisch, J., Schlipf, S., Wildgruber, D., & Kreifelts, B. (2013). Functional responses and structural connections of cortical areas for processing faces and voices in the superior temporal sulcus. *NeuroImage, 76,* 45-56.

Fazio, R. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, *15*, 115-141.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I. Turner, R., Friederici, A. D., & Koelsch, S. (2009). Universal recognition of three basic emotions in Music. *Current Biology, 19,* 573-576.

Goerlich, K. S., Aleman, A., & Martens, S. (2012). The sound of feelings: Electrophysiological responses to emotional speech in Alexithymia. *PloS One, 7,* e36951.

Hoekert, M., Vingerhoets, H., & Aleman, A. (2010). Results of a pilot study on the involvement of bilateral inferior frontal gyri in emotional prosody perception: An rTMS study. *BMC Neuroscience, 11*, 93

Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31,* 575-621.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132,* 297-326.

Mothes-Lasch, M., Mentzel, H. J., Miltner, W. H. R., & Straube, T. (2011). Visual attention modulates brain activation to angry voices. *The Journal of Neuroscience, 31,* 9594-9598.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*, 483-522.

Pessoa, L., McKenna, M., Gutierrez, E., & Ungerleider, L. G. (2002). Neural processing of emotional faces requires attention. *Proceedings of the National Academy of Sciences, 99,* 11458-11463.

Ross, E. D. (1981). The aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology, 38,* 561-569.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32,* 76-92.

Scheuerecker, J., Frodl, T., Koutsouleris, N., Zetzsche, T., Wiesmann, M., Kleemann, A. M., . . . Meisenzahl, E. M. (2007). Cerebral differences in explicit and implicit emotional processing--an fMRI study. *Neuropsychobiology, 56,* 32-39.

Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10,* 24-30.

Schupp, H. T., Öhman, A., Junghöfer, M., Weike, A. I., Stockburger, J., & Hamm, A. O. (2004). The facilitated processing of threatening faces: An ERP analysis. *Emotion, 4,* 189-200.

Sifneos, P. E. (1973). The prevalence of 'alexithymic' characteristics in psychosomatic patients. *Psychotherapy & Psychosomatics 22*, 255–262.

Van Bezooijen, R., Otto, S. A., & Heenan, T. A. (1983). Recognition of vocal expressions of emotion. *Journal of Cross-Cultural Psychology, 14*, 387-406.

Van der Velde, J., Servaas, M. N., Goerlich, K. S., Bruggeman, R., Horton, P., Costafreda, S. G., & Aleman, A. (2013). Neural correlates of alexithymia: A meta-analysis of emotion processing studies. *Neuroscience and Biobehavioral Reviews, 37*, 1774-1785.

Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Papers in Linguistics: International Linguistics: International Journal of Human Human Communication, 13,* 201-277.

Van Lancker, D., & Sidtis, J. J. (1992). The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. *Journal of Speech and Hearing Research, 35,* 963-970.

Wiethoff, S. W., Wildgruber, D., Grodd, W., & Ethofer, T. (2009). Response and habituation of the amygdala during processing of emotional prosody. *Neuroreport, 20,* 1356-1360.

Wildgruber, D., Ethofer, T., Grandjean, D., & Kreifelts, B. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology, 11,* 277-281.

Samenvatting

Hoe we iets zeggen – met welke toon, luidheid en spreeksnelheid – kan net zo belangrijk zijn als wat we zeggen (de verbale inhoud). Dit proefschrift richt zich op de vraag welk netwerk in de menselijke hersenen dergelijke 'prosodische' aspecten van spraak waarneemt en of de persoonlijkheidstrek *alexithymie* de waarneming van prosodie door dit neurale netwerk beïnvloedt.

Aan de hand van prosodie kunnen twee fundamenteel verschillende soorten informatie gecommuniceerd worden. Aan de ene kant kan prosodie gebruikt worden om emotionele informatie over te brengen – door de manier waarop we iets zeggen kunnen we bijvoorbeeld aangeven of we vrolijk zijn of boos. Aan de andere kant kan prosodie gebruikt worden om de taalkundige structuur van een spraakuiting over te brengen, door bijvoorbeeld aan te geven of een uiting een mededeling of juist een vraag is. Zowel emotionele als linguïstische prosodie maken gebruik van akoestische dimensies als fundamentele frequentie (toonhoogte), intensiteit (luidheid), duur en spectrale samenstelling om betekenis over te brengen. Dit proefschrift richt zich op de vraag welk netwerk in de menselijke hersenen deze twee communicatieve functies van prosodie analyseert en hoe (gebruik makend van welke opeenvolging van neurocognitieve bewerkingen).

Bijna iedereen zal beamen dat het moeilijk is om bijvoorbeeld een ruziënd stel in een restaurant te negeren – zelfs als we deze personen niet kunnen zien en niet kunnen verstaan. Deze observatie duidt erop dat het netwerk in onze hersenen dat emotionele prosodie analyseert, dit blijkbaar ook blijft doen als we niet de intentie hebben de emotionele prosodie te analyseren. In de experimentele psychologie worden dergelijke mentale processen die blijven optreden zelfs als men niet de intentie heeft het proces in te zetten, wel 'automatisch' genoemd. Een andere eigenschap van automatische processen is dat ze snel (efficiënt) zijn. Waarom zou waarneming van emotionele prosodie dan automatisch zijn? Darwin's evolutietheorie biedt een mogelijke verklaring. Het nauwkeurig waar kunnen nemen van de emotionele prosodie (en daarmee de emotionele toestand) van soortgenoten zou in onze evolutionaire geschiedenis voordelig geweest kunnen zijn – het nauwkeurig vaststellen van boosheid van een dominant groepslid zou het verschil betekend kunnen hebben tussen leven en dood (en daarmee het verschil tussen het wel of niet doorgeven van de genen naar de volgende generatie). Omdat waarneming van emotionele prosodie hiermee waarschijnlijk 'fitness value' heeft gehad, zou er in de hersenen door selectieve evolutionaire druk een aangeboren systeem geëvolueerd kunnen zijn dat emotionele prosodie automatisch analyseert. Er is wel voorgesteld dat een dergelijk gespecialiseerd neuraal netwerk met name aannemelijk is voor dreigende prosodie (met andere woorden 'boosheid') omdat het mogelijk schade aan het organisme aangeeft (zodat de vermijding ervan 'fitness value' zou kunnen hebben). In dit proefschrift werd daarom onderzocht of automatische waarneming van emotionele prosodie kan worden aangetoond en of dergelijke automatische waarneming een algemene eigenschap is van emotionele prosodische perceptie of specifiek is voor boosheid.

Als automaticiteit van de waarneming van emotionele prosodie inderdaad verklaard kan worden door het bestaan van een geëvolueerd neuraal systeem dat gespecialiseerd is in de waarneming van emoties van soortgenoten, dan zouden we geen automatische waarneming verwachten van een veel recenter uitgevonden akoestisch signaal waarmee ook emotie kan worden gecommuniceerd – *muziek*. Aan de andere kant is er gesuggereerd dat ook muziek 'fitness value' heeft gehad zodat er mogelijk ook een gespecialiseerd neuraal systeem bestaat voor de waarneming van emotionele muziek. Een alternatieve hypothese die is voorgesteld is dat muziek haar krachtige emotionele vervoeringsvermogen ontleent aan het imiteren en vervolgens overdrijven van emotionele prosodie zodat het mogelijk door het voor emotionele prosodie gespecialiseerde neurale netwerk automatisch waargenomen kan worden. In dit proefschrift werd daarom waarneming van emotionele prosodie en muziek direct vergeleken en onderzocht of automaticiteit van waarneming voor beide akoestische media kan worden aangetoond.

Hoewel het netwerk dat in onze hersenen emotionele prosodie waarneemt in grote lijnen hetzelfde is voor verschillende personen, zijn er ook subtiele individuele verschillen in de waarneming van emotionele informatie. Een persoonlijkheidstrek die geassocieerd is met emotionele waarnemingsstijl is *alexithymie*. Mensen die hoog scoren op deze persoonlijkheidsdimensie hebben moeite met het herkennen, identificeren en het onder woorden brengen van emoties en worden daarom door de omgeving wel als 'koud' of 'afstandelijk' gezien. Er is niet veel bekend over of alexithymie de waarneming van emotionele informatie in de hersenen beïnvloedt (en zo ja, hoe). Het lijkt echter waarschijnlijk dat alexithymie invloed heeft op de verwerking van emotionele prosodie in de hersenen. In dit proefschrift werd daarom tenslotte onderzocht of, en zo ja, in welke fase van het waarnemingsproces alexithymie de waarneming van prosodie in de hersenen beïnvloedt.

In Hoofdstuk 2.1 van dit proefschrift werd onderzocht of er verschillen zijn tussen de twee hersenhelften in het vermogen om emotionele en linguïstische prosodie waar te nemen, ofwel of er 'hemisferische specialisatie' is voor de waarneming van prosodie. Dit kan onderzocht worden door te onderzoeken of schade door een hersenbloeding in de ene hersenhelft de prestatie op prosodische waarnemingstaken meer verstoort dan vergelijkbare schade aan de andere hersenhelft vergeleken met gezonde controlepersonen. Uit een meta-analyse van dergelijke studies bleek dat schade aan beide hersenhelften zowel de waarneming van emotionele- als linguïstische prosodie verstoort. In het geval van emotionele prosodie leidt schade aan de rechter hemisfeer echter tot grotere achteruitgang in taakprestatie dan vergelijkbare schade aan de linker hemisfeer. Uit Hoofdstuk 2.1 bleek dus dat er relatieve rechtshemisferische specialisatie is voor de waarneming van emotionele prosodie, terwijl er voor linguïstische prosodie geen hemisferische specialisatie kan worden aangetoond.

In Hoofdstuk 2.2 werd onderzocht *hoe* hemisferische specialisatie voor de waarneming van prosodie zou kunnen ontstaan. Eén mogelijkheid is dat de rechter hersenhelft beter is in het analyseren van de akoestische eigenschappen van emotionele prosodie dan de linker hersenhelft. Een andere mogelijkheid is dat de rechter hersenhelft beter is in het op meer abstracte wijze analyseren van emotionele informatie. In Hoofdstuk 2.2 werden daarom identieke onzinwoorden aangeboden aan twee groepen proefpersonen. De onzinwoorden hadden klemtoon op de eerste- of de tweede lettergreep en tegelijkertijd treurige of boze prosodie. De ene groep proefpersonen moest de linguïstische prosodie categoriseren, en de andere groep juist de emotionele prosodie. Tijdens het categoriseren van de onzinwoorden werd elektrische activiteit van de hersenen gemeten met elektro-encefalografie. Als alleen de manier waarop de onzinwoorden geanalyseerd worden (linguïstisch versus emotioneel) hemisferische specialisatie bepaalt, dan zouden we voor de groep proefpersonen die de onzinwoorden naar emotie categoriseert grotere activiteit over de rechter dan linker hersenhelft moeten zien, maar voor de groep proefpersonen die dezelfde woorden talig (naar klemtoon) analyseert niet. Er werd echter geen verschil in activiteit tussen de hersenhelften gevonden voor beide groepen. Uit Hoofdstuk 2.2 bleek dus dat rechts hemisferische specialisatie niet goed verklaard kan worden door superioriteit van de rechter hersenhelft in het op abstractere wijze analyseren van emotionele informatie maar waarschijnlijk eerder door specialisatie in de waarneming van de akoestische kenmerken van emotionele prosodie.

In Hoofdstuk 2.3 werd onderzocht welk netwerk in de menselijke hersenen betrokken is bij de waarneming van emotionele prosodie. Eerder onderzoek had gesuggereerd dat de waarneming van emotionele prosodie in drie fases verloopt. In een eerste fase zouden elementaire akoestische eigenschappen van emotionele prosodie verwerkt worden in Heschl's gyrus (HG) en het middelste deel van de superieure temporale gyrus (m-STG). In een tweede fase zouden vervolgens deze akoestische eigenschappen geïntegreerd worden tot meer abstracte representaties. Het was echter niet duidelijk of deze fase plaatsvindt in de anterieure superieure temporale gyrus (a-STG) of de posterieure superieure temporale gyrus (p-STG). In een laatste fase zou emotionele prosodie abstract geëvalueerd worden en geïntegreerd met andere aspecten van de uiting (zoals de semantische betekenis) in de inferieure frontale gyrus (IFG). In Hoofdstuk 2.3 werd een meta-analyse van de functionele MRI (fMRI-) literatuur uitgevoerd om te onderzoeken welk netwerk daadwerkelijk bij de waarneming van emotionele prosodie is betrokken. Het bleek dat HG en m-STG enerzijds en IFG anderzijds inderdaad betrokken waren bij respectievelijk de eerste en laatste fase van het waarnemingsproces. Voorts bleek dat de p-STG en niet de a-STG waarschijnlijk het neurale substraat is van de tweede fase van emotionele prosodiewaarneming. Overeenkomstig de bevindingen van Hoofdstuk 2.1 werd gevonden dat *beide* hersenhelften betrokken zijn bij de waarneming van emotionele prosodie, maar dat er grotere betrokkenheid is van de rechter hersenhelft, wederom duidend op een relatieve rechtshemisferische specialisatie voor emotionele prosodie. Deze grotere rechtshemisferische betrokkenheid werd echter alleen gevonden voor HG en p-STG, wat suggereert dat hemisferische specialisatie voor emotionele prosodie verklaard kan worden door specialisatie van de rechter hemisfeer voor de waarneming van *akoestische* eigenschappen van emotionele prosodie, overeenkomstig de conclusie uit Hoofdstuk 2.2.

In Hoofdstuk 2.4 werd onderzocht of automaticiteit van waarneming kan worden aangetoond voor emotionele prosodie, en of dergelijke automaticiteit specifiek is voor dreigende prosodie (boosheid) of een algemene eigenschap is van emotionele prosodiewaarneming. Hiertoe werden proefpersonen twee taken aangeboden met identiek materiaal terwijl de hersenactiviteit werd gemeten met fMRI. Tijdens de eerste taak moesten proefpersonen categoriseren of de emotionele prosodie van onzinwoorden neutraal, verrast of boos klonk. Voor de tweede taak moesten proefpersonen een andere dimensie van dezelfde onzinwoorden beoordelen. Als we nu in gebieden zowel tijdens de eerste- als de tweede taak activiteit zien, is dit een aanwijzing dat in deze gebieden er automatische verwerking van emotionele prosodie plaatsvindt (want blijkbaar blijven deze gebieden ook emotionele prosodie analyseren als men niet de intentie heeft de emotionele prosodie te analyseren, aangezien men de instructie heeft gekregen om een andere dimensie van dezelfde stimulus te categoriseren). Het bleek dat er inderdaad automatische verwerking van emotionele prosodie aangetoond kon worden in de superieure temporale gyrus (STG), maar alleen voor boosheid. Hiermee zijn de resultaten uit Hoofdstuk 2.4 in overeenstemming met het idee van een geëvolueerd neuraal substraat voor de automatische waarneming van dreigend gevaar uit menselijk stemgeluid (d.w.z. boosheid).

In Hoofdstuk 2.5 werd onderzocht of er ook automatische waarneming van emotionele muziek kan worden aangetoond. Een ander kenmerk van automatische waarneming (dan het optreden van het waarnemingsproces zelfs als men niet de intentie heeft emotionele prosodie te analyseren) is dat automatische waarneming zeer snel plaatsvindt. Het zogenaamde 'affective priming paradigma' kan gebruikt worden om te testen of er inderdaad snelle waarneming plaatsvindt van emotionele prosodie of muziek. Binnen dit paradigma wordt eerst treurige of vrolijke emotionele prosodie of muziek gepresenteerd (de 'affectieve prime'). Vlak (200 milliseconden) daarna wordt een emotioneel woord gepresenteerd dat positief of negatief is. De woorden kunnen dus emotioneel congruent zijn met de affectieve prime (bv. vrolijke muziek en het woord 'GELUK') of emotioneel incongruent (bv. vrolijke muziek en het woord 'PIJN'). Proefpersonen moeten zo snel mogelijk beoordelen of een woord positief of negatief is. Als emotionele prosodie en muziek nu inderdaad snel (in minder dan 200 milliseconden) verwerkt worden, dan zouden deze affectieve primes de snelheid van het categorisatieproces moeten kunnen beïnvloeden, met een langzamere responsie voor incongruente dan congruente paren. Dit wordt het 'affectieve priming effect (APE)' genoemd. In Hoofdstuk 2.5 werd dit paradigma daarom toegepast terwijl de elektrische activiteit van de hersenen werd gemeten. Uit zowel de reactietijden als de metingen van elektrische hersenactiviteit bleek dat er wel automaticiteit van waarneming van emotionele prosodie kon worden aangetoond maar niet voor de waarneming van emotionele muziek. Als proefpersonen echter werden gevraagd een andere (niet affectieve-) dimensie van de woorden te beoordelen, verdwenen ook voor emotionele prosodie de aanwijzingen voor automaticiteit van waarneming. De resultaten van Hoofdstuk 2.5 bevestigen dus de conclusie uit Hoofdstuk 2.4 dat er automatische waarneming is van emotionele prosodie, maar ondersteunen niet de hypothese dat er vergelijkbare automatische waarneming is voor muziek. Hoewel strikt genomen speculatief, komt deze afwezigheid van automatische verwerking van emotionele muziek overeen met het idee dat waarneming van emotionele muziek (in tegenstelling tot emotionele prosodie) mogelijk geen 'fitness value' heeft gehad in de evolutionaire geschiedenis van de mens zodat er geen gespecialiseerd neuraal substraat voor is geëvolueerd. Voorts suggereren de resultaten dat automaticiteit voor de waarneming van emotionele prosodie relatief is – als de aandacht van de emotionele dimensie wordt afgeleid, zien we geen snelle (automatische-) waarneming meer van vrolijke en treurige prosodie.

In het tweede empirische deel van dit proefschrift werd onderzocht of de persoonlijkheidstrek alexithymie de waarneming van emotionele prosodie in de hersenen beïnvloedt, en zo ja hoe, d.w.z. in welke fase van het verwerkingsproces.

In Hoofdstuk 3.1 werd onderzocht of alexithymie de automatische waarneming van emotionele prosodie beïnvloedt zoals gemeten met het affectieve priming paradigma in Hoofdstuk 2.5. Alexithymie bleek niet de gedragsresultaten te beïnvloeden. Op neuraal niveau was er echter een reductie te zien van de elektrofysiologische index van automaticiteit van emotionele prosodiewaarneming door alexithymie. Deze resultaten wijzen er op dat alexithymie relatief vroege (automatische-)

waarneming van emotionele prosodie beïnvloedt zonder dat dit ook is terug te zien op gedragsniveau, mogelijk door strategische compensatie.

In Hoofdstuk 3.2 werd onderzocht of alexithymie de verwerking van emotionele prosodie beïnvloedt zoals onderzocht in Hoofdstuk 2.4. Het bleek dat alexithymie op gedragsniveau opnieuw geen invloed had. Op neuraal niveau bleek alexithymie echter gepaard te gaan met een verminderde responsie van de amygdala en de STG op emotionele prosodie, terwijl activiteit van de IFG niet werd beïnvloed. Deze resultaten duiden er weer op dat alexithymie met name vroege fases van het waarnemingsproces beïnvloedt. Het feit dat alexithymie de gedragsresultaten niet beïnvloedde suggereert dat er vervolgens mogelijk een strategisch compensatie-mechanisme wordt ingezet om voor dit tekort te compenseren.

Concluderend wijzen de resultaten uit dit proefschrift erop dat een bilateraal netwerk, bestaande uit HG, m-STG, p-STG en IFG, (emotionele) prosodie in drie verwerkingsstappen waarneemt. Hierbij is er geen hemisferische specialisatie voor linguïstische prosodie maar wel relatieve rechtshemisferische specialisatie voor de waarneming van emotionele prosodie. Het is aannemelijk dat deze relatieve rechtshemisferische specialisatie voor de waarneming van emotionele prosodie verklaard wordt door specialisatie voor de waarneming van de akoestische eigenschappen van emotionele prosodie in de HG en p-STG. Voorts kan inderdaad automatische waarneming van emotionele prosodie worden aangetoond, met name voor dreigende prosodie (overeenkomstig met fylogenetisch geïnspireerde modellen van automaticiteit van emotionele prosodiewaarneming) maar niet voor emotionele muziek. Automaticiteit van de waarneming van emotionele prosodie voor vrolijke en treurige prosodie blijkt echter relatief te zijn – er zijn geen aanwijzingen voor snelle verwerking van deze categorieën van emotionele prosodie als de aandacht van de emotionele prosodie wordt afgeleid. Tenslotte blijkt alexithymie inderdaad de waarneming van emotionele prosodie in het hierboven beschreven netwerk te beïnvloeden. Hierbij lijken met name vroege (en mogelijk automatische-) fases van het waarnemingsproces beïnvloed te worden zonder dat er ook op gedragsniveau effecten te zien zijn. Dit patroon van resultaten wijst op mogelijke secundaire strategische compensatie voor een primair vroeg emotioneel verwerkingstekort in alexithymie.

Curriculum Vitae

Jurriaan Witteman was born in Haarlem on the 7th of December 1978. He attended the Coornhert Lyceum in Haarlem from 1990-1996. Subsequently he studied medical biology and neuropsychology at the University of Amsterdam where he obtained a Msc in Clinical Neuropsychology in 2005 (Msc, first-class honours*)*. Subsequently, he has been a research assistant at the Psychiatry and Neurosurgery departments of the Utrecht University Medical Center, working on cognitive neuroscience projects. From 2007-2012 he worked on his PhD project "Towards a Cognitive Neuroscience of Prosody Perception and its Modulation by Alexithymia" at Leiden University. As of 2012 he has been working as a project coordinator on projects that are concerned with the influence of alcohol on the brain at the Dutch Institute for Alcohol Policy.