

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/30209> holds various files of this Leiden University dissertation

Author: Witteman, Jurriaan

Title: Towards a cognitive neuroscience of prosody perception and its modulation by alexithymia

Issue Date: 2014-12-18

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 ‘**export**’ (bold indicates a stressed syllable) versus the second syllable ‘**export**’ 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 ‘**neural**’ (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 (*alexia*) 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 microvolts), 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.

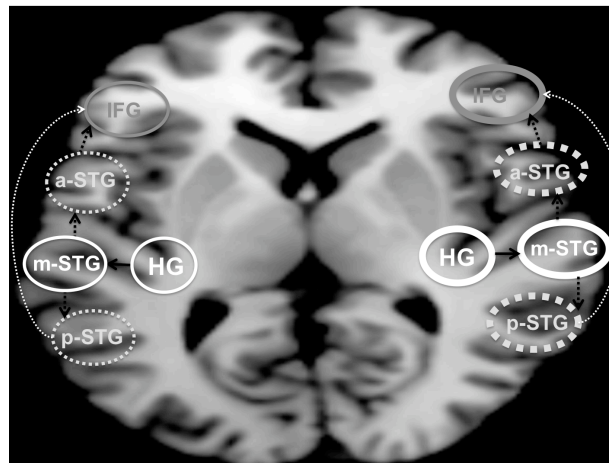


Figure 1. Schematic visualization of the prosody perception pathway as hypothesized by two dominant models of (emotional) prosody perception, superimposed on an axial slice of an MRI scan. Solid circles and arrows indicate regions and connections that are hypothesized to be important in (emotional) prosody 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 (Mériaux 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.
- Banase, 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 H₂¹⁵O-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.
- Mériaux, K., Wartenburger, I., Kazzner, 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. Schmidt, & 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.

