Aus der Universitätsklinik für Psychiatrie und Psychotherapie Tübingen

Abteilung Allgemeine Psychiatrie und Psychotherapie mit Poliklinik

Neurobiology of knowledge and misperception of lyrics by means of a novel method to induce concurring auditory perceptions by using Mondegreens and Soramimi

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Medizin

der Medizinischen Fakultät

der Eberhard Karls Universität

zu Tübingen

vorgelegt von

Beck Lidén, geb. Beck, Claudia Sibylle

2019

Dekan: Professor Dr. I. B. Authenried

1. Berichterstatter: Professor Dr. T. Ethofer

2. Berichterstatter: Professor Dr. M. Wilke

Tag der Disputation: 15.02.2019

Für T., der mein bester Freund und geliebter Weggefährte war und für den ich größten Respekt hege

Table of contents

Table of figures

1 Introduction

1.1 About (mis)perception

Who has not experienced a moment when listening to a song, one stops abruptly and wonders “did he really just sing this”? Maybe it was Michael Jackson and you just heard him singing in German? Maybe your brain tricked you about hearing German? Maybe you even realize it is unlikely that the US popstar is singing in German. You might even be aware that this could be a mere “production” of your brain. The perception that arrived in your consciousness (Jackson singing in an unlikely language) is a product of processing. The reason for this experience is the brain’s attempt to put everything into understandable boxes and its attempt to find similarities or other laws to simplify the sensory impression. As early as 1890 the philosopher Christian v. Ehrenfels wrote his essay “Ueber ‘Gestaltqualitäten’” where he proposed that perception includes qualities which are not only transported by the mere arrangement of sensory impressions. These qualitative perceptions he termed a quality of *Gestalt* (German “*Gestalt*” means shape or form). A melody would be such a quality of *Gestalt*, because the tones could be replaced by other tones – the essential quality is the relation between the tones. He raised the question whether perception – through understanding of a melody – would present more than the individual single tones (v. Ehrenfels, 1890). Ehrenfels was the pioneer of the philosophy of mind of “*Gestalt* psychology”. In 1923, Max

Wertheimer – together with Koffka and Köhler, the founders of the Berlin school of *Gestalt* psychology – formulated essential factors of how we create correlated and meaningful perceptions in constellations (Wertheimer, 1923). He called them *Gestalt* “factors”, but they were later often referred to as *Gestalt* laws, or laws of perceptual organization. The fundamental *Gestalt* law is the law of Prägnanz (German “Prägnanz”: “pithiness”) which states that the different elements of any given array are simply arranged in a stable and coherent form (Sternberg & Sternberg, 2011). Marini and Marzi (2016) explain, “the human brain makes sense of the multitude of sensory stimuli in the environment by grouping them into meaningful representations such as forms or objects”. These laws of perceptual organization describe the functional relations in the emergence of *Gestalt* (Wagemans et al., 2012). Examples include the human tendency to group things close to each other (law of proximity), group similar things (law of similarity) or highlight some things and put others in the background (figure-ground-perception) (Sternberg, 2012). The emerged structures, the “*Gestalten*” (German plural of *Gestalt*), are different than just the sum of their parts, or in Koffka’s words (1935): “the whole is something else than the sum of its parts”. Wagemans et al. (2012) conclude that the *Gestalten* “arise from continuous global processes in the brain, rather than combinations of elementary excitations”. Although the *Gestalt* psychology is often criticized for its lack of “quantitative measurements and precise mathematical models“ (Jäkel et al., 2016) and for being more descriptive than explicative (Friedenberg and Silverman, 2006), more than one century after v. Ehrenfels, Gestalt and its related theories do still inspire research (Wagemans et al., 2012). It is the subjective nature of phenomenal awareness seen in e.g. perceptual switching or perceptive illusions which has been described in *Gestalt* theories (Wagemans et al., 2012). And it is these perceptual illusions that were used in the below described research studies for revealing deeper understanding about the human brain. Today, more tools are available than were to the Gestaltists (Wagemans et al., 2012), including, for example, functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) which offer high spatial resolution, or electroencephalogram (EEG) and magnetoencephalography (MEG), which provide high temporal resolution (Gosseries et al., 2008).

1.2 How perception is altered

Many different models exist for explaining and investigating the process between the actual physical stimulus and what ends up as perceived “reality” in our consciousness. It is well established knowledge that the perception of stimulus is filtered and altered; i.e., it is processed. A rather old example of such an alteration in perception is the placebo effect which was already described by Plato (4th century BC). He stated that words have the ability to heal patients. Nowadays, the placebo effect refers to “any medical treatment surrounded by a psychosocial context that affects the therapeutic outcome“ (Benedetti et al. 2005). With the placebo effect, perception is positively altered in favor of an expectation. Plato referred to the spoken word that would cure, but effects of expectation on perception can be seen for all different kinds of sensory modalities, and are of course not only valid for the Placebo effect.

In the field of speech perception and how perception is altered by expectation, there is the example of the McGurk illusion. According to McGurk and MacDonald (1976), the matching of one sound ("ba") with the visual lip movements of another sound ("ga") gives the perception of a third sound ("da"). This demonstrates that speech perception is not a purely auditory process. Through audiovisual integration something completely new – here the sound "da" – can be perceived. Concerning alteration in vision, Ethofer et al. (2006) demonstrated how speech melody changes our judgment of emotional faces. A fearful or even neutral facial expression is perceived as more fearful when it is accompanied by a fearful voice. Expectations can also alter how intensely we perceive taste. Woods et al. (2011) showed that the expectation of a very sweet drink made it more likely a drink would be reported as sweet compared to the same drink without that sweet expectation. When it comes to the more bitter than sweet expectation of pain, Atlas et al. (2010) showed that expectation has an impact on pain perception. The participants of that study were told that two different tones (high and low) indicated whether they were to receive high or low intensity painful heat stimulation at their forearms. These so-called predictive cues had an impact on how strongly the participants perceived the pain. Heat was perceived as more painful when it was expected to be painful. Melloni et al. (2011) showed that expectation can not only alter perception, but even accelerate it. They measured event-related potentials correlated with when the participants "deciphered" an object (letters, numbers or symbols) as visible and showed that expectations sink the threshold of perception (Melloni et al., 2011). Similarly, Pinto et al. (2015) demonstrated that clear-cut preceding expectations can push access of visual stimuli to consciousness.

Another factor known to alter perception is knowledge. An easy way to show the influence of knowledge on perception is to observe the images below (see figure 1). Dolan et al. (1997) impoverished a picture (object or face), making recognition extremely difficult. However, after acquaintance with the original picture, the viewer can easily decipher the impoverished picture and recognize it. Dolan et al. explored the neural underpinnings of this process called rapid perceptual learning, a process which facilitates perception.

Figure 1: Impoverished images of an object (top row) and a face (bottom row) with their corresponding full grey scale versions to the right, adapted from Dolan et al. (1997).

Different examples on what can alter perception were presented, such as knowledge or expectation. It should be noted that perception can also fail due to expectation. Beck et al. (2016) introduced the example of a bridegroom who misunderstood the song lyric "*Gib mir die Hand, ich geb Dir die Meine*" (give me your hand, and I'll give you mine) at his wedding ceremony. He repeatedly understood "*Gib mir die Hand, ich geb Dir die Beine*" (give me your hand, and I'll give you my legs) and wondered about the meaning of this song within his wedding. "Hand" and "leg" belong to the semantic cluster of body parts and thus, the bridegroom made a (wrong) prediction. His expectation altered his perception – and perception failed him. That humans generate predictions is fundamental to how sensory systems work. This concept is called predictive coding and this framework offers explanations for many perceptual phenomena (Vetter, Sanders and Muckli, 2014). The framework is based on von Helmholtz (1867) and his early ideas about visual perception. Helmholtz stated an involuntary and pre-rational mechanism is part of the formation of visual impression. He referred to that mechanism as unconscious inference (German: *unbewusster Schluss*) as it follows its own laws and wields an imperious power over the human mind. Our eyes see something and our brain comes to an (unconscious) conclusion. As an example, he used the optical illusion of the sun rotating around the world. Although we know better, we see every evening how the sun (seemingly) goes down over the fixed horizon. Over a century later, theoretical and computational studies (Friston, 2005; Rao and Ballard, 1999; Mumford, 1992) precipitated the framework of predictive coding. The brain is referred to as a statistical inference machine which deduces the probable causes of sensory input (Dayan et al., 1994). Our brain deals with the overwhelming amount of sensory information in an efficient way: it makes continuous predictions, based on prior knowledge. The predicted sensory input gets matched with the actual observed input and possible mismatches are identified prediction errors. It is only those prediction errors which get further processed and constantly update further interpretation. Hence, it is only the discrepancy our brain has to take care of, all redundancies are removed (Rao & Ballard, 1999; Sohoglu and Davis, 2016). According to Picard and Friston (2014) expectations are the primary content of our percepts and thus, they are just fantasies made by the brain to interpret our sensations. For this reason, they refer to the brain as a “fantastic” organ.

1.3 Concurring perceptions and processing concepts

A way to further investigate perception is using stimulus material which induces so called concurring perceptions. Well known and largely employed within the field of vision research are reversible figures, like Rubin’s face-vase illusion (Rubin, 1921), the Necker cube or the Schröder staircase. The previously outlined *Gestalt* laws and more precisely the law of figure-ground-perception describe that while looking at a stimulus pattern, some things get emphasized, while others become background. For this reason, some stimulus patterns offer two different ways of interpretation. See figure 2 for an example of Rubin’s face-vase illusion where either two white faces or a black vase can be seen. See figure 3 for another illusion, the so-called duck-rabbit illusion, where either a duck looking to the left or a rabbit looking to the right can be seen (the duck´s beak turns into the rabbit´s ears).

Figure 2: An example of Rubin's face-vase illusion, modified from Kogo et al. (2015).

Figure 3: One of the earliest known duck-rabbit illusions from A. Hengeler published in Fliegende Blätter (1892). Adapted from the digital "Heidelberger historische Bestände" from the university library of Heidelberg.

What is the advantage of using these concurring perceptions in research? Though the physical stimulus itself stays untouched, different percepts reach consciousness. Thus, research can fully concentrate on what is happening in the brain. Despite the long popularity of these reversible figures, there is no consensus of how exactly the multistable perceptual nature of these figures might be explained. Long, Toppino and Mondin (1992) wrote about a "myriad" of empirical effects in the reversible-figure literature. Generally, two basic models are proposed (eg Long et al. 1992; Strüber & Stadler 1999): first, the passive and stimulus driven "bottom-up" manner due to neural fatigue and recovery within cortical structures (eg Köhler 1940; Nawrot & Blake 1989; von Grünau, Wiggin & Reed 1984). Second, active cognitive processes ("top-down") which are supposed to be conceptually driven and reflect analogue learning and problem solving (eg Girgus, Rock & Egatz 1977; Rock, Hall & Davis 1994). Toppino and Long (2005) put it the following way: “the brain finds successively alternative solutions to the puzzle which is created by a reversible figure”. Interestingly, systematic comparison on top-down influences to different categories of reversible figures showed more effective control of the reversability for content related figures (here: duck/rabbit) than perspective related (here: Schröder staircase) (Strüber & Stadler 1999). Besides studies supporting one of these two separate models, there are studies proposing a two-process model which is stimulus driven as much as conceptually driven (Long, Toppino, & Mondin 1992), and even a hybrid theoretical framework in which both types of processes contribute to figure reversal (Toppino & Long 2005). Investigating brain areas involved in bottom-up versus top-down processes, Zekveld et al. (2006) showed distinct roles for frontal and temporal regions. In his study examining impact of noise on speech intelligibilitiy, temporal regions underlie bottom-up processes and frontal areas more subserve top-down mechanisms.

1.4 Mondegreens and Soramimi as auditory concurring perceptions

Concerning speech content and its ambiguous perception, stimuli which can be perceived inaccurately are called “slips of the ear” (Bond, 2005). In the process of decoding speech content “listeners report hearing, as clearly and distinctly as any correctly perceived stretch of speech, something that does not correspond to the speakers' utterance” (Bond, 1999).

Slips of the ear representing misheard song lyrics or other popular collections (Bond, 2005) are known as “Mondegreens”. This term was coined by Sylvia Wright who misunderstood the last line of the Scottish folk ballad “The Bonny Earl O’Morray” (Wright 1945). Wright perceived “They have slain Earl O’Morray/and Lady Mondegreen” instead of “They have slain Earl O’Morray/and laid him on the green”. Mondegreens are constrained to the original language: in the just mentioned example, it was an English native speaker who perceived an English statement out of an English lyric.

Mishearings can also lead to homophonic/near homophonic translations into other languages. Homophonic means “that a word is pronounced like another word but with a different meaning or spelling” (Hornby, 1995). Usually this happens from a foreign language into the native language of the receiver. These cross-language misunderstandings have a long tradition in Japan and are therefore called “Soramimi” or according to Otake (2007) “Soramimi awaa”, which refers to the Japanes word for “mishearing”, literally “empty ear”. In other languages other terms were created for this phenomenon, often based on a prototypical popular example. In German, Soramimi have been called “Agathe Bauer”: a German first and family name whereby “Bauer” means “farmer” and it can be heard instead of “I got the power” in a Snap song. Another popular example and also sometimes name giving to the phenomenon in German is the line “Oma fiel ins Klo” (German for “grandmother fell into the toilet”) instead of “all my feelings grow” in Chris Norman’s “Midnight lady”. Also known is the line “da geht der Gärtner” (German for “there walks the gardener”) instead of Michael Jacksons actually singing “Dirty Diana”. In Dutch, the phenomenon has been named “Mama Appelsap” meaning “Momma Applejuice” which can be heard in a passage of the Micheal Jackson song “Wanna be starting something”.

Another, unique, opportunity for exploring hypotheses about how the human brain encodes language is bilingualism (Spivey & Marian 1999). As per Otake (2007), the basic mechanism of a Soramimi seems to be the same as when a bilingual activates a word: his native L1 lexicon is triggered by non-native L2 auditory. L1 stands for the first language lexicon and L2 for the second language lexicon, as bilinguals are supposed to possess two lexicons instead of a monolingual’s one lexicon (Otake 2007). The term “mental lexicon” is a psycholinguistic concept. According to Marslen-Wilson (1994) a mental lexicon is a listener's mental representation of what words sound like and what they mean, and is at the heart of the spoken language comprehension process. The mental lexicon can be compared to the plan of the London Underground System: it’s an understandable, practical illustration of a network, which is in reality much more complex (Aitchison, 2012).

Interestingly, Mondegreens and Soramimi can be induced through expressly drawing attention to a potentially mispercievable passage of a lyric. For example, this can be done by help of subtitles offering alternative lyrics. These induced slips of the ear are often amusing for the listener, and due to their entertainment value were sometimes featured on radio broadcasts. See further examples on www.youtube.com under the search keyword “misheard lyrics”.

Despite the popularity of Mondegreens and Soramimi – evidenced by their representation in radio shows and on the internet – there exists no scientific research to further investigate this phenomenon, and consequently their neuronal underpinning remains unexplored. This is especially remarkable as study participants report that these alternative perceptions can be persistent over time and can even make it difficult to actually hear the original line of the lyric. This stability makes it a useful tool to induce plasticity in the auditory system. An explanation for this absence of research can be the problem of inducing this phenomenon within the settings of a well-controlled experiment. Normally one encounters a misperception of a song in an unpredictable way. Additionally, there is an unknown variability subject to both stimulus-bound and interindividual factors. Thus, the first step for this work was to establish a method to induce misperceptions based on misheard lyrics.

1.5 Top down and bottom up

The primary strength of using ambiguous perception is that it elegantly enables investigation of processes like prior expectations or memory without changing any physical properties of the stimulus. In perception psychology terms, these are “top down” and “bottom up” approaches. The term of “bottom up” reflects the data driven properties of a stimulus. That could be for example, how loud the stimulus is, or whether words are sung or spoken. “Top down” reflects higher order cognitive processes such as intentions or expectations (Ethofer 2011). As we use Mondegreens and Soramimi as ambiguous material, the stimulus itself is not touched in its properties – all happens in the head of the participant and thus enables us to study top-down modulation of perception.

1.6 Methodology of MRI and fMRI

Magnetic resonance imaging (MRI) is a non-invasive imaging technique used both in research and in clinical medicine. The principles of nuclear magnetic resonance were developed in the 1940s. It is based on the fact that all atomic nuclei act as a magnet and spin around their axes. If an object is put into an external magnetic field, all the atomic nuclei in the object align their spin axes according to the magnetic field. In a next step, a radio frequency (RF) signal is sent at the object perpendicular to the magnetic field. This causes the spin axes to tip away from the magnetic field in a unique angle according to the object and the RF. After its so-called “relaxation time”, the spin axis returns to its parallel alignment with the magnetic field. Relaxation has two components: the transversal and the longitudinal relaxation. During the relaxation process, the nucleus sends out a radio signal which can be measured. The relaxation time differs between different tissues. In brain imaging hydrogen nuclei are used and they have different relaxation times in fat and water. Due to these unique radio transmissions, MRI images of the brain can be generated (Horowitz, 1995).

The method of functional magnetic resonance imaging (fMRI) uses the principle of magnetic resonance for showing activation in the brain. It works with the observation that the activity of neurons uses energy and is in consequence, followed by a local increase in blood flow for refilling the energy depots. The measurement of local changes in blood flow offers a sensitive index of local changes in the activity of neurons (Frith & Dolan, 1997). For measuring blood flow hemoglobin’s oxygenation characteristics are utilized: hemoglobin is diamagnetic when oxygenated, and paramagnetic when deoxygenated (Pauling and Coryell, 1936). In other words, paramagnetic deoxyhemoglobin in venous blood is used as an intrinsic contrast agent in MRI. Ogawa et al. (1990) accentuated this effect and discovered the so called blood oxygenation level-dependent (BOLD) contrast leading to functional MRI methodologies measuring regional neural activity.

1.7 Research questions

1.7.1 Creating a method to induce concurring auditory perceptions

For the purpose of inducing misperceptions based on misheard lyrics, we designed a behavioral study and used short cutouts of songs with the potential for being misperceived. We employed 41 English and 20 German songs which were previously used in radio broadcasts. In this behavioral study we called the attention to passages within the lyrics by using subtitles offering “alternative” lyrics. Our design addressed four aspects of misperceptions in songs: the participants’ familiarity with the lyrics, whether they already knew the slip of the ear beforehand, how they rated the wittiness of a misperception, and the participant’s language competences. We used a bilateral hypothesis for the influence of familiarity, as we had no a priori hypothesis whether familiarity would constitute a protective or susceptible factor for misperceptions. It is often reported that these misperceived lyrics are stable over time, so we hypothesized stronger effects if the participant was already familiar with the alternative lyrics beforehand. Concerning the wittiness of the alternate lyrics, we expected a positive correlation between the strength of the misperception and its rated wittiness: this expectation was based on the fact that emotional arousal strongly influences the encoding of new information (Cahill & McGaugh 1995) and thus, we expected that the joy brought by Mondegreens and Soramimi would enhance the encoding of the alternative lyrics. The last scientific question of our behavioral study was the influence of verbal fluency and linguistic competence. On the one hand, we hypothesized a positive association between high verbal fluency in the native language (German) and a raised occurrence of within-language misperceptions: we expected that the participants would use their wide vocabulary to compensate for ambiguous perceptions which might occur due to, for example, incorrect pronunciation of the singers. In other words, we expected that a high fluency in German would make it easier to misperceive. On the other hand, for across-language misperceptions, we assumed that a high foreign language competence in English would correlate with less across-language misperceptions. Thus, we hypothesized that the unknown foreign words were substituted by known words out of the mental lexicon of the native language. In other words, we expected a high English level to protect the participants from perceiving German words when listening to English songs.

The behavioral study is described in detail in the publication which forms the first part of the results of this thesis. This study also empowered us to do a balanced selection of the stimulus material then used for the next step, a functional magnetic resonance (fMRI) study.

1.7.2 Investigating neurobiology of knowledge and misperception of lyrics

The study consisted of two fMRI experiments and a behavioral experiment. FMRI experiment 1 set out to explore the neural correlates connected to knowledge of lyrics and fMRI experiment 2, the brain areas involved in misperceptions.

In fMRI experiment 1, during scanning, the trial participants had to rate how familiar they are with the lyrics of the song cutouts. We expected that assessing familiarity with lyrics would obviously engage processes as word retrieval and speech comprehension. Thus, we hypothesized activation in left-lateralized regions related to the speech processing network (Hickok and Poeppel, 2000; for a vast review on heard speech and fMRI see Price, 2012). Janata (2009) assumes that music and memories are associated with the medial prefrontal cortex (MPFC). His neuroimaging study using familiar and unfamiliar song excerpts showed activation in foci including the inferior frontal gyrus, the posterior superior temporal gyrus (STG) and the MPFC.

Before the participants went back in the scanner for fMRI experiment 2, they completed two runs (as behavioral experiments), where they listened to the whole set of cutouts subtitled with alternative lyrics. The intent was to train the participants, so that we could later induce misperceptions systematically during the subsequent scanning run. Furthermore, we could collect behavioral data about knowledge and wittiness of the stimuli. In all experiments, nota bene, the auditory input stays identical, as ambiguous stimulus material is used.

We put our expectations for the fMRI experiment 2 in line with the existing neuroimaging literature on ambiguous auditory stimuli. As previously described, the approach with ambiguous material competing for “perceptual dominance” (Andrews 2002) has been mainly used within the visual area. Andrews et al. (2002) employed visual illusions such as Rubin’s face-vase-illusion and suggested that activity in the fusiform gyrus “face area” stands for the perceived rather than barely the retinal stimulus and that the MR activity on a trial-to-trial basis was even statistically predictive of the subjects’ responses (face or vase). Auditory perception is capable of being tricked as well; one illustration is the octave illusion, which uses simple tone stimuli (Deutsch, 1974). Two tones which are an octave apart are played in alternation to both ears simultaneously. While the right ear gets the high tone, the left ear gets the low tone and then vice versa. This causes according to Deutsch (1974) various illusory percepts whereby participants most commonly hear a single tone which alternates by an octave as it alternates between the ears. In combination with neuroimaging, such simple tone stimuli were used by Brancucci et al. (2011). Their results indicate a bilateral involvement of auditory and frontal cortices in association with these illusionary percepts. More complex stimuli material includes the “speech to song transformation” (Deutsch, Henthorn & Lapidis 2011) where a spoken phrase sounds rather like song through repetition. This kind of stimuli material showed differential brain activation in bilateral auditory and inferior frontal cortices during the “song versus speech auditory illusion” (Tierney et al., 2013). However, the study (although using the same phrase as stimuli material) is based on the differences of perception between spoken and sung words and thus can only provide indirect evidence. We tried to explore which areas of the ordinary network for speech perception miscarry during misperceptions triggered by Mondegreens/Soramimi. Furthermore, we wanted to analyse whether these regions are subject to top-down modulation: our ambiguous auditory stimuli material makes us evaluate higher order cognitive processes as expectations without changing the stimulus itself (Beck Lidén et al., 2016). Consequently, we hypothesized that the two different percepts in our study are similarly associated with activity in the named brain areas. Furthermore, as a misperception is connected to a change in the perceived semantic meaning of a lyric, we expected a partial overlap in activity within areas related to the knowledge of lyrics (fMRI experiment 1).

Returning to the example from the beginning, assuming you really just heard Michael Jackson singing in German – would it not be interesting to know more about how this happened? Spanning the descriptions of the subjective nature of phenomenal awareness of the Gestaltists (Wagemans et al., 2012) to modern fMRI, our study, which forms the second part of this thesis, aimed to examine the neuronal underpinnings of the knowledge of lyrics and the network underlying auditory misperceptions.

2 Results

2.1 Mondegreens and Soramimi as a Method to Induce Misperceptions of Speech Content – Influence of Familiarity, Wittiness and Language Competence

The first part of the results is formed by the printout of an article by Beck, Kardatzki and Ethofer which was published in the online journal PloS one in 2014. The following is the detailed bibliographic data of the publication:

Beck, C., Kardatzki, B., & Ethofer, T. (2014). Mondegreens and Soramimi as a method to induce misperceptions of speech content – influence of familiarity, wittiness, and language competence. PloS One, 9(1), e84667. http://doi.org/10.1371/journal.pone.0084667

OPEN ACCESS Freely available online PLOS | ONE

Mondegreens and Soramimi as a Method to Induce Misperceptions of Speech Content – Influence of Familiarity, Wittiness, and Language Competence

Claudia Beck[1], Bernd Kardatzki[1], Thomas Ethofer[1,2]*

1 Department of Biomedical Magnetic Resonance, University of Tübingen, Tübingen, Germany, 2 University Clinic for Psychiatry and Psychotherapy, University of Tübingen, Tübingen, Germany

Abstract

Expectations and prior knowledge can strongly influence our perception. In vision research, such top-down modulation of perceptual processing has been extensively studied using ambiguous stimuli, such as reversible figures. Here, we propose a novel method to address this issue in the auditory modality during speech perception by means of Mondgreens and Soramimi which represent song lyrics with the potential for misperception within one or across two languages, respectively. We demonstrate that such phenomena can be induced by visual presentation of the alternative percept and occur with a sufficient probability to exploit them in neuroscientific experiments. Song familiarity did not influence the occurrence of such altered perception indicating that this tool can be employed irrespective of the participants' knowledge of music. On the other hand, previous knowledge of the alternative percept had a strong impact on the strength of altered perception which is in line with frequent reports that these phenomena can have long-lasting effects. Finally, we demonstrate that the strength of changes in perception correlated with the extent to which they were experienced as amusing as well as the vocabulary of the participants as source of potential interpretations. These findings suggest that such perceptional phenomena might be linked to the pleasant experience of resolving ambiguity which is in line with the long-existing theory of Hermann von Helmholtz that perception and problem-solving recruit similar processes.

Citation: Beck C, Kardatzki B, Ethofer T (2014) Mondegreens and Soramimi as a Method to Induce Misperceptions of Speech Content – Influence of Familiarity, Wittiness, and Language Competence. PLoS ONE 9(1): e84667. doi:10.1371/journal.pone.0084667

Editor: Kevin Paterson, University of Leicester, United Kingdom

Received September 5, 2013; **Accepted** November 18, 2013; **Published** January 8, 2014

Funding: The study was supported by the Open Access Publishing Fund of the University of Tübingen. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: thomas.ethofer@med.uni-tuebingen.de

Introduction

Our perception of the environment and ourselves is strongly shaped by our expectations. This is reflected in sayings such as "I did not believe my eyes" or in well-known psychological phenomena, such as the placebo effect [1,2] and the McGurk illusion [3]. Expectations can both result in acceleration [4] as well as alteration of perception across sensory modalities. For vision, such perceptual changes have been described for judgments of emotional faces if they are preceded by affective signals expressed by speech melody [5]. Similarly, expectations can modify the perceived intensity of gustatory [6] or painful perceptions [7]. An elegant method to examine top-down modulation of perception is to employ stimuli which can induce two concurring percepts which has been extensively employed in vision research using reversible figures, such as Rubin's vase-face illusion [8]. So far, much less is known about how expectations can alter perception in the auditory modality. For speech content, such ambiguous stimuli that can be perceived in two different ways are called 'slips of the ear' [9] which have been described as phenomena during which 'a listener reports hearing, as clearly and distinctly as any correctly perceived stretch of speech, something that does not correspond to the speaker's actual utterance' [10]. We will focus on slips of the ear as a means to study top-down modulation of perception.

We introduce a novel approach to induce such altered perception of verbal messages by means of misheard song lyrics. This phenomenon was named "Mondegreen" in reference to its first description by Sylvia Wright for the Scottish folk song "The Bonny Earl O'Morray" [11] in which the author understood in its last line "They have slain Earl O'Morray/and Lady Mondegreen" instead of "They have slain the Earl O'Morray/and laid him on the green". Interestingly, this phenomenon can be induced by explicitly calling the attention on passages of lyrics with the potential for misperception. In many cases, such induced "Mondegreens" are amusing for the listener and thus songs with possibly misheard passages were broadcasted by several radio stations in the recent past (for examples, please see www.youtube.com, search item: misheard lyrics).

In addition to Mondegreens which are restricted to the original language, mishearing of lyrics can also result in homophonic/near-homophonic translations into another language (typically into the native language of the listener). This phenomenon has its longest tradition in Japanese and is thus called "Soramimi" (which means mishearing in Japanese). However, Soramimi became recently also popular in other cultures and have been named after prototypical homophonic translations within their respective language: In German, Soramimi are called "Agathe Bauer" a combination of a German first and family name with "Bauer" meaning "farmer" misheard from the song of Snap "I got the power". In Dutch, this phenomenon became popular under the name "Mama Appelsap" (meaning momma apple juice) misheard from a passage of Michael Jackson's song "Wanna be starting something".

In spite of the popularity of Mondegreens and Soramimi with regular radio and television broadcasting devoted to this topic, no scientific studies have been conducted to examine this phenomenon. This is even more surprising as people often report that the altered perception can be quite persistent occurring every time the song is heard, making Mondegreens and Soramimi a valuable tool to induce plasticity within the auditory system. It is important to acknowledge, however, that the potential of induction of both Mondegreen and Soramimi is variable across the employed songs as well as listeners. Therefore, the first step towards establishing these phenomena as method for induced misperceptions is to investigate the variability depending on stimulus-bound as well as interindividual influence factors for the occurrence of within-language (Mondegreen) and across-language (Soramimi) misperceptions. In the current study, we focused on the familiarity of the participants with the employed song texts as well as prior knowledge and rated wittiness of the misperceived lyrics. We had no a priori hypothesis whether familiarity with the original song texts make the participants more persistent against or susceptible for induced misperceptions and thus tested possible influences of this factor using a bidirectional hypothesis. As it is often reported that such altered perception is stable across time, we expected stronger effects if the alternative lyrics were already known to the study participants. Emotion has been repeatedly demonstrated to have a strong impact on encoding of novel information (e.g., [12,13]). Therefore, we hypothesized that the joy typically induced by Mondegreens/Soramimi would result in enhanced encoding of alternative lyrics and thus a positive relationship between the strength on induced misperceptions and their wittiness was expected. Finally, we investigated the impact of verbal fluency and linguistic competence on occurrence of such phenomena. We predicted that high verbal fluency within the native language of the participants is associated with increased occurrence of within-language misperceptions as listeners with high verbal fluency might use their broad vocabulary to compensate ambiguous perception caused by inaccuracies in the pronunciation of the singer. In contrast, competence for the language of the songs with potential Soramimi (English) was expected to be a protective factor against across-language misperceptions into the native language of the participants (German) as we hypothesized that such misperceptions are partly driven by the fact that unknown foreign words are automatically replaced by known words of the native language of the listener.

Materials and Methods

Ethics Statement

The study was approved by the Ethical Committee of the University of Tuebingen (votum: 215/2012 BO2) and written informed consent was obtained from all participants. All study procedures were in line with the latest version of the Declaration of Helsinki.

Participants

23 healthy German native speakers (12 women, 11 men, mean age: 28.3±6.8 years, education: 16.4±2.8 years) participated in this study. All participants were right-handed according to the Edinburgh Handedness Inventory [14]. Mean verbal intelligence of the participants as obtained by a German vocabulary test (Mehrfach-Wortschatz-Intelligenztest B) was 112.9±11.5. Knowledge of English language was assessed using a language test based on a short form of the Test of English as a Foreign Language (Mini-TOEFL, http://www.ets.org/toefl) to determine the participants' grammar and vocabulary comprehension. On average, the participants correctly answered 15.4±5.1 out of 25 questions. Verbal intelligence and English language comprehension were significantly correlated ($r = .67$, two-tailed $p < .01$) reflecting the fact that they are similarly driven by the educational level of the participants. All values are given in mean ± standard deviation.

Stimuli

The stimulus set comprised short audio clips (mean duration: 17.7±3.0 sec) taken from 41 English and 20 German songs which were broadcasted by radio stations because of their potential to induce within- or across language misperceptions. All stimuli were normalized to same peak intensity.

Experimental Design

All stimuli were presented during three consecutive experimental runs in a fully randomized order. During experimental run 1, the participants judged their degree of familiarity with the songs on a four-point scale ('unknown', 'know melody', 'know refrain', 'know text') that was visually presented for five seconds after stimulus offset. After this judgment period, a fixation cross was shown for two seconds. Before starting experimental run 2, the participants were informed that the song texts can be misheard and that such alternative lyrics will be presented visually during the next experimental run. The task of the participants was to determine whether these alternative lyrics were already known to them (because they heard them in radio transmissions or even spontaneously misheard the lyrics in the presented manner) during a four second period which was followed by a presentation of a fixation cross for two seconds. In experimental run 3, participants were instructed to judge whether and how strongly they misheard the lyrics as presented visually during the second run on a four point scale ('not at all', 'slightly', 'moderate', 'strongly'). It was emphasized during instruction of the volunteers that the task is not to indicate whether they remember the alternative lyrics, but to indicate to what extent an alteration of perception occurs. Directly after judging the strength of possible misperceptions, the participants rated the wittiness of these misperceptions on a four point scale ('not at all', 'slightly', 'moderate', 'strongly'). For both ratings in this third experimental run, participants conveyed their decision during a four second interval and a fixation was shown for two seconds before the next stimulus was presented.

Data Analysis and Hypotheses

All values are given in mean ± standard error of the mean (SEM) unless otherwise specified. For correlation coefficients, we employed back-transformed mean and SEM values of individual Fisher Z scores.

We first determined the frequency and the variability (i.e., range and standard deviation, sd) of misperceptions across the employed stimuli as well as the study participants separately for within- and between-language misperceptions. We then tested whether familiarity with the employed songs as well as prior knowledge and wittiness of the alternative lyrics influenced the strength of induced misperceptions. In addition, we determined potential influences of interindividual differences in verbal intelligence and knowledge of English on these misperceptions.

Effect of familiarity (experimental run 1). We calculated individual correlation coefficients between song familiarity and strength of misperceptions separately for each subject. These individual correlation coefficients were then transformed to Fisher

Z scores. As we had no a priori assumption on the direction of the effect of familiarity (bidirectional hypothesis), we subsequently submitted these values to a two-tailed one-sample t-test.

Effect of prior knowledge of alternative lyrics (experimental run 2). We hypothesized that prior knowledge (i.e., previous encounters with the alternative lyrics in the media) increases the frequency of misperceptions (unidirectional hypothesis). To test this hypothesis, we statistically compared the frequency for misperceptions of previously known versus unknown alternative song texts using one-tailed paired t-tests. In this analysis, only participants (17 for within- and 20 for across-language trials) who knew at least one alternative song text of the respective types of misperceptions could be included (otherwise the frequency in previously known misperceptions cannot be defined due to a division by zero in these subjects).

In addition, we determined whether subjects who already know many alternative song texts are generally prone to such misperceptions. To this end, we investigated whether the probability to misperceive songs in which the alternative lyrics were not known to the subject increases with the amount of previously known alternative song texts by means of a correlation analysis (unidirectional hypothesis).

Effect of wittiness of alternative lyrics (experimental run 3). We calculated individual correlation coefficients between rated wittiness and strength of misperceptions separately for each subject. As the subjects judged the wittiness of perceived misperceptions, only those trials during which the subjects reported to have misperceived the lyrics were included in this analysis (8.5 ± 0.8 within-language trials, 13.7 ± 0.1 across-language trials). Furthermore, as correlation analyses require data that are distributed at least across two factor levels for both investigated variables, data of one participant had to be excluded for the across-language misperceptions and data of five participants had to be excluded from the within-language misperceptions for this analysis.

The individual correlation coefficients were then transformed to Fisher Z scores. As we hypothesized a positive relationship between these factors (unidirectional hypothesis), we subsequently submitted these values to a one-tailed one-sample t-test.

Effect of verbal intelligence and knowledge of English. We additionally evaluated whether verbal intelligence as well as knowledge of English influenced misperceptions. To this end, mean values of ratings for misperceptions were calculated separately for within- and across-language trials for each subject and correlation analyses of these values with verbal intelligence scores as obtained by the MWTB as well as correct responses in the Mini-TOEFL-test were conducted. As we predicted that high verbal fluency increases the occurrence of within-language misperceptions while linguistic competence in English protects against across-language misperceptions, we employed one-tailed t-tests to address these directional hypotheses.

Results

All data are freely available and can be downloaded as supplemental material (Data S1). Frequencies of normal song perception and misperceptions (slight, moderate, and strong) as well as missed responses are presented in Table 1 separately for within- and across-language trials. Subjects responded in more than 99% of the trials indicating that the four second period was sufficient to judge the strength of misperceptions. Within-language misperceptions occurred with a significantly higher frequency than across-language misperceptions (42.6% ±3.8% versus 33.5% ±3.8%, $t(22) = 2.62$, two-tailed $p < .05$). For both types of misperceptions, a large variability was found across the employed stimulus material ranging from 8.7% to 87.0% (sd: 20.3%) for within- and 4.4% –82.6% (sd: 21.1%) for between-language misperceptions. Similarly, these frequencies also varied considerably across participants ranging from 15.0% to 90% (sd: 18.3%) for within- and 9.8% to 85.4% (sd: 18.1%) for across-language misperceptions.

We then tested whether the familiarity of the participants with the stimulus material (experimental run 1), prior knowledge of the alternative lyrics (experimental run 2) and wittiness of these altered song texts (experimental run 3) influenced the strength of misperceptions.

For familiarity with the song texts as obtained in experimental run 1, no significant correlation was found for either within- ($r = .08 \pm .04$, $t(22) = 2.00$, two-tailed $p = .06$) or across-language misperceptions ($r = -.01 \pm .03$, $t(22) = -0.20$, two-tailed $p = .85$).

Prior knowledge of the alternative song texts as assessed in experimental run 2, however, was strongly predictive for both within- (72.1% ±9.6% versus 40.5% ±3.7%, $t(16) = 2.78$, one-tailed $p < .05$ for known versus unknown alternative lyrics, respectively) and across-language misperceptions (74.0% ±6.2% versus 30.0±4.0%, $t(19) = 7.08$, one-tailed $p < .001$ for known versus unknown alternative lyrics, respectively) in experimental run 3. However, prior knowledge cannot explain the generally higher frequency of within- than across-language misperceptions as the participants were familiar with more alternative song texts for across- than within-language trials (13.5% ±2.8% versus 6.7±1.2%, $t(22) = 2.40$, two-tailed $p < .05$). Furthermore, previous knowledge of alternative lyrics did not generally increase the probability of misperceptions as no significant correlation was found between the number of known alternative song texts and the probability to misperceive songs for which the alternative lyrics were not known to the subject ($r = .03$, one-tailed $p = .44$ for within- and $r = .07$, one-tailed $p = .38$ for across-language misperceptions).

We also asked whether the strength of misperceptions was influenced by the extent the participants perceived them as witty. This was indeed the case as the mean within-subject correlation coefficient for ratings of the strength of misperceptions and judgments of their wittiness was $r = .55 \pm .13$ ($t(17) = 4.23$, one-tailed $p < .001$) for within- and $r = .58 \pm .09$ ($T(21) = 6.31$, one-tailed $p < .001$) for across-language misperceptions.

Finally, we investigated possible influences of verbal intelligence and comprehension of English language on misperceptions. In agreement with our a priori hypothesis, the mean rating of misperceptions in within-language trials calculated separately for each subject correlated significantly with verbal intelligence ($r = .46$, one-tailed $p < .05$) of the participants. For across-language

Table 1. Percentage of within and across language misperceptions.

Strength of misperception	Within-language	Across-language
None	56.7%±3.8%	66.0%±3.8%
Slight	17.0%±2.7%	11.0%±1.5%
Moderate	10.7%±2.0%	11.5%±1.7%
Strong	15.0%±2.8%	11.0%±2.2%
Missed response	0.7%+0.4%	0.5%+0.3%

All values are given in mean + standard error.
doi:10.1371/journal.pone.0084667.t001

trials, however, the effect was opposite to our hypothesis as language competence as assessed with the Mini-TOEFL was also positively correlated with mean ratings for across-language trials, but failed to reach significance ($r = .28$, one-tailed $p = .10$).

Discussion

Induced alterations of perception as an experimental tool have a long-standing tradition in research within the visual domain (e.g. reversible figures, such as the Rubin's vase-face illusion [8]) and have been successfully employed to study cerebral processing of perceptual phenomena [15,16]. The primary strength of paradigms relying on such induced changes in perception is that they elegantly enable investigation of top-down influences (e.g., prior expectations, memory) without changing bottom-up parameters (i.e., physical parameters of the stimulus material). A principal advantage of examining altered perception in the auditory modality is that acoustic information is expressed in the temporal domain which offers the opportunity to exactly pinpoint the onset of events with altered perception. Thus, studying such phenomena in the auditory domain enables to overcome methodological obstacles typically encountered during visual presentation of reversible figures that induce spontaneous as well as unpredictable switches between two competing percepts without sufficient temporal stability necessitating modification of the stimuli, such addition of embossing [15].

In the current study, we propose a novel method for targeted induction of auditory misperceptions by means of Mondegreens/Soramimi. To this end, we visually presented the alternative lyrics during listening to the respective song parts and asked the participants to rate the strength of misperceptions in a succeeding run without additional visual cue. Altered perception occurred in 42.6% of the within- and 33.5% of the across-language trials. These results indicate that careful stimulus selection based on the current evaluation experiments would allow obtaining a set of stimuli for which altered perception can be expected in about half of the trials. A probability of 50% would be optimal for application of these stimuli in a classical 2×2 factorial design with induction (before versus after induced misperceptions) and perception (original versus alternative lyrics) as within-subject factors during neuroimaging studies.

We also evaluated potential stimulus-dependent and interindividual influence factors that might predict the occurrence of such phenomena. Familiarity with the stimulus material has been shown to have a profound impact on neural processing of musical stimuli [17]. In our study, the correlation between the familiarity of listeners with the presented musical stimuli and the occurrence of altered perception was close to zero indicating that these phenomena can be studied without potential biases of song familiarity. On the other hand, previous knowledge of the alternative lyrics strongly influenced whether or not misperceptions occurred which is in line with the reports of many of our study participants that induced misperceptions can result in long-lasting effects that occur each time the respective song is heard. Thus, it might be advisable to obtain data on familiarity with the alternative lyrics to be able to exclude such trials in neuroscientific experiments relying on a design comparing responses before versus after induced misperceptions.

A large body of evidence indicates greater allocation of attentional resources [18], enhancement of perceptual vividness [19], as well as better memory [20] to stimuli of affective value and similar effects have also been demonstrated for musical stimuli [21–23]. Therefore, we predicted that humor appreciation modulates the occurrence of induced misperceptions. Indeed, the rated wittiness of the alternative lyrics correlated significantly with the intensity of misperceptions explaining about 30% of the variance. This finding is in line with observations made in the visual domain where resolving ambiguities of visual percepts has been found to generate pleasant feelings similar to the experience of problem solving [24]. These findings underline the importance of affective modulation in perceptual processes including the induced misperceptions examined here and suggest that future studies of such phenomena should encompass ratings of wittiness to study the neural correlates underlying modulation of such perceptual phenomena.

In concordance with our a priori hypothesis, a positive relationship between verbal fluency in German on the one hand and the strength of induced within-language misperceptions on the other hand was found. This finding indicates that the more exhaustive vocabulary of participants with high verbal fluency can be automatically used to generate alternative solutions in ambiguous perceptual situations. Regarding across-language misperceptions, however, we did not find the expected negative, but rather a positive relationship for linguistic competence as assessed by the Mini-TOEFL. This finding has to be interpreted with caution as it failed to reach significance, but at least argues clearly against the assumption that such mishearings are simply driven by inadequate language knowledge of the listeners [25].

In summary, we propose a novel method for investigating top-down effects on processing of language stimuli by means of induced Mondegreens/Soramimi as experimental tool. We demonstrated that such phenomena occur with a sufficient frequency for application in neuroscientific experiments that aim to study processing of ambiguous stimuli and how such processing is modulated by expectations. Interestingly, the occurrence of induced misperceptions was independent of knowledge of the original, but not of the alternative percept which is in line with observations made in the visual domain for reversible figures demonstrating that we can get stuck in one interpretation until we are informed that there is an alternative interpretation. Finally, wittiness of the misheard lyrics as stimulus-dependent and verbal fluency as interindividual factor increased the strength of induced misperceptions indicating that this phenomenon depends on whether the alternative percept was experienced as pleasant as well as on the vocabulary of the participants as source of possible interpretations. These findings suggest that these phenomena are linked to the joyful feeling of resolving ambiguity and concur with the long-existing theories formulated by Hermann von Helmholtz who hypothesized that perception relies on similar processes as intellectual problem solving [26].

Supporting Information

Data S1 The supplemental data file includes information on analog data of the study participants as well as data analyses according to the subjects and stimuli (familiarity with the song text, familiarity with misperceptions, strength of misperceptions, wittiness uncorrected and corrected).
(XLSX)

Acknowledgments

The authors thank Alexys Lang for proofreading of the manuscript.

Author Contributions

Conceived and designed the experiments: CB TE. Performed the experiments: CB. Analyzed the data: CB BK. Contributed reagents/materials/analysis tools: BK. Wrote the paper: BC TE.

References

1. Plato (4th century BC) Charmides. 155C-156A p.
2. Dean ME (2006) 'An innocent deception': placebo controls in the St Petersburg homeopathy trial, 1829–1830. J R Soc Med 99: 375–376.
3. McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264: 746–748.
4. Melloni L, Schwiedrzik CM, Muller N, Rodriguez E, Singer W (2011) Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. J Neurosci 31: 1386–1396.
5. Ethofer T, Anders S, Erb M, Droll C, Royen L, et al. (2006) Impact of voice on emotional judgment of faces: an event-related fMRI study. Hum Brain Mapp 27: 707–714.
6. Woods AT, Lloyd DM, Kuenzel J, Poliakoff E, Dijksterhuis GB, et al. (2011) Expected taste intensity affects response to sweet drinks in primary taste cortex. Neuroreport 22: 365–369.
7. Atlas LY, Bolger N, Lindquist MA, Wager TD (2011) Brain mediators of predictive cue effects on perceived pain. J Neurosci 30: 12964–12977.
8. Rubin E (1915) Visuell Wahrgenommene Figuren. Copenhagen: Gyldenalske Boghandel.
9. Bond ZS (1999) Morphological errors in casual conversation. Brain Lang 68: 144–150.
10. Bond ZS (1999) Slips of the ear: Errors in the perception of casual conversation. New York: Academic Press.
11. Wright S (1954) The Death of Lady Mondegreen. Harper's Magazine 209: 48–51.
12. Cahill L, McGaugh JL (1995) A novel demonstration of enhanced memory associated with emotional arousal. Conscious Cogn 4: 410–421.
13. Kensinger EA (2004) Remembering emotional experiences: the contribution of valence and arousal. Rev Neurosci 15: 241–251.
14. Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9: 97–113.
15. Andrews TJ, Schluppeck D, Homfray D, Matthews P, Blakemore C (2002) Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion. Neuroimage 17: 890–901.
16. Hasson U, Hendler T, Ben Bashat D, Malach R (2001) Vase or face? A neural correlate of shape-selective grouping processes in the human brain. J Cogn Neurosci 13: 744–753.
17. Pereira CS, Teixeira J, Figueiredo P, Xavier J, Castro SL, et al. (2011) Music and emotions in the brain: familiarity matters. PLoS One 6: e27241.
18. Anderson AK (2005) Affective influences on the attentional dynamics supporting awareness. J Exp Psychol Gen 134: 258–281.
19. Cahill L, Anderson AK (2009) Emotional learning in humans; Squire L, editor. London: Elsevier.
20. Cahill L, Prins B, Weber M, McGaugh JL (1994) Beta-adrenergic activation and memory for emotional events. Nature 371: 702–704.
21. Platel H, Baron JC, Desgranges B, Bernard F, Eustache F (2003) Semantic and episodic memory of music are subserved by distinct neural networks. Neuroimage 20: 244–256.
22. Eschrich S, Munte TF, Altenmuller EO (2008) Unforgettable film music: the role of emotion in episodic long-term memory for music. BMC Neurosci 9: 48.
23. Stalinski SM, Schellenberg EG (2013) Listeners remember music they like. J Exp Psychol Learn Mem Cogn 39: 700–716.
24. Ramachandran VS, Rogers-Ramachandran D (2007) Ambiguities and Perception. Scientific American Mind 18: 18–20.
25. Meyer S, Ptok M (2011) [A phenomenon called Mondegreen]. HNO 59: 926–930.
26. Helmholtz v (1866) Concerning perceptions in general.

2.2 Neurobiology of knowledge and misperception of lyrics

The second part of the results is formed by the printout of an article by Beck Lidén, Krüger, Schwarz, Erb, Kardatzki, Scheffler and Ethofer which was published in the journal NeuroImage in 2016. The following is the detailed bibliographic data of the publication:

> *Beck Lidén, C., Krüger, O., Schwarz, L., Erb, M., Kardatzki, B., Scheffler, K., & Ethofer, T. (2016). Neurobiology of knowledge and misperception of lyrics. NeuroImage, 134, 12–21.*

NeuroImage 134 (2016) 12–21

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Neurobiology of knowledge and misperception of lyrics

Claudia Beck Lidén [a], Oliver Krüger [a], Lena Schwarz [a,b], Michael Erb [a], Bernd Kardatzki [a], Klaus Scheffler [a,c], Thomas Ethofer [a,b,c,*]

[a] *Department of Biomedical Magnetic Resonance, University of Tübingen, Otfried-Müller-Str. 51, 72076 Tübingen, Germany*
[b] *University Clinic for Psychiatry and Psychotherapy, University of Tübingen, Calwer Str. 14, 72076 Tübingen, Germany*
[c] *Max-Planck-Institute for Biological Cybernetics, Speemannstraße 38-40, 72076 Tübingen, Germany*

ARTICLE INFO

Article history:
Received 16 November 2015
Revised 30 March 2016
Accepted 31 March 2016
Available online 13 April 2016

Keywords:
Connectivity
DWI
fMRI
Lyrics
Misperception
Music

ABSTRACT

We conducted two functional magnetic resonance imaging (fMRI) experiments to investigate the neural underpinnings of knowledge and misperception of lyrics. In fMRI experiment 1, a linear relationship between familiarity with lyrics and activation was found in left-hemispheric speech-related as well as bilateral striatal areas which is in line with previous research on generation of lyrics. In fMRI experiment 2, we employed so called Mondegreens and Soramimi to induce misperceptions of lyrics revealing a bilateral network including middle temporal and inferior frontal areas as well as anterior cingulate cortex (ACC) and mediodorsal thalamus. ACC activation also correlated with the extent to which misperceptions were judged as amusing corroborating previous neuroimaging results on the role of this area in mediating the pleasant experience of chills during music perception. Finally, we examined the areas engaged during misperception of lyrics using diffusion-weighted imaging (DWI) to determine their structural connectivity. These combined fMRI/DWI results could serve as a neurobiological model for future studies on other types of misunderstanding which are events with potentially strong impact on our social life.

Introduction

"Gib mir die Hand, ich geb Dir die Meine" (give me your hand, I give you mine) was the refrain of a church hymn played during a wedding ceremony, but the bridegroom repeatedly understood "Gib mir die Hand, ich geb Dir die Beine" (give me your hand, I give you the legs) and was wondering about the meaning of these words in the context of his marriage. This is a typical example on how perception can fail due to expectations (in this case generated by the fact that hand and leg belong both to the semantic word field of body parts). Generation of predictions is a fundamental mechanism of sensory systems to maximize the conveyed information while minimizing the required resources which has been defined within the concept of predictive coding (Huang and Rao, 2011). Removal of predictable components by learning statistical regularities is a highly efficient way of coding sensory input, but can also result in perception which does not accurately reflect the physical stimulus. While in this case the misperception was clearly harmless and its enlightenment a reason for joy, misunderstandings caused by such extrapolations of predictive coding mechanisms can have a severe impact on one's social life. It is difficult, however, to examine such phenomena within the framework of a neuroimaging experiment as this requires both to control induction of wrong expectations and the onset of their effects on speech perception. To overcome this methodological obstacle we recently established a method to induce misperceptions on the basis of misheard lyrics (Beck et al., 2014). Lyrics represent speech stimuli which are typically perceived under adverse conditions due to often atypical pronunciation and background noise (instrumental music) and such suboptimal conditions decrease intelligibility (Mattys et al., 2012). Misperceptions within the language of the original lyrics are called "Mondegreen" in reference to their first description by Sylvia Wright for the Scottish folk song "The Bonny Earl O'Morray" (Wright, 1954) in which the author understood in its last line "They have slain Earl O'Morray/and Lady Mondegreen" instead of "They have slain the Earl O'Morray/and laid him on the green". Here, the mishearing is based on wrong segmentation of continuous speech into its word components, for a review of this phenomenon see Cutler and Butterfield, 1992). Non-native stimuli are particularly prone to perceptual illusions including 'nativization processes' (Calabrese, 2012). If mishearing of lyrics results in homophonic/near homophonic translations across languages (typically into the native language of the listener), they are called "Soramimi" (which means mishearing in Japanese where the observation of this phenomenon has its longest tradition). It has been shown that concurrent presentation of written text can serve as a top–down modulation enhancing the clarity of speech stimuli (Sohoglu et al., 2014). In the present study, we make use of this phenomenon to induce Mondegreens/Soramimi which enables us to experimentally control both the expectation as well as the exact onset of the resulting misperception.

* Corresponding author at: Department of Biomedical Magnetic Resonance, University of Tübingen, Otfried-Müller-Str. 51, 72076 Tübingen, Germany.

http://dx.doi.org/10.1016/j.neuroimage.2016.03.080
1053-8119/© 2016 Published by Elsevier Inc.

We conducted two functional magnetic resonance imaging (fMRI) experiments. In fMRI experiment 1, we aimed to clarify which neural structures underlie knowledge of lyrics. We measured knowledge of lyrics by asking the participants to rate their familiarity with the lyrics of song excerpts on a 4-point scale presented during scanning. A previous neuroimaging study (Janata, 2009) comparing familiar with unfamiliar song excerpts revealed activations in the left inferior frontal cortex and posterior superior temporal gyrus which have been associated with speech-related processes such as word retrieval and speech comprehension (for a review, see Price, 2012). On the other hand, adaptation effects due to repeated exposure of lyrics resulted in bilateral repetition suppression effects in the superior temporal cortex (Sammler et al., 2010).

Before fMRI experiment 2, we informed the study participants about ambiguous parts in these song excerpts with the intention to systematically induce misperceptions (i.e. perception of alternative lyrics not intended by singer, but made possible by the ambiguity of the acoustic information). Regarding the neural correlates of such altered speech perception, we based our hypothesis on the available neuroimaging literature on bistable auditory stimuli. This approach induces two different percepts in spite of physical stimulus constancy and has been repeatedly employed using visual stimuli (e.g. Rubin's face-vase illusion, Andrews et al., 2002; Hasson et al., 2001). In the auditory domain, first results based on simple tone stimuli to induce the octave illusion (Deutsch, 1974) point to a bilateral involvement of auditory and frontal cortices associated with the two concurring percepts (Brancucci et al., 2011). So far, little is known regarding the representation of such competing perceptions in more complex stimuli, such as lyrics. It should be noted, however, that a network specifically engaged to induce misperceptions in speech is not biologically plausible as it would offer no advantage from an evolutionary perceptive. Thus, the aim of this study is not to propose a 'misperception network', but rather clarify which parts of the regular speech perception network fail during misperceptions induced by Mondegreens/Soramimi. A dual-stream speech perception model has been proposed with a predominantly left-lateralized dorsal stream for auditory motor-integration and a bilaterally organized ventral stream for retrieval of meaning from sounds (Hickok, 2012; Hickok and Poeppel, 2000, 2007). As misheard lyrics are associated with a change in meaning, the ventral stream regions in bilateral superior/middle temporal gyrus are therefore potential candidates mediating induced misperceptions of speech stimuli. Regarding processing of music, comparison of sung versus spoken sentences revealed an overlapping, but differential representation within the temporal lobe (Schön et al., 2010) as well as the inferior frontal cortex (Merrill et al., 2012) and these areas also showed differential activation in the "song versus speech auditory illusion" (Tierney et al., 2013). The critical results of these studies are, however, dependent on activation differences to two types of stimuli (sung versus spoken words) and thus on a bottom-up effect. In the current study, we aimed to test whether these areas are also subject to top–down modulation (i.e. a differential activation to the same physical stimuli driven before versus after experimentally induced misperceptions). Moreover, as such misperceptions reflect changes of the perceived semantic meaning of lyrics, we hypothesized that altered activation partially overlaps with brain areas associated with knowledge of lyrics as determined in fMRI experiment 1.

Materials and methods

Ethics statement

The study was approved by the Ethical Committee of the University of Tübingen (votum: 215/2012BO2) and written informed consent was obtained from all participants. All study procedures were in line with the latest version of the Declaration of Helsinki.

Participants

20 healthy German native speakers (9 women, 11 men, mean age: 28.9 ± 5.4 years, education: 16.8 ± 2.6 years) participated in this study. All participants were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). Mean verbal intelligence of the participants, as obtained by a German vocabulary test (Mehrfach-Wortschatz-Intelligenztest B), was 123.8 ± 13.6. Knowledge of English language was assessed using a language test based on a short form of the Test of English as a Foreign Language (Mini-TOEFL, http://www.ets.org/toefl) to determine the participants' grammar and vocabulary comprehension. On average, the participants correctly answered 14.7 ± 5.4 out of 25 questions. Verbal intelligence and English language comprehension were significantly correlated ($r = 0.53$, two-tailed $p < 0.05$) reflecting the fact that they are similarly driven by the educational level of the participants. All values are given in mean ± standard deviation.

Stimuli

The stimulus set comprised short audio clips (mean duration: 17.7 ± 3.0 s) taken from 18 English and 18 German songs. These 36 songs were selected from a set of 41 English and 20 German songs which were evaluated in a behavioral experiment (Beck et al., 2014) to ensure that both types of stimuli were matched for degree of familiarity with the lyrics, strength of the induced misperception, and wittiness of the alternative lyrics as evaluated on a four-point scale as well as prior knowledge of the alternative lyrics as determined on a two-alternative forced choice decision (see Table 1). No significant differences were found for any of these stimulus characteristics (all $p > 0.10$). All stimuli were normalized to same peak intensity.

Experimental design

All stimuli were presented during four consecutive experimental runs in a fully randomized order. Experimental run 1 and 4 were carried out during fMRI while run 2 and 3 were conducted outside the scanner as behavioral experiments. Directly before each of the four runs, the participants were familiarized with the task by a short training run consisting of 2 stimuli which were not used during the experimental runs. The same two training stimuli were used in all four training runs. In all four experimental runs, the response scales were shown for five seconds and followed by a fixation cross presented for two seconds before onset of the next stimulus. During experimental run 1 (fMRI experiment 1) the participants judged their degree of familiarity with the lyrics on a four-point scale (+++: 'I know most/all of the lyrics', ++: 'I know some of the lyrics in addition to the refrain', +: 'I only know the refrain', 0: 'I do not even know the full refrain'). Before starting experimental run 2 (behavioral experiment 1), the participants were informed that the lyrics can be misheard and that such alternative lyrics will be presented visually on the screen while listening again to the same stimuli in the next two experimental runs. The task of the participants in experimental run 2 was to determine whether these alternative lyrics were already known to them (e.g., because they heard them in radio transmissions or even spontaneously misheard the lyrics in the presented manner) on a two-alternative scale. In experimental

Table 1
Characteristics of employed stimulus material.

	English songs	German songs
Familiarity with lyrics	1.44 ± 0.39	1.16 ± 0.81
Strength of induced misperceptions	1.94 ± 0.48	1.90 ± 0.53
Wittiness of alternative lyrics	1.27 ± 0.42	1.23 ± 0.37
Prior knowledge of alternative lyrics	16% ± 16%	9% ± 10%

All values in mean ± standard deviation.

run 3 (behavioral experiment 2), participants were instructed to rate the wittiness of these potential misperceptions on a four-point scale (+++, ++, +, 0). In experimental run 4 (fMRI experiment 2) the stimuli were presented without visual presentation of lyrics. The participants judged in the scanner if they heard the original lyrics or a misperception on a two-point scale (original, misperception). To avoid lateralization effects caused by motor responses, the arrangement of the symbols of the scales presented during the fMRI experiments was flipped in horizontal direction for half of the participants.

Data acquisition

Structural T_1-weighted images (TR = 2300 msec, TE = 4.18 msec, TI = 900 msec, voxel size: $1 \times 1 \times 1$ mm^3) and functional images (30 axial slices acquired in sequential descending order, slice thickness 4 mm + 1 mm gap, TR = 2.0 s, TE = 35 msec, voxel size: $3 \times 3 \times 5$ mm^3) were acquired with a 3 T scanner (Siemens TIM PRISMA, Erlangen, Germany). Time series consisted of 480 images per experimental run. For correction of image distortions, a field map (36 slices, slice thickness 3 mm + 0.75 mm gap, TR = 400 msec, TE(1) = 4.92 msec, TE(2) = 7.38 msec, voxel size: $3 \times 3 \times 3.75$ mm^3) was acquired. Diffusion-weighted images were acquired using a "Skejskal-Tanner" sequence (TR = 6.0 s, TE = 69 msec, flip angle = 90°, 50 axial slices, slice thickness: 2 mm + 0.5 mm gap, voxel size of $1.7 \times 1.7 \times 2.5$ mm^3, 2 acquisitions) along 30 independent directions using a b-value of 1000 s/mm^2. Additionally, an image with a b-value of 0 s/mm^2 was acquired for coregistration with the fMRI data.

Analysis of behavioral data

All values are given in mean ± standard error of the mean (SEM) unless otherwise specified. For correlation coefficients, we employed back-transformed mean and SEM values of individual Fisher Z scores. Kolmogorov–Smirnov-tests were employed to assess whether the tested variables are normally distributed. Normally distributed data were analyzed using T-tests corrected for potential heteroscedasticity. In case of non-normal distributed data, Mann–Whitney-U-tests and Wilcoxon tests were employed for comparison of independent and dependent samples, respectively, and resulting z scores are reported.

Analysis of fMRI data

Images were analyzed with statistical parametric mapping software (SPM8, Wellcome Department of Imaging Neuroscience, London, UK). Preprocessing comprised realignment, unwarping to correct for field distortions and to remove residual movement-related variance due to interactions between motion and field distortions (Andersson et al., 2001), slice time correction, normalization into MNI space (Montreal Neurological Institute, resampled voxel size: $3 \times 3 \times 3$ mm^3), and smoothing with a Gaussian filter (8 mm full width at half maximum).

For the first experimental run (fMRI experiment 1), hemodynamic responses were modeled using a box-car function convolved with the hemodynamic response function (hrf) corresponding to the duration of the presented stimuli. In addition, a parametric regressor reflecting the familiarity ratings of the participant was included in the first-level model. For the fourth experimental run (fMRI experiment 2), onsets were time-locked to the occurrence of the potential misperception and modeled as a stick function convolved with the hrf. The first-level model included separate regressors for stimuli which were perceived according to the original versus alternative lyrics. Moreover, a parametric regressor reflecting the wittiness ratings as obtained in the third experimental run (behavioral experiment 2) was included for these two conditions. As prior knowledge of alternative lyrics increases the probability of misperceptions (Beck et al., 2014) during the experiment, we also defined a first-level model in which trials with previously known alternative lyrics as obtained in the second experimental run (behavioral experiment 1) were modeled as separate regressor and thus excluded from the contrast of interest (see control analyses).

For both experiments, data from the individual first-level general linear models were employed to create contrast images for each subject which were then submitted to a second-level random-effect analysis to enable population inference.

The following contrasts were obtained:

fMRI experiment 1. To define areas responsive to the familiarity with lyrics we examined the main effect of the parametric regressor reflecting a linear increase in brain activation with familiarity ratings.

fMRI experiment 2. Areas reacting specifically during misperception were defined by the difference contrast of the two types of events (alternative > original lyrics) and were employed as regions-of-interest (ROI) in subsequent fMRI and DWI analyses. Average responses (mean ± standard error of beta values) obtained from these areas were extracted. In addition, we determined regions that showed a parametric response to the wittiness of potential misperceptions according to the ratings obtained in behavioral experiment 2.

All fMRI activation maps are reported using a height threshold of $p < 0.001$ (uncorrected). Correction for multiple comparisons ($p < 0.05$, corrected) across the whole brain was assessed at cluster level using random field theory ($k > 151$ and 93 voxels for fMRI experiment 1 and 2, respectively). Assignment of brain structures and brodmann areas (BA) to activation clusters was done using xjview (http://www.alivelearn.net/xjview8) and the automatic anatomic labeling atlas (AAL, Tzourio-Mazoyer et al., 2002).

Control analyses

The following three control analyses were carried out:

First, to clarify that differential effects were not caused by stimulus-driven differences (e.g. acoustic parameters), we used the first-level model with event-related responses time-locked to perception of original versus alternative lyrics (as identified from behavioral data obtained during fMRI experiment 2) on the data obtained during fMRI experiment 1. We extracted the responses to these two trial types within the clusters showing significant responses in fMRI experiment 2 from both fMRI experiments. This enabled us to plot fMRI responses of trials with and without misperception before and after the induction phase.

Second, to determine whether the activation differences between perceptions of alternative versus original lyrics are not dependent on long-term memory, but can be reliably detected for recently induced misperceptions (i.e. during behavioral experiment 1 and 2 outside the scanner), we analyzed the data of fMRI experiment 2 using a model for which the comparison of the two types of events was restricted to misperceptions that were not previously known to the subject (i.e. events for which the participant indicated that the alternative lyrics was already known during behavioral experiment 1 were modeled as an additional separate regressor of no interest).

Third, to examine potential differences in responses for within- versus across-language trials during fMRI experiment 2, we defined a model with separate regressors for events with perception of original and alternative lyrics for these two types of trials, respectively. Beta values of the ROIs were submitted to a two-factorial repeated-measures ANOVA with perception (alternative versus original lyrics) and language (within- versus across-language trials) as within-subject factors. Please note that the main effect of perception is not reported as the contrast between alternative and original lyrics perception was used to define the ROI. The aim of this analysis was to test whether the ROIs exhibit a main effect of language or an interaction between language and perception.

Analysis of DWI data

Diffusion-weighted data were analyzed using FSL4.04 (FMRIB Software Library, Oxford University, www.fmrib.ox.ac.uk/fsl). Preprocessing of diffusion-weighted images included eddy current correction and averaging across the two acquisitions. Voxel-wise estimates of fiber orientations and their uncertainty were calculated on the basis of a model that accounts for the possibility of crossing fibers within each voxel (Behrens et al., 2007; Behrens et al., 2003b). The two most probable directions within each voxel of the white-matter were determined for each individual subject. The cortical structures (gray matter as obtained by tissue segmentation of the mean normalized structural images of the study participants) within the ROIs were coregistered to individual DW-MRI space (resampled voxel size: $2 \times 2 \times 2$ mm^3) and used as seed volumes for probabilistic fiber tracking. No waypoint or target masks were employed and probabilistic fiber tracking results were thresholded at 1% of the maximum value to minimize false-positive fiber tracks. These thresholded probabilistic fiber tracking maps were then transformed to MNI space. Binary connectivity maps were generated from each resulting data set, and then added across participants. Fiber projections present in more than 75% of the participants were displayed.

Results

Behavioral data

Average ratings of familiarity with lyrics on a scale ranging from 0 to 3 obtained during the first fMRI run were 1.44 ± 0.12 for English and 1.02 ± 0.10 for German songs. Although these mean familiarity ratings were strongly correlated ($r = 0.92$) with those obtained during the prestudy (Beck et al., 2014), the difference in familiarity ratings between the two types of stimuli was slightly more pronounced, but failed to reach significance ($z = 1.41$, two-tailed $p = 0.16$).

Behavioral testing between the two fMRI runs revealed that the study participants were on average familiar with the alternative lyrics in $12.50 \pm 2.95\%$ of the across-language trials and $11.02 \pm 3.15\%$ of the within-language trials, with no significant difference between these two types of stimuli ($z = 0.68$, two-tailed $p = 0.50$). In addition, similar ratings for the wittiness of alternative lyrics were obtained for across- and within-language trials (2.52 ± 0.12 and 2.75 ± 0.08, respectively) which were also not significantly different ($T(34) = 1.70$, two-tailed $p = 0.10$).

During the second fMRI run subjects indicated misperceptions during $41.39 \pm 3.06\%$ of the across- and $34.44 \pm 3.64\%$ of the within-language trials. Again, the frequency of misperceptions was not significantly different for the two trial types ($T(34) = 1.45$, two-tailed $p = 0.15$). The statistical comparison of trials with and without misperceptions revealed no significant difference with respect to familiarity ratings (1.09 ± 0.11 versus 1.30 ± 0.11 for alternative versus original lyrics, respectively, $T(19) = 1.92$, two-tailed $p = 0.07$). However, previous knowledge of the alternative lyrics was strongly predictive for misperceptions ($57.88 \pm 5.82\%$ for known versus $33.94 \pm 3.18\%$ for unknown alternative lyrics, $\mathbf{z = 3.58}$, two-tailed $p < 0.001$). Three participants had to be excluded from this comparison because all alternative lyrics were unknown to them.

In addition, trials with induced misperceptions were characterized with higher ratings of wittiness than those without misperceptions (1.61 ± 0.10 versus 1.21 ± 0.08 for alternative versus original lyrics, respectively, $T(19) = 4.20$, two-tailed $p < 0.001$, scale ranging from 0 to 3). Significantly higher wittiness ratings were also observed if across- and within-language trials were analyzed separately (both $T(19) > 3.08$, two-tailed $p < 0.01$).

Brain activation

The parametric analysis of brain responses and familiarity ratings as obtained in the first fMRI experiment revealed a positive linear relationship in the left temporoparietal junction (BA 40) and a large cluster comprising left prefrontal cortex on its lateral aspect including Broca's area (BA 44–47) and medial aspect (BA 6, 8–10, 22, 24, 32), bilateral caudate nucleus/putamen, and right cerebellum (see Table 2 and Fig. 1). Comparison of trials during which the participants indicated perception of the alternative versus original lyrics in the second fMRI experiment (see Table 3 and Fig. 2) yielded activation in bilateral mediodorsal thalamus (mdTh)/medial geniculate body (MGB), superior and middle temporal gyrus along the middle part of the superior temporal sulcus (mSTS, BA 21, 22), inferior frontal gyrus (IFG, BA 9, 45, 47) including parts of the anterior insula, and supplemental motor area/dorsal anterior cingulate cortex (SMA/dACC, BA 6, 8, 9, 32). These activation clusters were used as ROIs for the subsequent control analyses. The parametric analysis between event-related responses and wittiness ratings yielded a significant activation in the left dACC (BA 24, 32; MNI coordinates: $x = -6$, $y = -18$, $z = 39$, Z score $= 3.66$, cluster size $= 123$ voxels) which partly overlaps (43 voxels) with the activation cluster in SMA/dACC defined by comparison of trials with alternative versus original lyrics.

Control analysis 1 revealed that beta estimates obtained during fMRI experiment 1 comparing trials with and without misperceptions revealed no significant activation differences in any of the ROIs (see Fig. 2, all $T(19) < 1.7$, all $p > 0.05$) indicating that increased activation was not driven by general differences of the stimulus material (i.e. acoustic parameters). Control analysis 2 which excluded trials with misperceptions that were previously known to the participants yielded very similar results (MNI coordinates, Z scores, and cluster sizes are given in brackets in Table 3). Control analysis 3 which estimated betas separately for within- and across-language trials of these clusters were submitted to an ANOVA with language (within- versus across-language trials) and perception (original versus alternative lyrics) as within-subject factor. Data from one participant was removed from control analysis 3 as only one misperception occurred for across-language trials (in this case the regressor modeling the misperception trial and the regressor modeling the parametric effect of wittiness would be perfectly correlated and are thus not estimable). In all six activation clusters, no effect of language (all $F(1,18) < 3.3$, all $p > 0.05$) or interactions between language and perception (all $F(1,18) < 4.1$, all $p > 0.05$) was found.

Structural connectivity

Probabilistic fiber tracking was conducted using voxels situated in the cortex of activation clusters defined by the comparison of trials with alternative versus originally perceived lyrics in fMRI experiment 2. Fiber tracking using bilateral mSTS as seed area revealed consistent connections with ipsilateral IFG as well as the auditory pathway terminating near the MGB (see Fig. 3a). Fiber tracking from bilateral IFG revealed connections with ipsilateral mSTS as well as SMA/dACC and mdTh (see Fig. 3b, c). Correspondingly, using bilateral SMA/dACC as seed areas demonstrated consistent fiber connections with ipsilateral IFG and mdTh (see Fig. 3d, e). A schematic overview of the network of brain regions involved in lyrical misperception in displayed in Fig. 3f.

Table 2
Areas showing a linear relationship between activation and familiarity with lyrics.

	MNI coordinates	Z score	Cluster size
Left frontal cortex/right cerebellum/striatum			5877*
Left inferior frontal gyrus	−48 −45 12	4.84	
Right caudate/putamen	15–15 24	4.72	
Right cerebellum	33–60 −42	4.64	
Left supplemental motor area/cingulate cortex	−3 18 60	4.52	
Left caudate nucleus/putamen	−18 9 9	4.39	
Left temporoparietal junction	−54 −45 30	4.84	539*

* $p < 0.05$, correct at cluster level ($k > 151$ voxels).

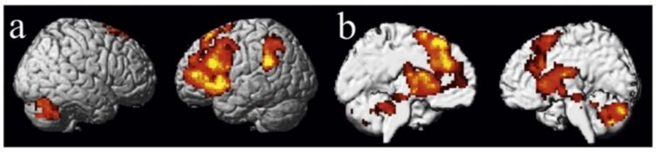

Fig. 1. Brain areas showing a linear relationship with familiarity with lyrics during fMRI experiment 1 (height threshold: $p < 0.001$, extent threshold $k > 151$ voxels, $p < 0.05$, corrected) rendered on the lateral (a) and medial (b) surface of both hemispheres.

Discussion

In the current study, we examined the network of brain areas implicated in knowledge and misperception of lyrics. In line with our hypothesis, a linear relationship of brain activation and familiarity ratings for lyrics implicated speech-related areas in the left hemisphere as relevant network for storage of knowledge for lyrics. Misperception of lyrics, however, activated a mostly symmetrical network in both hemispheres including MGB, mSTS, IFG, dACC/SMA and mdTh. We combined our fMRI results with DWI analysis to additionally investigate the structural connectivity of the involved networks, but restricted this approach to the brain regions activated in fMRI experiment 2 as numerous previous DWI studies (e.g. Catani et al., 2005; Glasser and Rilling, 2008; Saur et al., 2008) have been conducted to explore fiber connections of speech-related areas.

Knowledge of lyrics

While a few neuroimaging studies have examined general effects of lyrics on brain activity during music perception (Brattico et al., 2011; Herholz et al., 2012; Sammler et al., 2010) fMRI experiment 1 represents the first study that investigates the impact of familiarity with lyrics on music processing. A linear relationship between brain activity and individual familiarity ratings was found within predominantly left-hemispheric regions of the speech processing network (Hickok, 2012; Hickok and Poeppel, 2000, 2007) including medial and lateral aspects of the left frontal lobe as well as the left temporoparietal junction corresponding to the Sylvian parietal temporal region which has been implicated in auditory rehearsal/vocal imagery (Buchsbaum and D'Esposito, 2008; Buchsbaum et al., 2005; Hickok et al., 2003), covert singing (Koelsch et al., 2009), and covert humming (Pa and Hickok, 2008). In addition we observed activation within the right superior cerebellum which is in line with recent meta-analyses on word retrieval (Price, 2012) as well as subcortical–cortical models on speech perception (Kotz and Schwartze, 2010).

These findings obtained by a parametric approach are in line with previous results comparing activation to familiar and unfamiliar songs in a categorical subtraction design (Janata, 2009). Furthermore, the identified activation pattern strikingly resembles fMRI findings obtained during spontaneous lyrical improvisation of freestyle rap (Liu et al., 2012) indicating that retrieval of lyrics and their generation recruit similar neural sources supporting speech comprehension and word retrieval (for a review, see Price, 2012). It should be noted that the pattern of activated areas associated with familiarity of lyrics differs strongly from those obtained by adaptation designs showing a bilateral reduction responses in mSTS cortices (Sammler et al., 2010) indicating that long-term effects, such as familiarity with lyrics, and short-term processes due to repeated stimulus exposure of the same lyrical excerpt recruit different systems.

In addition, bilateral striatal and midbrain responses associated with lyrical familiarity were found (see yellow/red activation clusters of cerebral midline structures in Fig.1), replicating previous neuroimaging results obtained for familiarity with instrumental music (Trost et al., 2012) and song excerpts (Pereira et al., 2011). A link between familiarity and esthetic judgments in music ("we like what we know/we dislike what is new to us") has already been proposed in the early 20th century (Meyer, 1903) manifesting in a higher probability to rate a stimulus as pleasing if it is familiar (so called mere exposure effect, Peretz et al., 1998). This phenomenon occurs similarly for perception of melodies (Green et al., 2012) and dancing moves (Kirsch et al., 2015). Thus, it seems plausible to suggest that stimulus familiarity recruits dopaminergic components of the reward system during various aspects of music perception. It should be noted that we acquired the familiarity ratings during the first presentation of the stimuli (i.e. during fMRI experiment 1) to avoid potential changes of familiarity ratings or mere-exposure effects (Peretz et al., 1998) due to the repeated presentation during the course of the experiment. A potential drawback of this design is that we cannot exclude that the participants might have found it pleasing to rate a stimulus as familiar (and thus successfully documented their knowledge on lyrics to the experimenters) which might constitute an alternative explanation for the activation in the reward system. Therefore, future neuroimaging experiments on familiarity effects should obtain ratings outside the scanner (after fMRI acquisition) to address this issue. Besides that common component in the ventral striatum, however, familiarity evaluation seems to be linked to the specific task demands determined by the nature of the stimulus. This is reflected by activation of the anterior temporal cortex during perception of music (Trost et al., 2012), motion-sensitive cortex in the STS during judgment of dancing moves (Kirsch et al., 2015), or left hemispheric speech-related areas in the frontal and temporal lobe in our study.

Misperception of lyrics

The aim of fMRI experiment 2 was to clarify which areas show differential activations if experimentally induced misperceptions occur. As these misperceptions included an alteration of speech-related information, we expected that these activation changes occur within the regular speech perception network. In line with prediction with the dual-stream model for speech processing (Hickok, 2012; Hickok and Poeppel, 2000, 2007) activation differences were found in a mostly symmetrical network of temporal and frontal regions belonging to the ventral stream that maps sound to meaning rather than within the predominantly left lateralized dorsal stream that maps sound to actions.

Excluding trials with misperception of lyrics that were already known to the participants did not compromise these results, but rather yielded higher statistical values in spite of the fact that this analysis

Table 3
Stronger activation during perception of alternative versus original lyrics.

	MNI coordinates	Z score	Cluster size
Right IFG	36 24 −12 (36 24 −9)	5.17 (5.16)	700 (970)*
Bilateral ACC/SMA	−9 18 45 (3 12 69)	4.48 (4.75)	486 (602)*
Left IFG	−33 15 −6 (−33 15 0)	4.47 (4.84)	535 (1158)*
Right MTG	54 −21 −3 (54 −18 −6)	4.32 (4.85)	249 (363)*
Bilateral mediodorsal Thalamus	−3 −33 −3 (12 −6 -3)	4.26 (5.43)	126 (616)*
Left MTG	−63 −27 −9 (−57 −30 −6)	4.20 (4.92)	183 (410)*

* $p < 0.05$, corrected at cluster size ($k > 93$ voxels), MNI coordinates, Z scores, and cluster size obtained if trials for which the alternative song text was previously known to the participants were excluded are given in brackets.

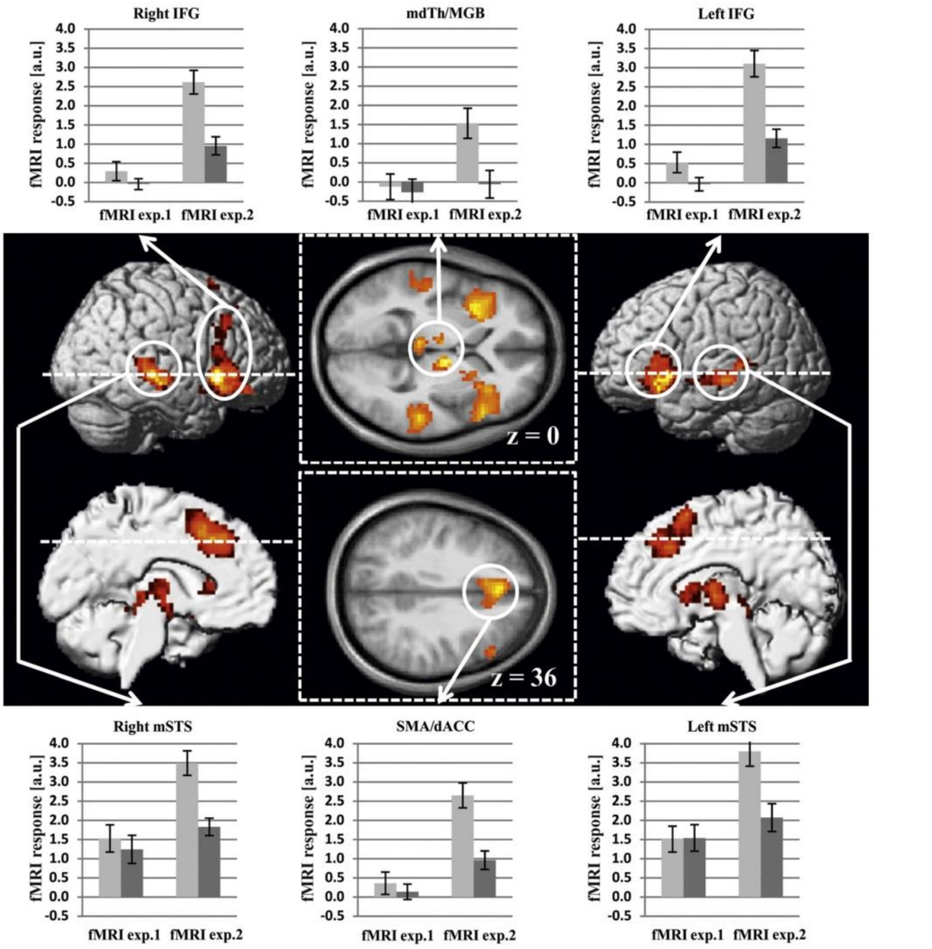

Fig. 2. Stronger response during perception of the alternative versus original lyrics in fMRI experiment 2 (height threshold: $p < 0.001$, extent threshold $k > 93$ voxels, $p < 0.05$, corrected) including ROI analysis in bilateral IFG, mSTS, dACC, and mdTh/MGB of trials with alternative versus original lyrics (mean ± standard error, light gray versus dark gray bars, respectively) plotted separately for fMRI experiment 1 (left bars) and fMRI experiment 2 (right bars). All fMRI responses are shown in arbitrary units (a.u.).

included fewer trials. A possible explanation for this observation is that previously known Soramimi/Mondegreens result in smaller activation changes than newly induced misperceptions due to habituation effects. However, as previous knowledge of alternative lyrics occurred only in a small number of trials and was inhomogeneously distributed across participants (e.g. some participants knew not a single Mondegreen/Soramimi before the experiment), it was impossible to investigate this systematically.

At the level of the temporal cortex, secondary auditory areas along the mSTS elicited increased responses during misperception which is in line with fMRI adaptation results on processing of lyrics and their integration with tunes (Sammler et al., 2010). Moreover, this area has been repeatedly linked to auditory perceptual phenomena such as the continuity illusion (Heinrich et al., 2008, 2011) and encoding of ambiguous sounds (Kilian-Hutten et al., 2011). Our findings on misperception of lyrics extend these previous findings obtained for relatively simple sounds (i.e., vowels and syllables) to more complex stimuli and suggest a more general role of mSTS cortices in processing ambiguous auditory information that requires rapid interpretation. This is also in line with clinical results demonstrating that damage to this area can induce auditory hallucinations and illusions (Wong et al., 2011).

Misperception of lyrics furthermore engaged bilateral IFG partly including Broca's area and its right-sided homolog. While the left-hemispheric activations are most probably due to the well-known role of Broca's area during semantic and syntactic processing (for a review, see Bookheimer, 2002) and overlapped with the frontal activation cluster as determined in fMRI experiment 1, activations in right IFG have been repeatedly observed during processing of pitch (e.g. Brancucci

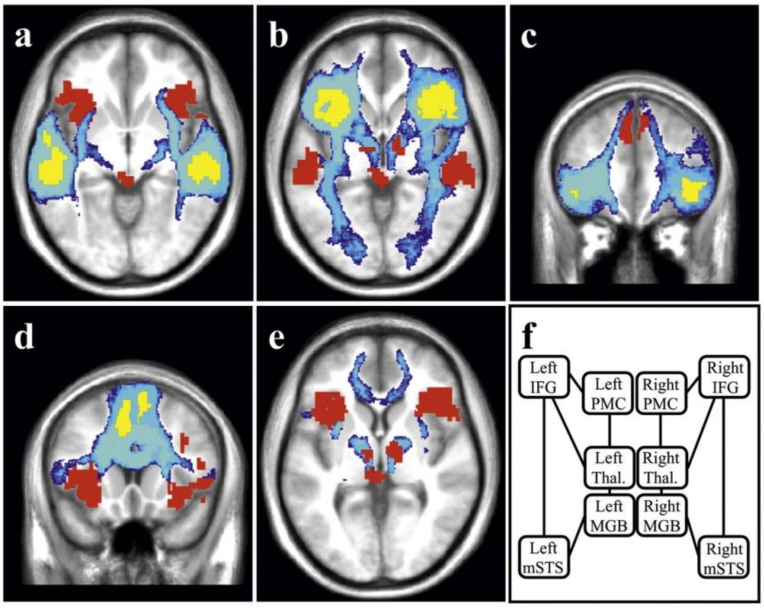

Fig. 3. Consistent structural connections (≥75% of participants) of the cortical ROIs defined in fMRI experiment 2 are displayed in blue/light blue using bilateral MTG (a), IFG (b, c), and PMC/dACC (d, e) as seed areas and displayed in a schematic diagram (f). Seed areas used for fiber tracking are shown in yellow while other ROIs are displayed in red.

et al., 2011; Zatorre et al., 1996), speech melody (e.g., Wildgruber et al., 2005), word processing in songs (Merrill et al., 2012), as well as various inhibition processes (Aron et al., 2004) including verbal suppression (Robinson et al., 2015). Thus, the bilateral IFG activation observed in the current study might reflect the necessity to embed a new semantic context within the given rhythm and melody which is in line with previous results indicating interaction of linguistic and musical processing in this area (Schön et al., 2010) and the observation that electrical stimulation of this area induces switching from speaking to singing during awake neurosurgery (Herbet et al., 2015).

The stronger activation of the SMA/dACC during alternative versus originally perceived trials is well in line with the known engagement of this area in error detection and conflict monitoring (for a review, see Bush et al., 2000). Concerning processing of music, this region has been implicated in singing (Perry et al., 1999) and musical imagery (Halpern, 2001; Herholz et al., 2012). Recently, a role for imagery of temporal patterns has been suggested for this area (Schaefer, 2014). In line with this proposed role, activity in SMA/dACC could reflect embedding of the alterative percept into the temporal structure of the instrumental music. Activation in this area also correlated with wittiness ratings of alternative lyrics concurring with findings made in the visual domain where resolving ambiguities of percepts has been found to generate pleasant feelings similar to the experience of problem-solving (Ramachandran and Rogers-Ramachandran, 2007) as well as the long-existing theory (Helmholtz, 1866) that perception and problem-solving recruit similar processes. Moreover, a similar correlation has been observed for SMA/dACC responses and intensity ratings of experienced chills (Blood and Zatorre, 2001). Such chills similarly reflect events during music perception which are highly reproducible within a given individual and are associated with pleasure. It should be noted, however, that the regression analysis of Blood and Zatorre (2001) revealed a much larger network than the one found in our study and also included reward-related areas in striatum and brainstem mediating the physiological changes associated with the feelings of "shivers-down-the-spine" in chills which do not occur in the misperception phenomena studied here.

Finally, subcortical structures including the MGB and mdTh also reacted stronger during trials with versus without misperception. Increased activation within the MGB most probably reflects modulatory feedback from auditory areas via reciprocal connections (Pickles, 2015), and circumscribed lesions of the MGB can induce complex auditory phenomena including illusions (Fukutake and Hattori, 1998). The involvement of the mdTh in the cognitive and social life of an individual has been suggested more than one hundred years ago (Monakow, 1895) and it has been recognized as a higher-order relay for executive functions (for a recent review, see Mitchell, 2015). Damage to this area results in consistent learning deficits and memory loss (diencephalic amnesia, Mitchell and Chakraborty, 2013) as well as language disturbances (Barbas et al., 2013) and thus it might serve as a hub for links with prefrontal structures during acquisition of novel auditory information.

Structural connectivity

We used DWI in combination with probabilistic fiber tracking to explore the structural connections within the network of brain areas specifically engaged during misperception of lyrics. Fiber tracking using the mSTS as seed region revealed fiber connections towards the MGB which correspond to the acoustic radiations which mostly project to the primary auditory cortex, but also higher-order acoustic areas (Pickles, 2015). This pathway terminated a few millimeters before reaching the activation cluster in the MGB which is most probably due to the well-known fact that in DWI the acoustic radiations from the MGB are obscured by the optic radiations of the lateral geniculate nucleus (Behrens et al., 2007). We also identified a projection of the mSTS area to the inferior frontal cortex running through the external/extreme capsule which accords with tracer studies in rhesus monkeys (Padberg et al., 2003; Seltzer and Pandya, 1989). Notably, this temporofrontal fiber connection has been previously identified in studies on processing vocal emotions (Ethofer et al., 2012; Fruhholz et al., 2015), suggesting that these structural connections subserve a more general role during processing of auditory information requiring rapid interpretation including both emotionally salient as well as ambiguous signals.

In addition, the IFG was also found to be connected with the SMA/dACC replicating previous neuroimaging data in humans (Ethofer et al., 2013) as well as tracer results in macaques (Petrides and Pandya, 2002; Petrides et al., 2012) and current subcortical–cortical framework models on speech perception (Kotz and Schwartze, 2010). The functional significance of this connection, however, is largely unknown. Animal experiments implicated this region in initiation of vocal behavior as a 'cingulate vocalization area' (Jürgens, 2009; Romanski, 2012). In humans, this area has further evolved as fMRI experiments evidenced its participation in higher-order cognitive processes, such as active emotional evaluation of voice descriptions in written text samples (Brück et al., 2014).

Both the IFG and the dACC were found to be connected to the mdTh. This concurs with DWI studies in humans which repeatedly showed that mdTh represent the part of the thalamus which is connected with most parts of the prefrontal cortex (Behrens et al., 2003a; Kumar et al., 2014) and it has been suggested that the mdTh reaches its peak dimensions and cytoarchitectonic complexity in parallel with the expansion of the prefrontal cortex (Jones, 2007). In addition to these general findings on connectivity of mdTh and prefrontal cortex, however, monkey data specifically demonstrated that the mdTh receives input from ventrolateral prefrontal cortex (McFarland and Haber, 2002) and is reciprocally connected to the ventromedial prefrontal cortex (Giguere and Goldman-Rakic, 1988) corresponding to the IFG and SMA/dACC area found to be activated in our study, respectively. Our DWI results indicate that the connections to mdTh also reach the MGB which could represent the anatomically substrate of top–down modulation of early structures during (mis-) perception of auditory information.

Limitations

The results of the current study need to be discussed in light of limitations of the employed response-related fMRI paradigms and general issues of probabilistic fiber tracking of DWI data: In fMRI experiment 1, we explicitly instructed the participants to rate their familiarity with the lyrics and not other aspects of the song excerpts. To emphasize this, each step on the four-point scale was assigned a description regarding the participant's knowledge of the lyrics of the presented song excerpts (see methods). Although the obtained distribution of activation patterns is very similar to those obtained during generation of lyrics (Liu et al., 2012), it is not unlikely that familiarity with lyrics is correlated with familiarity with other aspects of the songs (e.g., the musical structure or even semantic knowledge about the singer). To fully disentangle such effects it would be necessary to obtain a stimulus set which is orthogonal to these features (e.g. to obtain stimuli for which the participant is familiar with the lyrics, but not the musical structure and vice versa). This is particularly challenging as it had to be assured that this orthogonality is given for each and every participant and a large interindividual variability concerning this issue has to be expected.

In fMRI experiment 2, we employed a previously established paradigm relying on external induction of auditory misperceptions (Beck et al., 2014) that occur with a sufficient probability (30–40%) to study such phenomena in neuroscientific experiments. Thus, our experimental manipulation examines expectation-related alterations of perception. It should be noted, however, that even "spontaneous" misperceptions are often similarly driven by – albeit internally generated – expectations, e.g. due to semantic proximity as in the example reported in the introduction. We induced within- as well as across-language misperceptions. No significant activation differences were found for these two stimulus types at the whole-brain level or in the ROI analyses arguing for a recruitment of similar neural substrates irrespective of whether misperceptions include a switch between two different languages or not. This negative finding, however, should be carefully interpreted and it might be possible that smaller differences in brain activation are detectable if more stimuli are included in the paradigm.

We discussed our structural connectivity findings on the basis of previously published DWI studies in humans and tracer studies in monkeys. Probabilistic fiber tracking of DWI data relies on the identification of the axis along which diffusion is most likely, but no directional information can be inferred and results can be obscured by fiber crossings. In this aspect, DWI is clearly inferior to the methodology offered by anterograde and retrograde tracer studies in animals. Potential differences in neural architecture across species, especially in the prefrontal cortex (Petrides et al., 2012), limit the ability to draw inference on the human brain based on data obtained in macaques. This issue is particularly severe if the aim of the study involves neural substrates of speech and language and thus validation of monkey neuroanatomy using neuroimaging in humans can provide complementary information (Yeterian et al., 2012).

Conclusions

While familiarity of lyrics was associated with activation in reward-related areas and left-hemispheric speech regions, misperception of lyrics yielded a symmetric bilateral network in thalamus and frontotemporal cortex. Our combined fMRI/DWI results offer for the first time insights into the neural underpinnings of misperceptions which occur within the ventral stream of the regular speech perception network mapping sound to meaning. In this study, we used lyrics because these stimuli feature important factors that generally favor misperceptions of speech: atypical pronunciation resulting in ambivalent speech signals in combination with perception in adverse conditions due to presence of other acoustic signals (i.e. the instrumental music). Thus, although Mondegreens and Soramimi represent clearly harmless misperceptions, our results can serve as a model for future studies on the neural circuits of other types of misunderstandings which can have a great influence in the social life of humans.

Acknowledgments

This study was supported by the Deutsche Forschungsgemeinschaft (Grants: DFG ET 112/5-1 and ET 112/6-1).

References

Andersson, J., Hutton, C., Ashburner, J., Turner, R., Friston, K., 2001. Modeling geometric deformations in EPI time series. NeuroImage 13, 903–912.
Andrews, T.J., Schluppeck, D., Homfray, D., Blakemore, C., 2002. Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion. NeuroImage 17, 890–901.
Aron, A.R., Robbins, T.W., Poldrack, R.A., 2004. Inhibition and the right inferior frontal cortex. Trends Cogn. Sci. 8, 170–177.
Barbas, H., Garcia-Cabezas, M.A., Zikopoulos, B., 2013. Frontal-thalamic circuits associated with language. Brain Lang. 126, 49–61.
Beck, C., Kardatzki, B., Ethofer, T., 2014. Mondegreens and Soramimi as a method to induce misperceptions of speech content–influence of familiarity, wittiness, and language competence. PLoS One 9, e84667.
Behrens, T.E., Berg, H.J., Jbabdi, S., Rushworth, M.F., Woolrich, M.W., 2007. Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? NeuroImage 34, 144–155.
Behrens, T.E., Johansen-Berg, H., Woolrich, M.W., Smith, S.M., Wheeler-Kingshott, C.A., Boulby, P.A., Barker, G.J., Sillery, E.L., Sheehan, K., Ciccarelli, O., Thompson, A.J., Brady, J.M., Matthews, P.M., 2003a. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nat. Neurosci. 6, 750–757.

Behrens, T.E., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003b. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. 50, 1077–1088.

Blood, A.J., Zatorre, R.J., 2001. Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proc. Natl. Acad. Sci. U. S. A. 98, 11818–11823.

Bookheimer, S., 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annu. Rev. Neurosci. 25, 151–188.

Brancucci, A., Franciotti, R., D'Anselmo, A., Della Penna, S., Tommasi, L., 2011. The sound of consciousness: neural underpinnings of auditory perception. J. Neurosci. 31, 16611–16618.

Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., Tervaniemi, M., 2011. A functional MRI study of happy and sad emotions in music with and without lyrics. Front. Psychol. 2, 308.

Brück, C., Kreifelts, B., Gossling-Arnold, C., Wertheimer, J., Wildgruber, D., 2014. 'Inner voices': the cerebral representation of emotional voice cues described in literary texts. Soc. Cogn. Affect. Neurosci. 9, 1819–1827.

Buchsbaum, B.R., D'Esposito, M., 2008. The search for the phonological store: from loop to convolution. J. Cogn. Neurosci. 20, 762–778.

Buchsbaum, B.R., Olsen, R.K., Koch, P., Berman, K.F., 2005. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. Neuron 48, 687–697.

Bush, G., Luu, P., Posner, M.I., 2000. Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn. Sci. 4, 215–222.

Calabrese, A., 2012. Auditory representations and phonological illusions: a linguist's perspective on the neuropsychological bases of speech perception. Afr. J. Neurol. Sci. 25, 355–381.

Catani, M., Jones, D.K., ffytche, D.H., 2005. Perisylvian language networks of the human brain. Ann. Neurol. 57, 8–16.

Cutler, A., Butterfield, S., 1992. Rhythmic cues to speech segmentation: evidence from juncture misperception. J. Mem. Lang. 31, 218–236.

Deutsch, D., 1974. An auditory illusion. Nature 251, 307–309.

Ethofer, T., Bretscher, J., Gschwind, M., Kreifelts, B., Wildgruber, D., Vuilleumier, P., 2012. Emotional voice areas: anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. Cereb. Cortex 22, 191–200.

Ethofer, T., Bretscher, J., Wiethoff, S., 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.

Fruhholz, S., Gschwind, M., Grandjean, D., 2015. Bilateral dorsal and ventral fiber pathways for the processing of affective prosody identified by probabilistic fiber tracking. NeuroImage 109, 27–34.

Fukutake, T., Hattori, T., 1998. Auditory illusions caused by a small lesion in the right medial geniculate body. Neurology 51, 1469–1471.

Giguere, M., Goldman-Rakic, P.S., 1988. Mediodorsal nucleus: areal, laminar, and tangential distribution of afferents and efferents in the frontal lobe of rhesus monkeys. J. Comp. Neurol. 277, 195–213.

Glasser, M.F., Rilling, J.K., 2008. DTI tractography of the human brain's language pathways. Cereb. Cortex 18, 2471–2482.

Green, A.C., Baerentsen, K.B., Stodkilde-Jorgensen, H., Roepstorff, A., Vuust, P., 2012. Listen, learn, like! Dorsolateral prefrontal cortex involved in the mere exposure effect in music. Neurol. Res. Int. 2012, 846270.

Halpern, A.R., 2001. Cerebral substrates of musical imagery. Ann. N. Y. Acad. Sci. 930, 179–192.

Hasson, U., Hendler, T., Ben Bashat, D., Malach, R., 2001. Vase or face? A neural correlate of shape-selective grouping processes in the human brain. J. Cogn. Neurosci. 13, 744–753.

Heinrich, A., Carlyon, R.P., Davis, M.H., Johnsrude, I.S., 2008. Illusory vowels resulting from perceptual continuity: a functional magnetic resonance imaging study. J. Cogn. Neurosci. 20, 1737–1752.

Heinrich, A., Carlyon, R.P., Davis, M.H., Johnsrude, I.S., 2011. The continuity illusion does not depend on attentional state: FMRI evidence from illusory vowels. J. Cogn. Neurosci. 23, 2675–2689.

Helmholtz, v., 1866. Concerning the Perceptions in General, third ed.

Herbet, G., Lafargue, G., Almairac, F., Moritz-Gasser, S., Bonnetblanc, F., Duffau, H., 2015. Disrupting the right pars opercularis with electrical stimulation frees the song: case report. J. Neurosurg. 1–4.

Herholz, S.C., Halpern, A.R., Zatorre, R.J., 2012. Neuronal correlates of perception, imagery, and memory for familiar tunes. J. Cogn. Neurosci. 24, 1382–1397.

Hickok, G., 2012. The cortical organization of speech processing: feedback control and predictive coding the context of a dual-stream model. J. Commun. Dis. 45, 393–402.

Hickok, G., Buchsbaum, B., Humphries, C., Muftuler, T., 2003. Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J. Cogn. Neurosci. 15, 673–682.

Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn. Sci. 4, 131–138.

Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402.

Huang, Y., Rao, R.P., 2011. Predictive coding. Wiley Interdiscip. Rev. Cogn. Sci. 2, 580–593.

Janata, P., 2009. The neural architecture of music-evoked autobiographical memories. Cereb. Cortex 19, 2579–2594.

Jones, E.G., 2007. The Thalamus 2 Volume Set. Cambridge University Press.

Jürgens, U., 2009. The neural control of vocalization in mammals: a review. J. Voice 23, 1–10.

Kilian-Hutten, N., Valente, G., Vroomen, J., Formisano, E., 2011. Auditory cortex encodes the perceptual interpretation of ambiguous sound. J. Neurosci. 31, 1715–1720.

Kirsch, L.P., Dawson, K., Cross, E.S., 2015. Dance experience sculpts aesthetic perception and related brain circuits. Ann. N. Y. Acad. Sci. 1337, 130–139.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Muller, K., Gruber, O., 2009. Functional architecture of verbal and tonal working memory: an FMRI study. Hum. Brain Mapp. 30, 859–873.

Kotz, S.A., Schwartze, M., 2010. Cortical speech processing unplugged: a timely subcortico-cortical framework. Trends Cogn. Sci. 14, 392–399.

Kumar, V., Mang, S., Grodd, W., 2014. Direct Diffusion-based Parcellation of the Human Thalamus Brain Struct Funct.

Liu, S., Chow, H.M., Xu, Y., Erkkinen, M.G., Swett, K.E., Eagle, M.W., Rizik-Baer, D.A., Braun, A.R., 2012. Neural correlates of lyrical improvisation: an FMRI study of freestyle rap. Sci. Rep. 2, 834.

Mattys, S.L., Davis, M.H., Bradlow, A.R., Scott, S.K., 2012. Speech recognition in adverse conditions: a review. Lang. Cogn. Process. 27, 953–978.

McFarland, N.R., Haber, S.N., 2002. Thalamic relay nuclei of the basal ganglia form both reciprocal and nonreciprocal cortical connections, linking multiple frontal cortical areas. J. Neurosci. 22, 8117–8132.

Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., Friederici, A.D., 2012. Perception of words and pitch patterns in song and speech. Front. Psychol. 3, 76.

Meyer, M., 1903. Experimental studies in the psychology of music. Am. J. Psychol. 14, 192–214.

Mitchell, A.S., 2015. The mediodorsal thalamus as a higher order thalamic relay nucleus important for learning and decision-making. Neurosci. Biobehav. Rev.

Mitchell, A.S., Chakraborty, S., 2013. What does the mediodorsal thalamus do? Front. Syst. Neurosci. 7. http://dx.doi.org/10.3389/fnsys.

Monakow, C., 1895. Experimentelle und pathologisch-anatomische Untersuchungen über die Haubenregion, den Sehhügl und die Regio subthalamica, nebst Beiträgen zur Kenntniss früh erworbener Gross- und Kleinhirndefekte. Arch. Psychiatr. Nervenkr. 27 (1–128), 386–479.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Pa, J., Hickok, G., 2008. A parietal–temporal sensory-motor integration area for the human vocal tract: evidence from an fMRI study of skilled musicians. Neuropsychologia 46, 362–368.

Padberg, J., Seltzer, B., Cusick, C.G., 2003. Architectonics and cortical connections of the upper bank of the superior temporal sulcus in the rhesus monkey: an analysis in the tangential plane. J. Comp. Neurol. 467, 418–434.

Pereira, C.S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S.L., Brattico, E., 2011. Music and emotions in the brain: familiarity matters. PLoS One 6, e27241.

Peretz, I., Gaudreau, D., Bonnel, A.M., 1998. Exposure effects on music preference and recognition. Mem. Cogn. 26, 884–902.

Perry, D.W., Zatorre, R.J., Petrides, M., Alivisatos, B., Meyer, E., Evans, A.C., 1999. Localization of cerebral activity during simple singing. Neuroreport 10, 3979–3984.

Petrides, M., Pandya, D.N., 2002. Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. Eur. J. Neurosci. 16, 291–310.

Petrides, M., Tomaiuolo, F., Yeterian, E.H., Pandya, D.N., 2012. The Prefrontal Cortex: Comparative Architectonic Organization in the Human and the Macaque Monkey Brains. Cortex; a Journal Devoted to the Study of the Nervous System and Behavior Vol. 48 pp. 46–57.

Pickles, J.O., 2015. Auditory Pathways: Anatomy and Physiology. Elsevier.

Price, C.J., 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. NeuroImage 62, 816–847.

Ramachandran, V.S., Rogers-Ramachandran, D., 2007. Ambiguities and perception. Sci. Am. Mind 18, 18–20.

Robinson, G.A., Cipolotti, L., Walker, D.G., Biggs, V., Bozzali, M., Shallice, T., 2015. Verbal suppression and strategy use: a role for the right lateral prefrontal cortex? Brain 138, 1084–1096.

Romanski, L.M., 2012. Integration of faces and vocalizations in ventral prefrontal cortex: implications for the evolution of audiovisual speech. Proc. Natl. Acad. Sci. U. S. A. 109 (Suppl. 1), 10717–10724.

Sammler, D., Baird, A., Valabregue, R., Clement, S., Dupont, S., Belin, P., Samson, S., 2010. The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. J. Neurosci. 30, 3572–3578.

Saur, D., Kreher, B.W., Schnell, S., Kummerer, D., Kellmeyer, P., Vry, M.S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., Weiller, C., 2008. Ventral and dorsal pathways for language. Proc. Natl. Acad. Sci. U. S. A. 105, 18035–18040.

Schaefer, R.S., 2014. Images of time: temporal aspects of auditory and movement imagination. Front. Psychol. 5, 877.

Schön, D., Gordon, R., Campagne, A., Magne, C., Astesano, C., Anton, J.L., Besson, M., 2010. Similar cerebral networks in language, music and song perception. NeuroImage 51, 450–461.

Seltzer, B., Pandya, D.N., 1989. Frontal lobe connections of the superior temporal sulcus in the rhesus monkey. J. Comp. Neurol. 281, 97–113.

Sohoglu, E., Peelle, J.E., Carlyon, R.P., Davis, M.H., 2014. Top-down influences of written text on perceived clarity of degraded speech. J. Exp. Psychol. Hum. Percept. Perform. 40, 186–199.

Tierney, A., Dick, F., Deutsch, D., Sereno, M., 2013. Speech versus song: multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. Cereb. Cortex 23, 249–254.

Trost, W., Ethofer, T., Zentner, M., Vuilleumier, P., 2012. Mapping aesthetic musical emotions in the brain. Cereb. Cortex 22, 2769–2783.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM

using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15, 273–289.
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.
Wong, V.S., Adamczyk, P., Dahlin, B., Richman, D.P., Wheelock, V., 2011. Cerebral venous sinus thrombosis presenting with auditory hallucinations and illusions. Cogn. Behav. Neurol. 24, 40–42.
Wright, S., 1954. The death of Lady Mondegreen. Harper's Mag. 209, 48–51.
Yeterian, E.H., Pandya, D.N., Tomaiuolo, F., Petrides, M., 2012. The cortical connectivity of the prefrontal cortex in the monkey brain. Cortex; A Journal Devoted to the Study of the Nervous System and Behavior. Vol. 48, pp. 58–81.
Zatorre, R.J., Meyer, E., Gjedde, A., Evans, A.C., 1996. PET studies of phonetic processing of speech: review, replication, and reanalysis. Cereb. Cortex 6, 21–30.

3 Discussion

3.1 A new method for inducing concurring auditory perception

Within the visual domain there exists a long-standing tradition of using ambiguous stimuli material such as Ruben's face-vase illusion. It also has been used for knowing more about the cerebral processing involved in these concurring percepts (e.g. Andrews et al. 2002; Hasson et al. 2001). Ambiguous percepts provide elegant means of examining top-down processes without touching the stimulus material, as the process takes place in the mind of the participant. However, the visual approach implies several methodological difficulties as unintended rapid switches of perception, which make further modification of the stimuli, e.g. embossing, necessary (Andrews et al. 2002). By using auditive stimuli material, these obstacles can be bypassed. In literature, there was no evidence found for a previous study using Mondegreens/Soramimi for inducing auditory misperceptions. The behavioral study was designed as pre-study for the subsequent fMRI study. We were aiming to know more about the influence factors and gain a valid set of stimuli material. For this purpose, our trial participants listened to 61 song cutouts which were subtitled with an alternative lyric. In a final run without visual subtitles, the participants rated the strength of the experienced misperception.

3.1.1 Misperception of lyrics

We succeeded in altering perception in 42,6% (within language), and 33,5% (across language) of the trials (Beck et al., 2014). A later selection would be necessary to aim to reach an optimal 50 % induction, so the selection could be used in a classical 2x2 factorial design in a neuroimaging study. There were 20 German (for inducing within language misperception) and 41 English (for inducing across language misperception) song cutouts used, so they were not equally presented, which might be a weakness comparing the results. In a further study the amount of German and English cutouts should be equal. There was also a large variability in the induced misperception concerning stimuli material and participants. One can also criticize that 60 song cutouts are too many and participants get tired and thus have no optimal surroundings for learning and remembering. The stimulus material should be selected to achieve a more balanced, homogenous standard.

3.1.2 Familiarity with the lyrics

We examined several potential influence factors of the strength of misperceptions as the familiarity with the music. According to Pereira et al. (2011), familiarity has a strong impact on music's neural processing and appears to be an essential factor in making the listener emotionally engaged with music. On the basis of our data, we can report a close to zero correlation between the familiarity with the stimulus and the occurrence of a misperception and thus, there seem to be no potential biases of song familiarity expected in further studies (Beck et al., 2014).

3.1.3 Prior knowledge of alternative lyrics

By contrast, prior knowledge of possible alternate lyrics profoundly influenced the outcome of misperceptions (Beck et al., 2014). This is in line with reports from many of our trial participants stating that once induced, misperceptions can be persistent over time and can occur whenever the concerned piece of music is heard. It is also in line with findings from the visual domain where prior knowledge has massive influence on perception, as above described with Dolan et al.´s impoverished banana (1997). Hence, for neuroimaging studies dependent on a "before versus after induction" design, it would be favorable to know the participants' familiarity with possible alternate lyrics beforehand. Interestingly, prior knowledge of alternate lyrics did not across-the-board destine a person for misperception – although there are interindividual differences, our results suggest that the induction of misperceptions is not a pure individual phenomenon. Remembering that it was easier for the participants to misperceive within (German) than across (English) language, one could wonder if the prior knowledge explained the difference in frequency. This was not the case, as the participants knew more English than German misperceptions beforehand.

3.1.4 Wittiness of alternative lyrics

We found a significant correlation of the rated wittiness of the alternative lyrics with the intensity of misperceptions (Beck et al., 2014). It is well established, that stimuli of affective value have special impact. According to Anderson (2005), the emotional significance enhances the allocation of attentional resources. Emotion alters encoding (Schmitz, De Rosa & Anderson, 2009), fortifies perceptual vividness (Todd et al. 2012) and influences the long term memory (Cahill and McGaugh 1995). Related effects have been shown for musical stimuli. Listeners can not only recognize and perceive emotions in music, they also response to the music with the experience of emotions (Swaminathan and Schellenberg, 2015). Specific pieces of music can even evoke strong emotions and, maybe together with these emotions, the music can be remembered even years later (Eschrich, Münte & Altenmüller 2008). The aforementioned authors (2008) demonstrated that very positively rated musical excerpts were remembered better, and it has also been shown that listeners like music they remember and remember music they like (Stalinski & Schellenberg 2013). Additionally Aubé et al. (2013) demonstrated that emotional expressions transported by music, as well as faces and vocalizations, can influence memory. Consequently, we predicted that the occurrence of misperceptions is affected by emotion; more precisely, by humor appreciation which is the “affective experience of mirth” (Moran et al. 2004). In fact, the rated wittiness of the alternate lyrics is significantly correlated with the intensity of misperceptions. This agrees with findings from the visual domain, where solving visual ambiguities creates pleasant feelings just as solving a problem does (Ramachandran and Rogers-Ramachandran, 2007). In the words of the authors, “such paradoxical figures evoke wonder, delight and frustration at the same time—a microcosm of life itself” (q.v. 2007). Our findings confirm once more the importance of affective modulation in perceptual processes. Thus, further studies relating to perceptual phenomena should cover wittiness ratings for further investigation of the neuronal correlates underlying their modulation (Beck et al., 2014).

3.1.5 Linguistic competence

Ratings of misperceptions in within language trials correlated significantly with verbal intelligence in German (Beck et al., 2014). In other words: the larger the vocabulary of the participant, the easier it was to misperceive. This finding was in line with our hypothesis. Zareva (2007) did studies on the mental lexicon working with word association tests with native, advanced and intermediate speakers. He concludes that “the larger the vocabulary, the better connected it should be expected to be, especially in terms of number and variety of connections”. Our findings can be interpreted along similar lines: the larger the vocabulary, the easier it is for the participant to find another connection or word association than the original one. Also the across-language findings indicate similar results: there is a positive correlation between language competence in English and ratings of misperception, although these findings failed to reach significance (Beck et al., 2014) and thus should be interpreted with caution. It seems to contradict our a priori hypothesis that high language competence in English would be a protective factor against misperceptions. It also seems to contradict the proposition that mishearings are only driven by the listener´s inadequate language competence as proposed by Meyer and Ptok (2011) (Beck et al., 2014). The rate of misperceptions was higher within than across language. In other words, it was easier to misperceive from German to German, which could not be explained with prior knowledge (see above) nor with the plain wittiness ratings, as the participants rated the English misperceptions slightly funnier than the German ones. One possible interpretation might be that a German native speaker does naturally have a higher language competence and a larger mental lexicon in his native language and thus have more creative possibilities to choose from. The mishearings themselves were all in German, but perhaps the participants had better direct access to the native mental lexicon when listening to a German lyric instead of an English one. Yet another conclusion could be, that it was simply easier for a German speaker to remember the German alternate lyrics presented to them. More studies are needed to see if the results are reproducible.

3.2 Neurobiology of knowledge and misperception of lyrics

With the gained data out of the behavioral study, we were able to construct a balanced stimulus set. The 61 song cutouts (41 English and 20 German ones) were reduced to 36 song cut-outs (18 English and 18 German ones) usable in a 2x2 factorial design. That means, “before- versus after-induction” of the (mis)perception and “original versus alternative perception” can be examined within further neuroimaging studies.

In other words, a method for inducing misperceptions in a sufficiently predictable way was established and can be used within an fMRI study. We centered on finding out more about the neural correlates involved in knowledge and misperception of lyrics.

According to our hypothesis we found a linear relationship between brain activation and lyric-familiarity within the left hemispheric speech related areas, indicating them as important networks for the knowledge of lyrics (Beck Lidén et al., 2016). Whereas the misperception of lyrics showed activation in symmetrical areas in both hemispheres involving mSTS (middle superior temporal sulcus), MGB (medial geniculate body), IFG (inferior frontal gyrus), mdTh (medial dorsal thalamic nuclei) and SMA/dACC (supplemental motor area/dorsal anterior cingulate cortex) (Beck Lidén et al., 2016).

3.2.1 Neurobiology of knowledge of lyrics

There exist a number of neuropsychological and neuroimaging studies that investigated song processing. A “song system of the human brain” was proposed by Brown (2004) and a bihemispheric network is suggested for vocal production – interestingly without difference whether words are spoken or intoned (Özdemir et al., 2006). Linked to song familiarity, there are the findings of a study with a patient with severe expressive aphasia with an almost complete lesion of the left hemisphere, showing that the patient had a better word production in singing than in speaking, but only when dealing with familiar lyrics (Straube et al., 2008).

However, these studies investigated the production of songs rather than song perception. There exist several studies concerned with music perception and its general effects on brain activity. Comparisons between song and speech revealed that they have many features in common and in their perception cardinally similar brain regions are involved (Merrill et al., 2012). Brattico and collegues (2011) worked on musical emotions and used music with and without lyrics. Herholz et al. (2012) investigated the neuronal underpinnings of recognizing and encoding of heard and imagined music, and Sammler et al. (2010) elucidated the processing of lyrics and tunes of unfamiliar songs.

However, our study is the first study that examines the influence of lyric familiarity on the processing of music. There is a linear relationship between the individual ratings of familiarity and brain activity in mostly left hemispheric areas of the speech processing network (Hickok and Poeppel, 2000; 2007; Hickok, 2012). The activated areas include lateral and medial parts of the left frontal lobe and the Sylvian parietal temporal area (Beck Lidén et al., 2016) whose role in speech and vocal imagery has been largely documented. More specific, in line findings are connected to auditory retrieval, rehearsal and the imagery of vocals (Buchsbaum et al., 2005; Hickok et al., 2003) as well as covert humming (Pa and Hickok, 2008) and covert singing (internal rehearsal of tones) (Koelsch et al., 2009).

Moreover, activation patterns within the right superior cerebellum are in accordance with subcortico-cortical frameworks on speech perception (Kotz and Schwartze, 2010) as well as Price´s meta-analysis on word retrieval (Price, 2012). Very similar activation patterns as in our parametric approach were revealed in a categorical comparison between familiar and unfamiliar song excerpts (Janata, 2009). In addition, our findings saliently equal fMRI activations produced during spontaneous lyrical improvisation while practising freestyle rap (Beck Lidén et al., 2016; Liu et al., 2012). This points to the interpretation that the generation of lyrics during the creative process of freestyle rapping uses neural networks similar to those involved in the retrieval of lyrics, itself an important component to the process of recognising familiar lyrics.

Furthermore, earlier neuroimaging findings connected to the familiarity with song excerpts (Pereira et al., 2011) and instrumental music (Trost et al., 2012) were replicated with the midbrain and bilateral striatal activations found in lyric-familiarity (Beck Lidén et al., 2016). Interestingly, there is a clear discrepancy between our lyric-familiarity activated regions and those regions involved in adaptation designs where the same stimulus is repeatedly offered (Sammler et al., 2010). This could indicate that short-term adaptation processes recruit other neural networks than those involved with long-term effects such as lyric familiarity. (Beck Lidén et al., 2016).

As early as 1903 Meyer proposed a positive effect of familiarity on aesthetic judgment of music. In other words, mere repeated exposure towards a stimulus enhances an individual's attitude towards it (Zajonc, 1968). This effect is referred to as mere exposure effect and the relation between frequency and effect has been verified in many later studies (Peretz et al., 1998). Consistent neuroimaging findings for this effect in music listening were found and titled with "Listen! Learn! Like!" (Green et al., 2012). The effect has also been proven in motor familiarity framed in a study about dancing moves (Kirsch et al., 2015). Consequently, it appears plausible to propose that the familiarity with a stimulus recruits parts of the reward system during the different aspects of perceiving music (Beck Lidén et al., 2016). That can also explain why – besides the common aspect in the reward linked ventral striatum – the evaluation of familiarity is connected to specific tasks due to the nature of the stimulus. That is to say, there is activity during perceiving music in the anterior temporal cortex (Trost et al., 2012), during the evaluation of dancing moves in the motion linked cortex in the STS (Kirsch et al., 2015) and in the presented study the speech-related left hemispheric regions in the temporal and frontal lobe (Beck Lidén et al., 2016).

3.2.2 Neurobiology of misperception of lyrics

In the behavioral pre-study (for launching the new method) the rate of induced misperceptions was, as described earlier, at 34% for across language misperception (Soramimi) and 43% for within language misperceptions

(Mondegreens). As already described, the stimulus set was hand-selected from 61 to 36 song excerpts. In the fMRI-study, participants stated misperceptions in 41% of the across language and in 34% of the within language misperceptions. The most preferable induction rate of 50% was missed, but misperceptions can be induced in a sufficient way with the method. The difference between the studies regarding the misperception rates of Soramimi and Mondegreens can possibly be explained by the selection process of the stimuli material. The song excerpts for the Soramimi got selected from 41 down to 18 song excerpts and thus consisted of the funnier and consequently better working excerpts. That can explain the improvement of their misperception rate from 34% during the behavioral study to 41% during the fMRI study. The decline of the Mondegreen-misperception rate (43% to 34%) can be explained analogously: 18 out of 20 song excerpts got selected and one of the removed excerpts was very strong in misperception and wittiness. It got removed nevertheless for achieving a balanced set and can thus explain the decline.

In those trials with occurred misperceptions, the aim was to demonstrate which brain regions present differential activations. As a misperception implies a modification of speech-related information, an activation change within the components of the regular speech perception network seems plausible. The aim is not to identify a specific misperception network. A brain network which is specifically recruited for misperceptions does not appear biologically plausible as it would offer no advantage from an evolutionary perspective.

Differences in activation were detected in a mostly symmetrical network of temporal and frontal regions. This is in line with the dual-stream model of speech processing (Hickok and Poeppel, 2000; 2007; Hickok, 2012). More precisely, it is linked with the bilateral ventral stream associated with processing sound to meaning, in contrast to the left-lateralised dorsal stream where sound to action gets mapped. This was even emphasized when those trials with already earlier known misperceptions were excluded. It is thinkable, that those earlier known Mondegreens/Soramimi evoke less activation changes than the freshly induced ones, maybe due to habitation effects. In the presented study, this effect could not be further investigated systematically, as it only happened in a small number of trials and was unevenly spread over the participants.

Closer analysis of the temporal cortex revealed that secondary auditory regions along the mSTS showed more activation during misperceptions. This is in line with earlier fMRI results connected with processing and integration of lyrics and tunes of unfamiliar songs (Sammler et al., 2010) as well as the perceptual interpretation of ambiguous sounds (Kilian-Hutten et al., 2011). In the latter study, the purely subjective interpretation of an ambiguous sound can be decoded from the fMRI-measured activation patterns in the auditory areas of the superior temporal cortex. The present findings are also accordant with fMRI-findings from auditory perceptual grouping phenomena such as the continuity illusion (Heinrich et al., 2008; 2011). As early as 1950, Miller and Licklider discovered that discontinuous auditory signals may be perceived as continuous if the temporal gaps are filled with noise. This effect is called continuity illusion, temporal induction or phonemic restoration (Warren, 1970; Husain et al., 2005). The auditory system fills in the missing/blocked out parts for inducing continuity (Bidelman & Patro, 2016).

These earlier findings are mostly based on syllables or vowels as stimulus material. The present findings extend these simple sounds to the more complex stimulus material of misperceived lyrics and thus propose a more universal function of the mSTS in the processing of ambiguous auditory perceptions (Beck Lidén et al., 2016). This is accordant with clinical findings where damage in the middle temporal gyrus show auditory illusions and hallucinations (Wong et al., 2011).

In addition, activation patterns connected to misperception of lyrics include bilateral IFG, partly engaging Broca’s area and its homologue on the right hemisphere (Beck Lidén et al., 2016). The left sided results may be explained by the widely known function of Broca’s area within semantic processing (Bookheimer, 2002) and interleave with the activations found connected to the knowledge of lyrics, as described above. Activations on the right hemisphere have been shown earlier in the processing of pitch information, though there is evidence that the processing of pitch gets left-lateralized when pitch patterns are phonetically significant to the listener, as it is in tonal languages as Mandarin (Zatorre & Gandour, 2008). During the processing of pitch using the “octave illusion” as auditory ambiguous stimulus material, activation of the right IFG was found (Brancucci et al., 2011). Merrill (2012) studied on the differences of speech and song and found involvement of the right IFG in melody processing as well as Wildgruber (2005) linked the comprehension of emotional intonation (affective prosody) to a distinct network on the right hemisphere. The found bilateral activation of the IFG can indicate that the new semantic context of the misperceived lyric has to be integrated within the original melody and rhythm (Beck Lidén et al., 2016). This is in keeping with prior results showing widespread interactions during the musical as well as phonological processing of sung words within these bilateral areas (Schön et al., 2010).

Furthermore, the increase in activation of the SMA/dACC during misperception is accordant with the prior known role of that region as an “error detection and correction device” (Bush et al., 2002). This region is also named within studies about the processing of music in simple singing (Perry et al., 1999) as well as musical imagery (Halpern, 2001; Herholz et al., 2012). Anterior cingulate areas are also shown to be part of involuntary musical imagery (“earworms” aka song excerpts looping in your mind) (Farrugia et al., 2015). Areas including the SMA showed activation in a study associated with internal timekeeping (temporal imagery) (Schaefer, 2014). Activation in this area might point to the necessity of embedding the alternative lyric into the temporal pattern of the instrumental music (Beck Lidén et al., 2016). The found activation also correlates with how funny the participants rated the alterative lyrics (Beck Lidén et al., 2016). This is in keeping with prior findings, where sorting out visual ambiguities creates pleasant feelings analogue to problem-solving (Ramachandran & Rogers-Ramachandran, 2007) and concurs with Helmholtz´ long standing theory that perception is similar to the process of solving problems (von Helmholtz, 1867).

Furthermore, activation patterns are also solidifying prior research on this region linked to the highly pleasurable sensation of “chills” while perceiving music (Blood & Zatorre, 2001).

Lastly, differences in activation during misperception were also found in subcortical regions enclosing the mdTh and the MGB (Beck Lidén et al., 2016). The latter is part of the auditory thalamic nucleus (Møller, 2011) and not only a pure relay, but active modifying factor (“like a funhouse mirror”) in the auditory pathway between cochlea and cortex (Bartlett, 2013). Damage to this area can create auditory illusions as hyperacusis (sensibility to sounds) and palinacousis (a sort of auditory perseveration) (Fukutake & Hattori, 1998). The stronger reaction of the MGB during misperceptions might mirror modulatory auditory feedback coming via reciprocal connections (Pickles & O., 2015).

The mdTh works as a thalamic relay of higher order contributing to learning and decision-making as reviewed by Mitchell (2015). If damaged, thalamic amnesia (Mair et al., 2015), deficits in processing new information (Mitchell & Chakraborty, 2013) and language disturbances (Radanovic & Scaff, 2003; for a review, see Barbas et al., 2013). Thus, the region might be involved in the acquisition of new auditory information and might serve as a hub for connections with the prefrontal cortex (Beck Lidén et al., 2016).

3.2.3 Limitations

The findings have to be discussed considering the limitations of the used response-related fMRI models: For avoiding inducing mere exposure effects (Peretz et al., 1998) as mentioned above, the participants rated the familiarity during the first exposure with the stimulus in the scanner. This has the disadvantage that the patient could have gotten a certain reward out of demonstrating the knowledge of lyrics towards the experimenters and this could offer an alternative explication for the activation patterns within the reward system. Further neuroimaging experiments should take this into account by acquiring the familiarity ratings outside the scanner to avoid this effect (Beck Lidén et al., 2016). Furthermore, the participants were given explicit instruction to estimate their familiarity with the offered lyrics and not, for example, their familiarity with the music or other factors of the song excerpts. The activation pattern of lyric familiarity was strikingly similar to the generation of lyrics (during freestyle rapping) (Liu et al., 2012). This emphasizes the point, that it was the familiarity with the lyrics being rated by the participants. However, for fully detangling the relationship between the lyric familiarity versus the familiarity with the musical structure, an orthogonal set of stimuli is needed where the participant is familiar with the lyrics, but not the musical structure, and the other way round (Beck Lidén et al., 2016). Designing and obtaining such a stimulus set might be challenging due to strong variability between the participants.

Concerning the misperceptions, within- (Mondegreens) as well as across-language (Soramimi) misperceptions were induced. Analyses could not show any significant differences in activation between these stimulus types. This negative finding implies that similar neural networks are used regardless of whether language switch occurs or not. However, smaller differences in activation might be detected in future research, for example in models using more stimuli.

It also has to be noted that the used method for externally inducing misperceptions is based on expectation-related alterations of perception. One might argue that this is not fully comparable with spontaneous misperceptions, but even those are often analogically created by expectations, though those are internally (and not externally) generated.

3.3 Conclusions

In the presented studies, a novel method to externally induce auditory misperceptions based on ambiguous song excerpts (Mondegreens/Soramimi) was introduced. Lyrics within these song excerpts were used, because they offer as stimuli the right factors for promoting misperceptions: unusual (sung) pronunciation plus for a correct understanding unfavourable additional acoustic signals (background music). The natural potential for misperceptions of the used song excerpts enabled investigating top-down phenomena without changing any bottom-up parameters as the physical properties of the stimuli.

Misperceptions occurred with a sufficient probability. The familiarity of the song is irrelevant, which is a benefit in future studies, as the participant´s knowledge of the song/music is not relevant. The prior knowledge of alternative lyrics has, on the contrary, a strong influence. It would be interesting to further investigate the effects and dynamics of Mondegreens/Soramimi in relation to time. For example, the participants out of this study could participate at a follow-up study creating a pool with individuals who learned the misperceptions more or less around the same time compared to naïve participants and also compared to obtained data of the presented studies. That would give more sound insight of the effect reported by several participants and even testing personnel that the misperceptions seem to stick in mind.

Verbal fluency increased the strength of misperception suggesting that the size of vocabulary serves as a source for possible interpretations (Beck et al., 2014). Wittiness of the misperceived song excerpts also increased the strength of misperceptions. This indicates that the alternative percept was experienced as pleasant and resolving the offered ambiguity is connected with a joyful feeling (Beck et al., 2014). This stands in accordance with Helmholtz´s theories (1867) suggesting that perception is based on comparable processes as intellectual problem solving.

The method came to use within an fMRI study. The familiarity of lyrics was connected with activation in reward-associated regions and left-hemispheric speech areas. Thinking about the example from the introduction with Michael Jackson, which you might have heard singing, we know now which parts of the brain get active when you hear him mistakenly singing in German. The misperception of lyrics revealed a symmetric network over both hemispheres in frontotemporal cortex and thalamus (Beck Lidén et al., 2016). These findings provide for the first time a better understanding of the neural substrates underlying misperception of lyrics which arise in the ventral stream (sound to meaning) of the common speech perception network.

Misperceptions of song lyrics are harmless and can even be a source of amusement, but the results out of these studies can serve as a neurobiological model for other types of misunderstandings, which can represent events with potentially strong impact on human social life.

4 Summary

Expectations and previous knowledge can severely influence our perception. This so called top-down modulation in the processing of perception has been largely investigated in the visual modality, using ambiguous stimuli material such as reversible figures. We launch a novel method to study this issue in the auditory field. The method is based on Mondegreens and Soramimi: song lyrics which can be misperceived within one language or through switching between two languages. We proved through a behavioral experiment that misperceptions can be externally induced by visual presentation of the alternate percept as subtitles. Misperceptions appear with a sufficient probability for further use in neuroscientific experiments. The method comes into effect in a subsequent study including two functional magnetic resonance imaging (fMRI) experiments to fathom the neural correlates of knowledge and misperception of lyrics.

The familiarity of the song is irrelevant, but the prior knowledge of alternative lyrics has a strong influence on the strength of misperception. Verbal fluency, seen as source for possible interpretations, as well as wittiness of the alternate perception increased the strength of misperception. This is in accordance with Hermann von Helmholtz´s theories that perception is based on comparable processes as problem solving.

The first fMRI experiment revealed a linear correlation between familiarity of lyrics and activation in left-hemispheric speech-associated as well as bilateral striatal regions which is accordant with prior research on lyric-generation. The second fMRI experiment elucidated a bilateral network connected to the misperception of lyrics. This network included middle temporal and inferior frontal regions as well as anterior cingulate cortex (ACC) and medio-dorsal thalamus Activation in the ACC is also related to how amusing participants rated the misperceptions. This is solidifying prior research on this region linked to the pleasurable sensation of chills while listening to music.

Misperceptions of song lyrics are mostly harmless, but our results could serve as a more general neurobiological model for additional studies on other types of misunderstandings which can present events with potentially powerful impact on our social life (Beck et al., 2014; Beck Lidén et al., 2016).

5 Deutsche Zusammenfassung (German Summary)

Erwartungen und vorheriges Wissen können unsere Wahrnehmung beeinflussen. Diese so genannte top-down Modulation in der Verarbeitung von Wahrnehmung wurde im visuellen Bereich, besonders unter Zuhilfenahme von zweideutigem Stimulusmaterial wie etwa Kippfiguren, ausgiebig untersucht. Wir lancierten eine neue Methode, um diese Thematik im auditiven Bereich zu beleuchten. Die Methode basierte auf so genannten Mondegreens und Soramimi. Das sind Liedtexte, die innerhalb einer Sprache oder bei Wechsel in eine andere Sprache, missverstanden werden können. Durch ein Verhaltensexperiment wurde gezeigt, dass Fehlwahrnehmungen („Verhörer") von außen erzeugt werden können. Dazu wurden den Probanden die Verhörer mit Hilfe von Untertiteln zeitgleich zum gehörten Lied visuell präsentiert. Fehlwahrnehmungen traten mit einer ausreichenden Wahrscheinlichkeit auf, um in weiteren neurowissenschaftlichen Experimenten zur Anwendung zu kommen. Auf der ersten Studie basierte eine Folgestudie, die zwei fMRI (funktionelle Kernspintomographie) Experimente beinhaltete. Die Folgestudie ergründete die neuronalen Korrelate von Kenntnis und Fehlwahrnehmung von Liedtexten.

Die Bekanntheit des Liedes war nicht von Bedeutung, doch die vorherige Kenntnis des alternativen Liedtextes (Verhöres) hatte einen ausgeprägten Einfluss auf die Stärke der Fehlwahrnehmung. Verbale Sprachkompetenz – als Quelle für mögliche alternative Interpretationen angesehen – genauso wie die Witzigkeit des Verhörers, steigerte die Stärke der Fehlwahrnehmung. Dies stimmt überein mit den Theorien von Hermann von Helmholtz, wonach Wahrnehmung auf vergleichbaren Prozessen basiert wie das Lösen von Problemen.

Das erste fMRI Experiment ergab eine lineare Korrelation zwischen der Bekanntheit der Lieder und der Aktivierung in linkshemisphärischen sprachassoziierten und beidseitigen striatalen Arealen. Dies deckt sich mit früherer Forschung zum Thema Erzeugung von Liedtexten. Das zweite fMRI Experiment zur Fehlwahrnehmung von Liedtexte zeigte eine beidseitige Aktivierung von Gehirnarealen in Form eines Netzwerkes. Dieses beinhaltete mittlere temporale und inferiore frontale Arealen, genauso wie den anterioren cingulären Cortex

(ACC) und den medio-dorsalen Thalamus. Die Aktivierung im ACC erfolgte auch in Relation dazu, wie unterhaltsam die Probanden einen Verhörer bewertet hatten. Das bestätigte frühere Forschungsarbeiten zu diesem Areal, das ebenfalls aktiviert wurde, wenn beim Musikhören angenehme Empfindungen („wohlige Schauer“) erlebt werden.

Während das Verhören von Liedtexten zumeist harmlos ist, können Missverständnisse Ereignisse darstellen, die einen entscheidenden Einfluss auf unser soziales Leben haben. Mit Hilfe von Verhörern in Liedtexten und unseren Ergebnissen konnte ein allgemeineres, neurobiologisches Modell geschaffen werden für künftige Studien zur Erforschung von Missverständnissen (Beck et al., 2014; Beck Lidén et al., 2016).

6 Svensk sammanfattning (Swedish summary)

Förväntningar och tidigare kunskaper kan påtaglig påverka vår perception. Denna så kallade top-down (uppifrån och ned) moduleringen i bearbetningen av perceptioner har tidigare huvudsakligen undersökts inom det visuella området med hjälp av tvetydigt stimuli material så som ”dubbeltydiga bilder”. Vi presenterar en ny metod för att studera detta inom hörselområdet. Denna metod baserar sig på Mondegreens och Soramimi vilket är sångtexter som kan missuppfattas inom samma språk eller förväxlas mellan två språk. Vi bevisade genom en beteendestudie att missuppfattningar kan utlösas utifrån genom visuell presentation av alternativa uppfattningar i form av undertexter.

Missuppfattningar uppkommer med tillräcklig statistisk sannolikhet för att kunna använda detta i ytterligare neurovetenskapliga experiment. Metoden används i en efterföljande studie med två experiment med funktionell magnetkameraundersökning (fMR) för att förstå de neurala korrelaten till kunskap om och missuppfattningar i sångtexter.

Hur förtrogen man är med en sång är irrelevant, men tidigare kunskap om den alternativa texten har en stark påverkan på styrkan av missuppfattningen. Verbal språkkompetens – en möjlig källa till alternativa intryck – liksom vitsigheten av en alternativ uppfattning ökar styrkan av missuppfattningen. Detta stämmer väl med Hermann von Helmholtz teorier att perceptionen är baserad på processer som är jämförbara med problemlösning.

Det första fMR-experimentet avslöjade en linjär korrelation mellan förtrogenheten av sångtexten och aktivering av vänstra hjärnhalvans språkrelaterade område liksom bilaterala striatala regioner. Detta överensstämmer med tidigare forskning kring aktiv produktion av sångtexter. Det andra fMR-experimentet upptäckte ett bilateralt nätverk som är associerat med missuppfattning av sångtexter. Nätverket inkluderade mitt-temporala och inferiort frontala regioner samt anteriora cingulära kortex (ACC) och medio-dorsala thalamus. Aktivering i ACC var också relaterad med hur roligt den medverkande skattade missuppfattningen. Detta konsoliderar tidigare forskning om att denna region är associerad med rysningar under det att man lyssnar på musik.

Missuppfattningar av sångtexter är huvudsakligen harmlöst, men våra resultat kan användas som en mer generell neurobiologisk modell för andra studier kring missuppfattningar vilket kan ha potentiell kraftig påverkan på vårt sociala liv (Beck et al., 2014; Beck Lidén et al., 2016).

7 References

Aitchison, J. (2012). *Words in the Mind: An Introduction to the Mental Lexicon* (Fourth). Malden, USA: John Wiley & Sons.

Anderson, A. K. (2005). Affective influences on the attentional dynamics supporting awareness. *Journal of Experimental Psychology: General*, *134*(2), 258–281. https://doi.org/10.1037/0096-3445.134.2.258

Andrews, T. J., Schluppeck, D., Homfray, D., Matthews, P., & Blakemore, C. (2002). Activity in the Fusiform Gyrus Predicts Conscious Perception of Rubin's Vase–Face Illusion. *NeuroImage*, *17*(2), 890–901. https://doi.org/10.1006/nimg.2002.1243

Atlas, L. Y., Bolger, N., Lindquist, M. A., & Wager, T. D. (2010). Brain mediators of predictive cue effects on perceived pain. *Journal of Neuroscience*, *30*(39), 12964–12977.

Aubé, W., Peretz, I., & Armony, J. L. (2013). The effects of emotion on memory for music and vocalisations. *Memory (Hove, England)*, *21*(8), 981–90. https://doi.org/10.1080/09658211.2013.770871

Barbas, H., García-Cabezas, M. Á., & Zikopoulos, B. (2013). Frontal-thalamic circuits associated with language. *Brain and Language*, *126*(1), 49–61. https://doi.org/10.1016/j.bandl.2012.10.001

Bartlett, E. L. (2013). The organization and physiology of the auditory thalamus and its role in processing acoustic features important for speech perception. *Brain and Language*, *126*(1), 29–48. https://doi.org/10.1016/j.bandl.2013.03.003

Beck, C., Kardatzki, B., & Ethofer, T. (2014). Mondegreens and Soramimi as a method to induce misperceptions of speech content - influence of familiarity, wittiness, and language competence. *PLoS One*, *9*(1), e84667. https://doi.org/10.1371/journal.pone.0084667

Beck Lidén, C., Krüger, O., Schwarz, L., Erb, M., Kardatzki, B., Scheffler, K., & Ethofer, T. (2016). Neurobiology of knowledge and misperception of lyrics. *NeuroImage*, *134*, 12–21. https://doi.org/10.1016/j.neuroimage.2016.03.080

Benedetti, F., Mayberg, H. S., Wager, T. D., Stohler, C. S., & Zubieta, J.-K. (2005). Neurobiological Mechanisms of the Placebo Effect. *Journal of Neuroscience*, *25*(45), 10390–10402. https://doi.org/10.1523/JNEUROSCI.3458-05.2005

Bidelman, G. M., & Patro, C. (2016). Auditory perceptual restoration and illusory continuity correlates in the human brainstem. *Brain Research*, *1646*, 84–90. https://doi.org/10.1016/J.BRAINRES.2016.05.050

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(20), 11818–11823. https://doi.org/10.1073/pnas.191355898

Bond, Z. S. (1999). Morphological errors in casual conversation. *Brain and Language*, *68*(1–2), 144–50. https://doi.org/10.1006/brln.1999.2071

Bond, Z. S. (2005). Slips of the ear. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 290–310). Oxford, UK: Blackwell Publishing Ltd.

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, *25*(1), 151–188. https://doi.org/10.1146/annurev.neuro.25.112701.142946

Brancucci, A., Franciotti, R., D'Anselmo, A., Della Penna, S., & Tommasi, L. (2011). The Sound of Consciousness: Neural Underpinnings of Auditory Perception. *Journal of Neuroscience*, *31*(46), 16611–16618. https://doi.org/10.1523/JNEUROSCI.3949-11.2011

Brattico, E., Alluri, V., Bogert, B., Jacobsen, T., Vartiainen, N., Nieminen, S., & Tervaniemi, M. (2011). A Functional MRI Study of Happy and Sad Emotions in Music with and without Lyrics. *Frontiers in Psychology*, *2*, 308. https://doi.org/10.3389/fpsyg.2011.00308

Brown, S., Martinez, M. J., Hodges, D. A., Fox, P. T., & Parsons, L. M. (2004). The song system of the human brain. *Cognitive Brain Research*, *20*(3), 363–75. https://doi.org/10.1016/j.cogbrainres.2004.03.016

Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human Dorsal and Ventral Auditory Streams Subserve Rehearsal-Based and Echoic Processes during Verbal Working Memory. *Neuron*, *48*, 687–697. https://doi.org/10.1016/j.neuron.2005.09.029

Bush, G., Vogt, B. A., Holmes, J., Dale, A. M., Greve, D., Jenike, M. A., & Rosen, B. R. (2002). Dorsal anterior cingulate cortex: a role in reward-based decision making. *Proceedings of the National Academy of Sciences*, *99*(1), 523–528. https://doi.org/10.1073/pnas.012470999

Cahill, L., & McGaugh, J. L. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and Cognition*, *4*(4), 410–421.

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz Machine. *Neural Computation*, *7*(5), 889–904. https://doi.org/10.1162/neco.1995.7.5.889

Deutsch, D. (1974). An auditory illusion. *Nature*, *251*(5473), 307–309. https://doi.org/10.1121/1.1919587

Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *The Journal of the Acoustical Society of America*, *129*(4), 2245–52. https://doi.org/10.1121/1.3562174

Dolan, R. J., Fink, G. R., Rolls, E., Booth, M., Holmes, A., Frackowiak, R. S., & Friston, K. J. (1997). How the brain learns to see objects and faces in an impoverished context. *Nature*, *389*(6651), 596–599.

Eschrich, S., Münte, T. F., & Altenmüller, E. O. (2008). Unforgettable film music: The role of emotion in episodic long-term memory for music. *BMC Neuroscience*, *9*(1), 48. https://doi.org/10.1186/1471-2202-9-48

Ethofer, T. (2011). *Functional and structural networks for bottom-up and top-down processing of emotional prosody*. Medizinische Habilitationsschrift, Universität Tübingen.

Ethofer, T., Anders, S., Erb, M., Droll, C., Royen, L., Saur, R., … Wildgruber, D. (2006). Impact of voice on emotional judgment of faces: An event-related fMRI study. *Human Brain Mapping*, *27*(9), 707–714.

Farrugia, N., Jakubowski, K., Cusack, R., & Stewart, L. (2015). Tunes stuck in your brain: The frequency and affective evaluation of involuntary musical imagery correlate with cortical structure. *Consciousness and Cognition*, *35*, 66–77. https://doi.org/10.1016/j.concog.2015.04.020

Friedenberg, J., & Silverman, G. (2006). *Cognitive science : an introduction to the study of mind*. Sage Publications.

Frith, C., & Dolan, R. J. (1997). Brain mechanisms associated with top-down processes in perception. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *352*(1358), 1221–30. https://doi.org/10.1098/rstb.1997.0104

Fukutake, T., & Hattori, T. (1998). Auditory illusions caused by a small lesion in the right medial geniculate body. *Neurology*, *51*(5), 1469–1471.

Girgus, J. J., Rock, I., & Egatz, R. (1977). The effect of knowledge of reversibility on the reversibility of ambiguous figures. *Perception & Psychophysics*, *22*(6), 550–556. https://doi.org/10.3758/BF03198762

Gosseries, O., Demertzi, A., Noirhomme, Q., Tshibanda, J., Boly, M., Op de Beeck, M., … De Tiege, X. (2008). [Functional neuroimaging (fMRI, PET and MEG): what do we measure?]. *Revue Medicale de Liege*, *63*(5–6), 231–237.

Green, A. C., Bærentsen, K. B., Stodkilde-Jorgensen, H., Roepstorff, A., Vuust, P., Baerentsen, K. B., … Vuust, P. (2012). Listen, learn, like! Dorsolateral prefrontal cortex involved in the mere exposure effect in music. *Neurology Research International*, *2012*, 846270. https://doi.org/10.1155/2012/846270

Halpern, A. R. (2001). Cerebral substrates of musical imagery. *Annals of the New York Academy of Sciences*, *930*, 179–92.

Hasson, U., Hendler, T., Ben Bashat, D., & Malach, R. (2001). Vase or face? A neural correlate of shape-selective grouping processes in the human brain. *Journal of Cognitive Neuroscience*, *13*(6), 744–753.

Heinrich, A., Carlyon, R. P., Davis, M. H., & Johnsrude, I. S. (2008). Illusory vowels resulting from perceptual continuity: a functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *20*(10), 1737–1752. https://doi.org/10.1162/jocn.2008.20069

Heinrich, A., Carlyon, R. P., Davis, M. H., & Johnsrude, I. S. (2011). The continuity illusion does not depend on attentional state: FMRI evidence from illusory vowels. *Journal of Cognitive Neuroscience*, *23*(10), 2675–2689. https://doi.org/10.1162/jocn.2011.21627

Hengeler, A. (1892). Fliegende Blätter (97.1892 (Nr. 2449-2474)). https://doi.org/10.11588/diglit.2137#0147

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience*, *24*(6), 1382–1397. https://doi.org/10.1162/jocn_a_00216

Hickok, G. (2012). Journal of Communication Disorders The cortical organization of speech processing : Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, *45*(6), 393–402. https://doi.org/10.1016/j.jcomdis.2012.06.004

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory–Motor Interaction Revealed by fMRI: Speech, Music, and Working Memory in Area Spt. *Journal of Cognitive Neuroscience*, *15*(5), 673–682. https://doi.org/10.1162/089892903322307393

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*. https://doi.org/10.1016/S1364-6613(00)01463-7

Hornby, A. S. (1995). *Oxford Advanced Learner's Dictionary*. (J. Crowther, Ed.) (fifth). Oxford, UK: Oxford University Press.

Horowitz, A. L. (1995). *MRI physics for radiologists* (third). New York: Springer-Verlag. https://doi.org/10.1007/978-1-4684-0333-6

Husain, F. T., Lozito, T. P., Ulloa, A., & Horwitz, B. (2005). Investigating the Neural Basis of the Auditory Continuity Illusion. *Journal of Cognitive Neuroscience*, *17*(8), 1275–1292. https://doi.org/10.1162/0898929055002472

Jäkel, F., Singh, M., Wichmann, F. A., & Herzog, M. H. (2016). An overview of quantitative approaches in Gestalt perception. *Vision Research*, *126*, 3–8. https://doi.org/10.1016/j.visres.2016.06.004

Janata, P. (2009). The neural architecture of music-evoked autobiographical memories. *Cerebral Cortex*, *19*(11), 2579–2594. https://doi.org/10.1093/cercor/bhp008

Kilian-Hutten, N., Valente, G., Vroomen, J., & Formisano, E. (2011). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *Journal of Neuroscience*, *31*(5), 1715–1720. https://doi.org/10.1523/JNEUROSCI.4572-10.2011

Kirsch, L. P., Dawson, K., & Cross, E. S. (2015). Dance experience sculpts aesthetic perception and related brain circuits. *Annals of the New York Academy of Sciences*, *1337*(1), 130–139. https://doi.org/10.1111/nyas.12634

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: An FMRI study. *Human Brain Mapping*, *30*(3), 859–873. https://doi.org/10.1002/hbm.20550

Koffka, K. (1935). *Principles of Gestalt psychology [E-book]*. Oxon: Routledge. Retrieved Mai 17, 2018, from https://books.google.se/books?id=tWxHAQAAQBAJ&printsec=frontcover&hl=sv&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false.

Kogo, N., Hermans, L., Stuer, D., van Ee, R., & Wagemans, J. (2015). Temporal dynamics of different cases of bi-stable figure-ground perception. *Vision Research*, *106*, 7–19. https://doi.org/10.1016/j.visres.2014.10.029

Köhler, W. (1940). *Dynamics in Psychology*. New York: Liveright Publishing Corporation.

Kotz, S. A., & Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. *Trends in Cognitive Sciences*, *14*(9), 392–399. https://doi.org/10.1016/j.tics.2010.06.005

Liu, S., Chow, H. M., Xu, Y., Erkkinen, M. G., Swett, K. E., Eagle, M. W., … Braun, A. R. (2012). Neural correlates of lyrical improvisation: an FMRI study of freestyle rap. *Scientific Reports*, *2*, 834. https://doi.org/10.1038/srep00834

Long, G. M., Toppino, T. C., & Mondin, G. W. (1992). Prime time: fatigue and set effects in the perception of reversible figures. *Perception & Psychophysics*, *52*(6), 609–616. https://doi.org/10.3758/BF03211697

Mair, R. G., Miller, R. L. A., Wormwood, B. A., Francoeur, M. J., Onos, K. D., & Gibson., B. M. (2015). The neurobiology of thalamic amnesia: Contributions of medial thalamus and prefrontal cortex to delayed conditional discrimination. *Neuroscience & Biobehavioral Reviews*, *54*, 161–174. https://doi.org/10.1016/j.neubiorev.2015.01.011

Marini, F., & Marzi, C. A. (2016). Gestalt Perceptual Organization of Visual Stimuli Captures Attention Automatically: Electrophysiological Evidence. *Frontiers in Human Neuroscience*, *10*(August), 1–10. https://doi.org/10.3389/fnhum.2016.00446

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*(1), 3–33. https://doi.org/10.1037/0033-295X.101.1.3

MCGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Melloni, L., Schwiedrzik, C. M., Müller, N., Rodriguez, E., & Singer, W. (2011). Expectations change the signatures and timing of electrophysiological correlates of perceptual awareness. *Journal of Neuroscience*, *31*(4), 1386–1396. https://doi.org/10.1523/JNEUROSCI.4570-10.2011

Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., & Friederici, A. D. (2012). Perception of words and pitch patterns in song and speech. *Frontiers in Psychology*, *3*, 76. https://doi.org/10.3389/fpsyg.2012.00076

Meyer, M. (1903). Experimental Studies in the Psychology of Music. *The American Journal of Psychology*, *14*(3/4), 192. https://doi.org/10.2307/1412315

Meyer, S., & Ptok, M. (2011). A phenomenon called Mondegreen. *HNO*, *59*(9), 926–930. https://doi.org/10.1007/s00106-011-2335-7

Miller, G. A., & Licklider, J. C. R. (1950). The Intelligibility of Interrupted Speech. *The Journal of the Acoustical Society of America*, *22*(2), 167–173. https://doi.org/10.1121/1.1906584

Mitchell, A. S. (2015). The mediodorsal thalamus as a higher order thalamic relay nucleus important for learning and decision-making. *Neuroscience and Biobehavioral Reviews*, *54*, 76–88. https://doi.org/10.1016/j.neubiorev.2015.03.001

Mitchell, A. S., & Chakraborty, S. (2013). What does the mediodorsal thalamus do? *Frontiers in Systems Neuroscience*, *7*, 37. https://doi.org/10.3389/fnsys.2013.00037

Møller, A. R. (2011). Anatomy and Physiology of the Auditory System. In *Textbook of Tinnitus* (pp. 51–68). New York, NY: Springer New York. https://doi.org/10.1007/978-1-60761-145-5_8

Moran, J. M., Wig, G. S., Adams, R. B., Janata, P., & Kelley, W. M. (2004). Neural correlates of humor detection and appreciation. *NeuroImage*, *21*, 1055–1060. https://doi.org/10.1016/j.neuroimage.2003.10.017

Mumford, D. (1992). On the computational architecture of the neocortex - II The role of cortico-cortical loops. *Biological Cybernetics*, *66*(3), 241–251. https://doi.org/10.1007/BF00198477

Nawrot, M., & Blake, R. (1989). Neural integration of information specifying structure from stereopsis and motion. *Science*, *244*(4905), 716–718. https://doi.org/10.1126/science.2717948

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(24), 9868–72. https://doi.org/10.1073/pnas.87.24.9868

Otake, T. (2007). Interlingual Near Homophonic Words and Phrases in L2 Listening : Evidence From Misheard Song Lyrics. *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, (August), 777–780.

Özdemir, E., Norton, A., & Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *NeuroImage*, *33*(2), 628–635. https://doi.org/10.1016/j.neuroimage.2006.07.013

Pa, J., & Hickok, G. (2008). A parietal-temporal sensory-motor integration area for the human vocal tract: Evidence from an fMRI study of skilled musicians. *Neuropsychologia*, *46*(1), 362–368. https://doi.org/10.1016/j.neuropsychologia.2007.06.024

Pauling, L., & Coryell, C. D. (1936). The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin. *Proceedings of the National Academy of Sciences*, *22*(4), 210–216. https://doi.org/10.1073/pnas.22.4.210

Peretz, I., Gaudreau, D., & Bonnel, A.-M. (1998). Exposure effects on music preference and reognition. *Memory & Cognition*, *26*(5), 884–902. https://doi.org/10.3758/BF03201171

Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1999). Localization of cerebral activity during simple singing. *NeuroReport*, *10*(18), 3979–3984. https://doi.org/10.1097/00001756-199912160-00046

Picard, F., & Friston, K. (2014). Predictions, perception, and a sense of self. *Neurology*, 1–7. https://doi.org/10.1212/WNL.0000000000000798

Pickles, J. O., & O., J. (2015). Auditory pathways: Anatomy and physiology. In G. G. Celesia & G. Hickok (Eds.), *Handbook of Clinical Neurology* (Vol. 129, pp. 3–25). Elsevier B.V. https://doi.org/10.1016/B978-0-444-62630-1.00001-9

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8), 13. https://doi.org/10.1167/15.8.13

Plato. (n.d.). *Charmides*.

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–47. https://doi.org/10.1016/j.neuroimage.2012.04.062

Radanovic, M., & Scaff, M. (2003). Speech and language disturbances due to subcortical lesions. *Brain and Language*, *84*(3), 337–352. https://doi.org/10.1016/S0093-934X(02)00554-0

Ramachandran, V. S., & Rogers-Ramachandran, D. (2007). Ambiguities and Perception. *Scientific American Mind*, *18*(5), 18–20.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Rock, I., Hall, S., & Davis, J. (1994). Why do ambiguous figures reverse? *Acta Psychologica*, *87*(1), 33–59.

Rubin, E. (1921). *Visuell wahrgenommene Figuren [E-book]*. Copenhagen: Gyldenalske Boghandel. Retrieved August 26, 2015, from https://ia700401.us.archive.org/0/items/visuellwahrgenom01rubiuoft/visuellwahrgenom01rubiuoft.pdf.

Sammler, D., Baird, A., Valabregue, R., Clement, S., Dupont, S., Belin, P., & Samson, S. (2010). The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. *Journal of Neuroscience*, *30*(10), 3572–3578. https://doi.org/10.1523/JNEUROSCI.2751-09.2010

Schaefer, R. S. (2014). Images of time: temporal aspects of auditory and movement imagination. *Frontiers in Psychology*, *5*, 877. https://doi.org/10.3389/fpsyg.2014.00877

Schmitz, T. W., De Rosa, E., & Anderson, A. K. (2009). Opposing Influences of Affective State Valence on Visual Cortical Encoding. *Journal of Neuroscience*, *29*(22), 7199–7207. https://doi.org/10.1523/JNEUROSCI.5387-08.2009

Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J.-L., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *NeuroImage*, *51*(1), 450–461. https://doi.org/10.1016/J.NEUROIMAGE.2010.02.023

Silva Pereira, C., Teixeira, J., Figueiredo, P., Xavier, J., Luís Castro, S., & Brattico, E. (2011). Music and Emotions in the Brain: Familiarity Matters. *PLoS ONE*, *6*(11). https://doi.org/10.1371/journal.pone.0027241

Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(12), E1747-56. https://doi.org/10.1073/pnas.1523266113

Spivey, M. J., & Marian, V. (1999). Cross Talk Between Native and Second Languages: Partial Activation of an Irrelevant Lexicon. *Psychological Science*, *10*(3), 281–284. https://doi.org/10.1111/1467-9280.00151

Stalinski, S. M., & Schellenberg, E. G. (2013). Listeners remember music they like. *J Exp Psychol Learn Mem Cogn*, *39*(3), 700–716. https://doi.org/10.1037/a0029671

Sternberg, J. R., & Sternberg, K. (2011). *Cognitive Psychology*. (K. S. Robert J. Sternberg, Ed.), *Science* (sixth). Wadsworth: Wadsworth Publishing Co Inc. https://doi.org/10.1126/science.198.4319.816

Straube, T., Schulz, A., Geipel, K., Mentzel, H.-J., & Miltner, W. H. R. (2008). Dissociation between singing and speaking in expressive aphasia: The role of song familiarity. *Neuropsychologia*, *46*(5), 1505–1512. https://doi.org/10.1016/j.neuropsychologia.2008.01.008

Strüber, D., & Stadler, M. (1999). Differences in top - Down influences on the reversal rate of different categories of reversible figures. *Perception*, *28*(10), 1185–1196. https://doi.org/10.1068/p2973

Swaminathan, S., & Schellenberg, E. G. (2015). Current Emotion Research in Music Psychology. *Emotion Review*, *7*(2), 189–197. https://doi.org/10.1177/1754073914558282

Tierney, A., Dick, F., Deutsch, D., & Sereno, M. (2013). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cerebral Cortex*, *23*(2), 249–254. https://doi.org/10.1093/cercor/bhs003

Todd, R. M., Talmi, D., Schmitz, T. W., Susskind, J., & Anderson, A. K. (2012). Psychophysical and Neural Evidence for Emotion-Enhanced Perceptual Vividness. *Journal of Neuroscience*, *32*(33), 11201–11212. https://doi.org/10.1523/JNEUROSCI.0155-12.2012

Toppino, T. C., & Long, G. M. (2005). Top-down and bottom-up processes in the perception of reversible figures: Toward a hybrid model. In *Dynamic Cognitive Processes* (pp. 37–58). Springer Tokyo.

Trost, W., Ethofer, T., Zentner, M., & Vuilleumier, P. (2012). Mapping aesthetic musical emotions in the brain. *Cerebral Cortex*, *22*(12), 2769–2783. https://doi.org/10.1093/cercor/bhr353

Vetter, P., Sanders, L. L. O., & Muckli, L. (2014). Dissociation of Prediction from Conscious Perception. *Perception*, *43*(10), 1107–1113. https://doi.org/10.1068/p7766

von Ehrenfels, C. (1890). Ueber “Gestaltqualitäten.” *Vierteljahrsschrift Für Wissenschaftliche Philosophie*, 249–292. Retrieved from http://gallica.bnf.fr/ark:/12148/bpt6k94141n/f255.image

von Grünau, M. W., Wiggin, S., & Reed, M. (1984). The local character of perspective organization. *Perception & Psychophysics*, *35*(4), 319–324. https://doi.org/10.3758/BF03206335

von Helmholtz, H. (1867). *Handbuch der physiologischen Optik [E–book]*. Leipzig: Leopold Voss. Retrieved Mai 04, 2017, from https://archive.org/stream/handbuchderphysi00helm#page/n15/mode/2up

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, *138*(6), 1172–217. https://doi.org/10.1037/a0029333

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392–3.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. In K. Koffka, W. Köhler, M. Wertheimer, K. Goldstein, & H. Gruhle (Eds.), *Psychologische Forschung - Zeitschrift fuer Psychologie und ihre Grenzwissenschaften* (pp. 301–350). Berlin: Julius Springer. Retrieved March 30, 2017, from http://gestalttheory.net/download/Wertheimer1923_Lehre_von_der_Gestalt.pdf.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., & Ackermann, H. (2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage*, *24*(4), 1233–1241. https://doi.org/10.1016/J.NEUROIMAGE.2004.10.034

Wong, V. S. S., Adamczyk, P., Dahlin, B., Richman, D. P., & Wheelock, V. (2011). Cerebral Venous Sinus Thrombosis Presenting With Auditory Hallucinations and Illusions. *Cognitive and Behavioral Neurology*, *24*(1), 40–42. https://doi.org/10.1097/WNN.0b013e318217b0cd

Woods, A. T., Lloyd, D. M., Kuenzel, J., Poliakoff, E., Dijksterhuis, G. B., & Thomas, A. (2011). Expected taste intensity affects response to sweet drinks in primary taste cortex. *Neuroreport*, *22*(8), 365–369.

Wright, S. (1954). The Death of Lady Mondegreen. *Harper's Magazine*, *209*(1254), 48–51.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology: Monograph Supplement*, *9*.

Zareva, A. (2007). Structure of the second language mental lexicon: how does it compare to native speakers' lexical organization? *Second Languare Research*, *23*(2), 123–153.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1493), 1087–104. https://doi.org/10.1098/rstb.2007.2161

Zekveld, A. A., Heslenfeld, D. J., Festen, J. M., & Schoonhoven, R. (2006). Top-down and bottom-up processes in speech comprehension. *NeuroImage*, *32*(4), 1826–36. https://doi.org/10.1016/j.neuroimage.2006.04.199

8 Author Contribution (Erklärung zum Eigenanteil)

8.1 General

This dissertation was performed within the Department of Biomedical Magnetic Resonance, University of Tübingen, under the supervision of Prof. Thomas Ethofer. I declare that I independently wrote the manuscript and used no sources other than those indicated, nor any aids other than those permissible. I appropriately declared all citations.

8.2 Concerning the results of “Mondegreens and Soramimi as a Method to Induce Misperceptions of Speech Content”

Study conception and design in cooperation with Prof. Thomas Ethofer. Claudia Beck Lidén performed the experiments. Analysis of the data in cooperation with Bernd Kardatzki who also contributed with analysis tools. The paper was written in collaboration with Prof. Thomas Ethofer.

8.3 Concerning the results of “Neurobiology of knowledge and misperception of lyrics”

Study conception and design in cooperation with Prof. Thomas Ethofer. Performance of the experiments in cooperation with Prof. Thomas Ethofer.

Analysis of the data in cooperation with Bernd Kardatzki, Dr. Michael Erb, Oliver Krüger and Lena Schwarz. Prof. Klaus Scheffler provided support with respect to applied MR sequences. The sections concerning structural connectivity (diffusion-weighted data and probabilistic fiber tracking) are necessarily excluded from the scientific research questions of this thesis, as these parts were analyzed solely by Oliver Krüger and Prof. Thomas Ethofer. The paper was written in collaboration with Prof. Thomas Ethofer, excluding the parts concerning the structural connectivity which were written solely by Oliver Krüger and Prof. Thomas Ethofer.

9 Abbreviations

ACC	Anterior cingulate cortex
BOLD	Blood oxygenation level-dependent
fMRI	Functional magnetic resonance imaging
IFG	Inferior frontal gyrus
mdTh	Medial dorsal thalamic nuclei
MGB	Medial geniculate body
MPFC	Medial prefrontal cortex
mSTS	Middle superior temporal sulcus
PET	Positron emission tomography
RF	Radio frequency
SMA/dACC	Supplemental motor area/dorsal anterior cingulate cortex
STG	Superior temporal gyrus

10 Acknowledgements

I would like to thank Prof. Thomas Ethofer for his steady support even if important settings changed during those years. I respect you for many things: your integrity, your intellect, your optimism and your wittiness. It was a pleasure and an honor to have the opportunity to be able to work with you, and our subject even included research on wittiness…

Warm and strong thanks go to the members of the Section of Experimental MRI of the Central Nervous System as there are Bernd Kardatzki, Dr. Michael Erb and Prof. Uwe Klose. Without you and your help and support with all upcoming possible problems (especially all the technical ones, data analyses, programming and many others) – this research would simply not have emerged. Your help was invaluable. On top of that, I always felt welcomed at the “Section”. Additionally, thanks go to Franziska Hösl who made the lab an organized and simply good place to be. I would also like to thank Lena Schwartz and Oliver Kruger for their support and contribution with their skills and knowledge as co-authors of the NeuroImage paper. Special thanks go to Prof. Klaus Scheffler for appropriating the necessary MRI facilities.

I also would like to thank Dr. László Koszytu and Dr. Karin Huisman from the surgery and gynecology departments in Hudiksvall for providing time off from clinical duties, so that I could finish this thesis.

I owe my gratitude to Dr. Andrea Fuchs for her critical approach and constructive as well as inspiring support.

Dipl. Ing. Alexander Eldracher-Beck, thank you for being my big brother. I can always count on you. I am also deeply indebted to Dr. Sylvia Straub who supported me during all those years with her steady friendship and warm offers of a home away from home each time I come back to Stuttgart/Tübingen.

Particular thanks go to Alexys Lang for proofreading of the manuscript.

Special thanks also to all the study participants; without you, there would not be any studies.

Last but not least, my warmest and deepest thanks go to my husband Oskar Lidén for his support, love and belief in me. You are the jackpot in my life-lottery.