

Cross-modal effects in the construction of perception

Citation for published version (APA):

Kilian-Hütten, N. J. (2012). *Cross-modal effects in the construction of perception*. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20120203nk>

Document status and date:

Published: 01/01/2012

DOI:

[10.26481/dis.20120203nk](https://doi.org/10.26481/dis.20120203nk)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Niclas Kilian-Hütten

CROSS-MODAL EFFECTS IN THE
CONSTRUCTION OF PERCEPTION



CROSS-MODAL EFFECTS IN THE CONSTRUCTION OF PERCEPTION

Niclas J. Kilian-Hütten

No part of this publication may be reproduced, stored in a automated data system or transmitted in any form or by any means, electronic, mechanical or photocopying, recording, or otherwise, without the prior written permission of the author.

© Niclas J. Kilian-Hütten, Maastricht, 2012

ISBN 978-90-8570-972-5

Cross-modal effects in the construction of perception

Proefschrift

Ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. mr. G.P.M.F. Mols
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op vrijdag 03 februari 2011 om 16.00 uur

door

Niclas J. Kilian-Hütten

geboren op 3 oktober 1981 te Düren

Promotor

Prof. dr. E. Formisano

Copromotor

Prof. dr. R. Goebel

Beoordelingscommissie

Prof. dr. B. Jansma (voorzitter)

Dr. Nienke van Atteveldt

Dr. Michael Beauchamp (University of Texas)

Prof. dr. P. de Weerd

The work presented in this thesis was supported by a
Netherlands Organization for Scientific Research (NWO) TopTalent Grant.

CONTENTS

Chapter 1	General introduction	7
Chapter 2	Auditory cortex encodes the perceptual interpretation of ambiguous sound	25
Chapter 3	Brain activation during audiovisual exposure anticipates future perception of ambiguous speech	45
Chapter 4	Correspondence between emotion perception biases in clinical depression and musically induced sad mood	69
Chapter 5	Neural binding of separate cognitive aspects of depression through musically induced sad mood	85
Summary		113
Samenvatting		117
Acknowledgements		121
Curriculum Vitae		129

CHAPTER 1

GENERAL INTRODUCTION

Reality is merely an illusion, albeit a very persistent one.

Albert Einstein

When we move around in the world, we heavily rely on our senses to provide us with an accurate representation of reality. We feel that, under normal circumstances, we see, hear, taste, smell or touch reality. However, philosophers and psychologists have known for a long time that this notion of a direct projection of reality is an illusion. Instead, a multitude of factors, including the interaction between sensory modalities, shapes our constructive perception of the world.

In order to illustrate this point, I will devote the first part of this introduction to a short synopsis of the ideas of constructivism. Following from this, I will differentiate the notions of sensation and perception in psychology and neuroscience and shortly summarize the state of research in this domain. In the remaining paragraphs I will review the current state of multisensory neuroscience and introduce a number of methods essential for the understanding of this work (functional magnetic resonance imaging (fMRI), multivoxel pattern analysis (MVPA), independent component analysis (ICA), and ICA-based psychophysiological interaction (PPI)).

Constructivism

Distinguishing between knowing something and not knowing something seems to be a straightforward feat. Furthermore, our society seems to be based on a clear understanding of the concept of knowledge and about ways of acquiring it, for instance in schools and classes. In stark contrast to this matter-of-course handling stands the historical observation that a multitude of philosophical schools, religious ideas, and applied frameworks of thought have generated an equally large number of opinions and conceptualizations of just what knowledge entails. What most of these concepts have in common is that they presuppose that there is an ontological reality, or truth, that we may discover through faith, logical rationalism, or scientific methodology.

Constructivism, as first advocated by Giambattista Vico in the 18th century, and prominently espoused and re-discovered by Jean Piaget and, in its radical form, by Ernst von Glasersfeld (von Glasersfeld, 1984), proposes a different take on this concept. While it distances itself from solipsism, which would deny the very existence of a world outside the mind of the subject, it does deny our *access* to an abstract, objective, ontological reality. In

other words, constructivism emphasizes the subjective, point-of-view, relative, and context-based aspects in the observer's construction of reality/knowledge. Thus, stimuli in the environment influence the organism by eliciting changes in its structural dynamics, but do so differentially depending on its current structural status. Perception is thus seen as a reciprocal process between a living entity and its environment (Maturana, 1980). A very apprehensible illustration of these concepts is provided by the consideration of ambiguous perceptual illusions. The famous image shown in figure 1 ("All is Vanity" by Charles Allan Gilbert', 1892) can be interpreted in two distinct ways- it either shows a lady looking at herself in the mirror or a large skull. Thus, the same, physically constant, external input can be perceived differentially; a phenomenon which led psychologists to distinguish between the concepts of *sensation* and *perception*.

Sensation, perception, and multistable percepts

Hermann von Helmholtz was the first to introduce the distinction between sensation and perception. While traditionally perception had been conceived of as a passive representation of the outside world, Helmholtz stressed the importance of unconscious, knowledge-based inferences on the basis of (potentially insufficient or irrelevant) information transmitted via the senses. Thus, percepts can be conceptualized as hypotheses; knowledge acting as the source for resolving sensational ambiguities (Gregory, 1997).

The counter-intuitive indirectness of perception can be illustrated considering basic physiological perceptual mechanisms. Color is not a property of a given object, but is the interpretation our brain creates from calculating ratios of excitations of three types of photoreceptors, which are themselves based on the reflection of light of different wavelengths. Similar mechanisms underlie other phenomenal properties of external objects. Sensations are transmitted via neuronal action potentials and may be influenced by hierarchically higher cortical regions. To stay with the example of color vision; it has, for instance, been shown that activity in cortical language areas may influence activity in color-processing regions, affecting color-discrimination behavior (Ting Siok, et al., 2009).

Such so-called top-down influences of bottom-up (i.e. stimulus-driven) information processing lie at the heart of constructive perception. However, traditionally, studies in psychology and neuroscience have relied on distinct stimuli in order to elicit distinct percepts. This logic is comprehensible given the nature of our everyday experience.

Yet, it inextricably confounds sensation and perception. In order to differentiate between them and to investigate the psychological and neural basis of “pure” perception, bi- or multistable phenomena have been reliably employed in visual research (Sterzer, Kleinschmidt, & Rees, 2009). Multistable perception refers to alternating perceptual interpretations of the same ambiguous stimulus, as illustrated in the famous example of the Necker cube (figure 2). These phenomena yield the possibility of disentangling sensation and perception, since the physical stimulus input is constant, while the perceptual interpretation changes. Neural correlates of multistable perceptual switches in vision have been found in hierarchically lower, sensory regions, as well as in higher-order frontal and parietal regions (Sterzer, et al., 2009). However, it is unclear how exactly bottom-up and top-down influences interact in the creation of a constructive, conscious percept and where the percept itself is actually coded. One model of interest in this context is Lamme and Roelfsema’s theory of recurrent processing, which posits that a first, feedforward sweep of neuronal processing rapidly analyzes stimulus features, while recurrent processing (i.e. feedback from higher-order to primary cortices) is necessary for a stimulus to enter consciousness, i.e. for a percept to be formed (Lamme & Roelfsema, 2000).

A multitude of different sources of information may conceivably be reflected in these recurrent processes and, thus, affect perceptual processing. For instance, semantic context (Doehrmann & Naumer, 2008; Rumelhart & McClelland, 1982), motivation (Balcetis & Dunning, 2006), memory (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006), familiarity and exposure (Moreland & Zajonc, 1982), expectation (Summerfield & Egner, 2009), attention (Boynton, 2005), personality (Chen & Dalton, 2005), mood (Schmitz, De Rosa, & Anderson, 2009), and information input to other sensory modalities (Shimojo & Shams, 2001) are all factors that have been shown to affect perception individually and in interaction.

Multisensory interactions at different levels

Cross-modal effects in basic perception

The main focus of this thesis lies on this latter point; the influence exerted by the interplay between different sensory modalities at very concrete and at more abstract levels of perception. There is a wealth of evidence demonstrating behaviorally beneficial multisensory interaction and integration in humans (Bulkin & Groh, 2006; Shimojo & Shams, 2001). Congruent multisensory inputs typically lower detection thresholds

(Frassinetti, Bolognini, & Ladavas, 2002), shorten reaction times (Forster, Cavina-Pratesi, Aglioti, & Berlucchi, 2002; Schroger & Widmann, 1998), and decrease saccadic eye movement latencies (Hughes, Reuter-Lorenz, Nozawa, & Fendrich, 1994). Congruency here refers to temporal, spatial, and semantic stimulus characteristics (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004). Incongruent multisensory input has opposite consequences (Sekuler, Sekuler, & Lau, 1997).

An increasing number of neuroscientific studies have begun investigating the neural substrates of cross-modal integration in cats, primates and in human subjects (Beauchamp, 2004; Beauchamp, 2005; Calvert, Campbell, & Brammer, 2000; Green & Angelaki, 2010; Schroeder & Foxe, 2005; Stein & Stanford, 2008; van Atteveldt, Formisano, Goebel, & Blomert, 2004). Neuroimaging studies in humans have consistently identified higher-order regions as loci of audiovisual and audiotactile integration, such as intraparietal regions, or most prominently, superior temporal sulcus (STS). However, more recently, cross-modal interactions have been also observed in locations traditionally regarded as unimodal sensory cortex (Alais, Newell & Mamassian, 2010; Koelewijn, Bronkhorst & Theeuwes, 2010).

One of the most researched topics in the domain of multisensory integration is the nature of a particular class of neurons, which receive input from, and respond to, different sensory modalities. Interestingly, these multisensory neurons exhibit a response pattern that mirrors the behavioral benefits of multisensory integration described above. More specifically, these neurons fire in response to stimuli presented to different modalities individually, demonstrate supra-additive firing patterns in response to congruent multisensory input, and sub-additive firing patterns in response to incongruent multisensory stimulation. They have most extensively been researched in cats' and macaques' superior colliculi and cerebral cortices, including the homologue of the human STS (Meredith & Stein, 1983, 1986; Ohshiro, Angelaki, & DeAngelis, 2011; Schroeder, et al., 2001), but direct electrophysiological evidence for multisensory neurons in human auditory cortex has also accumulated (Calvert, Spence, & Stein, 2004; Ghazanfar & Schroeder, 2006).

While the traditional view of multisensory integration posits that unisensory processing streams in the brain converge in higher-order integrative regions and then influence processing in unisensory areas, it has to be noted that recently evidence has also accumulated for early, feedforward cross-modal influences (Foxe & Schroeder, 2005).

Neither of these possibilities, however, would deny the importance of multisensory integration in the construction of conscious perception.



Figure 1. *All is Vanity* by Charles Allan Gilbert, 1892. The observer can either see a lady looking at herself in the mirror or a large skull.

One especially striking and well-known example in the domain of multisensory integration is the McGurk effect (McGurk & MacDonald, 1976). In contrast to the effects discussed above, here the addition of the visual modality not only affects stimulus detection

or discrimination, but it actively alters the identity of the percept. More specifically, in a typical experiment, an auditory disyllable (e.g., /aba/) is paired with an incongruent visual disyllable (/aga/), pronounced by a speaker. The addition of this incongruent visual input alters the listeners' auditory percept (/ada/). Neuroimaging (Jones & Callan, 2003; Nath & Beauchamp, 2011; Sekiyama, Kanno, Miura, & Sugita, 2003), magnetoencephalography (Fingelkurts, Krause, Mottonen, & Sams, 2003; Keil, Muller, Ihssen, & Weisz, 2011; Mottonen, Krause, Tiippana, & Sams, 2002) and transcranial magnetic stimulation TMS studies (Beauchamp, Nath, & Pasalar, 2010) have generally converged on the view that left superior temporal sulcus (STS) and its functional connections to other regions can be regarded as the neural substrate of McGurk perception.

In a behavioral study, Bertelson, Vroomen, and De Gelder (2003) demonstrated that a paradigm based on the McGurk effect could even be employed to induce lasting perceptual after-effects, resulting in enduring perceptual biases. The major difference in stimulation, compared to other studies employing the McGurk effect, was the use of an ambiguous auditory stimulus, intermediate between /aba/ and /ada/. It was demonstrated that exposure to this ambiguous stimulus dubbed onto a video of a face pronouncing /aba/ or /ada/ selectively increased the proportion of corresponding responses in subsequent audio-only forced-choice trials. This direct cross-modal perceptual effect yields the unique possibility of investigating neural responses to auditory stimuli that are subject to two distinct perceptual interpretations, while being physically constant. This phenomenon can be regarded as the auditory analogue of bistable visual stimuli and represents the basis for the work presented in chapters 2 and 3 of this thesis. In these chapters, it is investigated how the varying perceptual interpretations of physically identical auditory stimuli can be predicted on the basis of brain data.

Cross-modal effects in emotional processing

The auditory and visual systems do not exclusively interact in such a direct, stimulus-driven way. On a more abstract level, cross-modal effects in emotional processing have been demonstrated in a multitude of different contexts. For instance, emotional vocal intonations have been shown to influence the perception of emotional facial expressions (Collignon, et al., 2008; Ethofer, et al., 2006; Pourtois, de Gelder, Bol, & Crommelinck, 2005; Stienen, Tanaka, & de Gelder, 2011) and vice versa (de Gelder, Pourtois, & Weiskrantz, 2002). Similar interactions have been demonstrated for the integration of signals from

vocalizations and bodily expressions (van den Stock, de Jong, Hodiamont, & de Gelder, 2011), as well as from auditory sources and the perceived emotion of a (visually presented) crowd (McHugh, Kearney, Rice, & Newell, 2011).

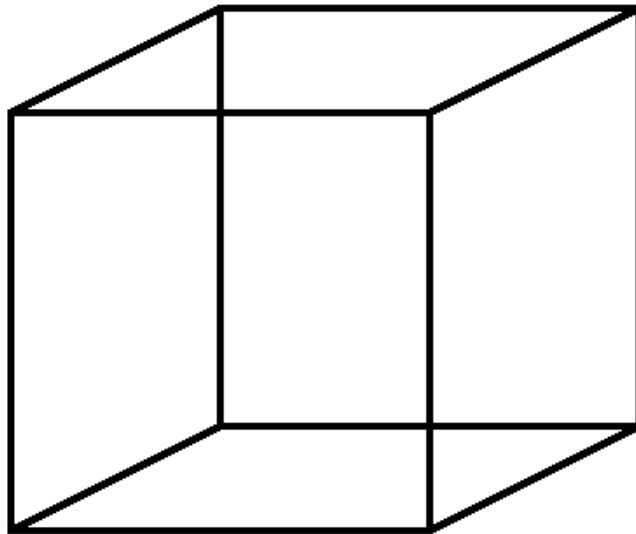


Figure 2. The Necker cube illusion (Necker, L. A., 1832, “Observations on some remarkable optical phaenomena seen in Switzerland; and on an optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid,” London and Edinburgh Philosophical Magazine and Journal of Science, Vol. 1, No. 5, pp. 329–337). Either the bottom-left or the top-right surface can be seen as being in front. The perceptual interpretation may switch spontaneously, making this an example of a bistable phenomenon.

One especially potent source of cross-modal emotional influence is music. While it has recently also been demonstrated that visual information (from the observation of the performing musicians) can influence the pitch-related and emotional perception of music itself (Thompson, Russo, & Livingstone, 2010; Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2011), the most striking effect follows the opposite trajectory. Music is an extremely powerful emotional stimulus and has been shown to influence consumer behavior (North, Hargreaves, Juslin, & Sloboda, 2010), pain experience (Huang, Good, & Zauszniewski, 2010), sexual arousal (ter Kuile, Both, & van Uden, 2010), emotional and

cognitive processing (Forgas & East, 2008), and even aid therapeutical efforts in such diverse disorders as Parkinson's (Pacchetti, et al., 2000) or visual neglect (Soto, et al., 2009). It has further been demonstrated that the emotional valences transported via faces and music, respectively, interact and modify cortical multisensory activation depending on their congruency (Jeong, et al., 2011). Furthermore, it has been found that emotional music can effectively influence the perception of emotional faces (Bouhuys, Bloem, & Groothuis, 1995; Jolij & Meurs, 2011; Logeswaran & Bhattacharya, 2009) and modulate the strength, and event-related responses, of an emotional negativity bias in face perception (Chen, Yuan, Huang, Chen, & Li, 2008). These cross-modal emotional effects are the basis for the work presented in chapters 4 and 5. This part of the thesis examines in how far music can influence implicit visual emotion perception of facial expressions, mirroring cognitive symptoms of clinical depression. Furthermore, it is investigated how neural networks processing different cognitive aspects of clinical depression interact dynamically in healthy participants undergoing musical mood induction.

Functional magnetic resonance imaging (fMRI)

Besides behavioral measures, functional magnetic resonance imaging (fMRI) is the main method used in the research presented in this thesis. Functional MRI is a non-invasive measure of oxygenated blood flow, which is related to neuronal activation. When neurons become active, they consume oxygen and glucose. The capillary system (likely instigated by astrocyte cells) overcompensates for this consumption, resulting in an over-proportional increase of oxygenated blood in regions that were just active. Thus, counter-intuitively, brain regions that respond to a certain sensory stimulation or are involved in a relevant cognitive process receive a large influx of oxygenated blood. This effect represents the indirect measure of neuronal activity that is picked up by the fMRI signal. In order to do so, participants enter a strong constant magnetic field (3 Tesla in this research) which aligns the protons in hydrogen atoms (which are abundantly present in the human body). A radiofrequency pulse is then sent in, which places the protons in a high-energy state and lets them oscillate in phase. The fMRI receiver coil then measures local energy fields that are emitted when the protons return to equilibrium (the orientation of the external magnetic field). The influence exerted by the influx of oxygenated blood results from the fact that deoxygenated hemoglobin is paramagnetic, which leads to local magnetic field

inhomogeneities in the capillary bed, which in turn speeds up MR signal decays. Conversely, the influx of oxygenated hemoglobin in response to neuronal activity results in a larger oxy/deoxy-hemoglobin ratio, and thus in a slower decay of the MR signal, effectively enhancing the signal compared to a non-active region (baseline; DeYoe, Bandettini, Neitz, Miller, & Winans, 1994). Typically, the hemodynamic response is very slow (in the range of several seconds), which makes the temporal resolution of fMRI relatively low. However, it provides excellent spatial resolution, being able to detect changes in the Blood Oxygenation Level Dependent (BOLD) signal with a specificity of few millimeters or less.

Analysis methods employed

Detailed descriptions of the methods employed in the studies of this thesis are provided in the individual chapters. However, it seems useful to provide a conceptual summary of the somewhat more advanced analysis methods in this introduction. Traditionally, the statistical analysis of fMRI data sets came forth from the positron emission topography (PET) analysis rationale, which rests on the notion of subtraction logic. This approach entails the presentation of different experimental conditions in the scanner, which differ in the cognitive process of interest. A statistical linear model, the general linear model (GLM), can then be applied to the data. In these regression models, experimental conditions are represented by individual predictors. For the time course of each voxel, the GLM then aims to predict the variation in terms of a weighted combination of these predictors (idealized fMRI responses). A strong weight, or beta value, for a given predictor indicates that the tested voxel exhibits strong activation during the experimental condition as compared to baseline. Post-hoc contrasts can then be applied to statistically compare the activation during different experimental conditions. Predictors in a GLM do not need to be simple on/off conceptualizations of experimental conditions. In chapter 3 I describe a variant of traditional GLM-based fMRI analysis, which entails differentially weighting a predictor based on behavioral values obtained later in time. This behaviorally-weighted contrast, which is explained in more detail in the chapter in question, hence identifies brain regions whose activation patterns predict future behavioral tendencies.

Recently, novel analysis techniques have become increasingly popular in the field of neuroimaging. Multivoxel pattern analysis (MVPA) is one such technique which applies machine learning methods to the multivariate analysis of fMRI data sets (Formisano, De

Martino, & Valente, 2008; Haxby, et al., 2001; Haynes & Rees, 2005). A pattern classification algorithm is typically trained, using a large set of trials, to associate a given experimental condition or cognitive state with a distributed pattern of fMRI responses. The trained classifier can then be tested on new, unseen fMRI patterns to decode the associated cognitive state. In other words, rather than predicting brain responses from experimental conditions, as in the GLM, MVPA can ‘predict’ the experimental condition from the brain response, which is why it has often been denoted as a ‘brain reading’ approach. This technique is employed in chapter 2 to decode conscious perceptual interpretations of physically ambiguous auditory stimuli from the activation of auditory cortical areas.

Another method that represents an alternative approach to GLM-based analyses is independent component analysis (ICA). While GLM approaches are hypothesis-driven, ICA is a data-driven method, which does not require explicit modeling of experimental conditions. Instead, the observed data are explained in terms of a linear mixture of spatiotemporal sources, based on spatial or temporal independence. The resulting independent components represent networks of brain regions that are functionally connected over the course of an experimental run (or whatever time course the ICA is applied to). These components, thus, may explain different aspects of the experimental task (auditory component, visual component, motor component etc.) or may represent noise due to large blood vessels, motion, or other sources. A challenge in applying ICA to fMRI data is the selection and interpretation of relevant components. In chapter 5, I present a combination of ICA with another functional connectivity analysis technique, psychophysiological interaction (PPI). A PPI analysis typically asks; given the time course of a region of interest (ROI), which other regions are functionally connected more strongly to this ROI during one, as compared to another, experimental condition. Consequently, the combination with ICA investigates which regions are functionally connected more strongly to a given *network* (component) during one, as compared to another, experimental condition.

References

- Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing Perceiving*, 23(1), 3-38.
- Balcetis, E., & Dunning, D. (2006). See what you want to see: motivational influences on visual perception. *J Pers Soc Psychol*, 91(4), 612-625.
- Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nat Neurosci*, 7(11), 1190-1192.
- Beauchamp, M. S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Curr Opin Neurobiol*, 15(2), 145-153.
- Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *J Neurosci*, 30(7), 2414-2417.
- Bertelson, P., Vroomen, J., & De Gelder, B. (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychol Sci*, 14(6).
- Bouhuys, A. L., Bloem, G. M., & Groothuis, T. G. (1995). Induction of depressed and elated mood by music influences the perception of facial emotional expressions in healthy subjects. *J Affect Disord*, 33(4), 215-226.
- Boynton, G. M. (2005). Attention and visual perception. *Curr Opin Neurobiol*, 15(4), 465-469.
- Bulkin, D. A., & Groh, J. M. (2006). Seeing sounds: visual and auditory interactions in the brain. *Curr Opin Neurobiol*, 16(4), 415-419.
- Calvert, G., Spence, C., & Stein, B. E. (2004). *The Handbook of Multisensory Processes*: MIT Press.
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol*, 10(11), 649-657.
- Chen, D., & Dalton, P. (2005). The effect of emotion and personality on olfactory perception. *Chem Senses*, 30(4), 345-351.
- Chen, J., Yuan, J., Huang, H., Chen, C., & Li, H. (2008). Music-induced mood modulates the strength of emotional negativity bias: An ERP study. *Neuroscience Letters*, 445(2), 135-139.

- Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., et al. (2008). Audio-visual integration of emotion expression. *Brain Res*, *1242*, 126-135.
- de Gelder, B., Pourtois, G., & Weiskrantz, L. (2002). Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proc Natl Acad Sci U S A*, *99*(6), 4121-4126.
- DeYoe, E. A., Bandettini, P., Neitz, J., Miller, D., & Winans, P. (1994). Functional magnetic resonance imaging (fMRI) of the human brain. *J Neurosci Methods*, *54*(2), 171-187.
- Doehrmann, O., & Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain Res*, *1242*, 136-150.
- Ethofer, T., Anders, S., Erb, M., Droll, C., Royen, L., Saur, R., et al. (2006). Impact of voice on emotional judgment of faces: an event-related fMRI study. *Hum Brain Mapp*, *27*(9), 707-714.
- Fingelkurts, A. A., Krause, C. M., Mottonen, R., & Sams, M. (2003). Cortical operational synchrony during audio-visual speech integration. *Brain Lang*, *85*(2), 297-312.
- Forgas, J. P., & East, R. (2008). How real is that smile? Mood effects on accepting or rejecting the veracity of emotional facial expressions. *Journal of Nonverbal Behavior*, *32*(3), 157-170.
- Formisano, E., De Martino, F., & Valente, G. (2008). Multivariate analysis of fMRI time series: classification and regression of brain responses using machine learning. *Magn Reson Imaging*, *26*(7), 921-934.
- Forster, B., Cavina-Pratesi, C., Aglioti, S. M., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Exp Brain Res*, *143*(4), 480-487.
- Foxe, J. J., & Schroeder, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *Neuroreport*, *16*(5), 419-423.
- Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp Brain Res*, *147*(3), 332-343.

- Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn Sci*, 10(6), 278-285.
- Green, A. M., & Angelaki, D. E. (2010). Multisensory integration: resolving sensory ambiguities to build novel representations. *Curr Opin Neurobiol*, 20(3), 353-360.
- Gregory, R. L. (1997). Knowledge in perception and illusion. *Philos Trans R Soc Lond B Biol Sci*, 352(1358), 1121-1127.
- Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nat Neurosci*, 9(11), 1367-1368.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539), 2425-2430.
- Haynes, J. D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci*, 8(5), 686-691.
- Huang, S.-T., Good, M., & Zauszniewski, J. A. (2010). The effectiveness of music in relieving pain in cancer patients: A randomized controlled trial. *International Journal of Nursing Studies*, 47(11), 1354-1362.
- Hughes, H. C., Reuter-Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. *J Exp Psychol Hum Percept Perform*, 20(1), 131-153.
- Jeong, J. W., Diwadkar, V. A., Chugani, C. D., Sinsoongsud, P., Muzik, O., Behen, M. E., et al. (2011). Congruence of happy and sad emotion in music and faces modifies cortical audiovisual activation. *Neuroimage*, 54(4), 2973-2982.
- Jolij, J., & Meurs, M. (2011). Music alters visual perception. *PLoS One*, 6(4).
- Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport*, 14(8), 1129-1133.
- Keil, J., Muller, N., Ihssen, N., & Weisz, N. (2011). On the Variability of the McGurk Effect: Audiovisual Integration Depends on Prestimulus Brain States. *Cereb Cortex*, in press.
- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta Psychol (Amst)*, 134(3), 372-384.
- Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci*, 23(11), 571-579.

- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Exp Brain Res, 158*(4), 405-414.
- Logeswaran, N., & Bhattacharya, J. (2009). Crossmodal transfer of emotion by music. *Neurosci Lett, 455*(2), 129-133.
- Maturana, H. V., F. (1980). *Autopoiesis and Cognition: The Realization of the Living* (Vol. 42). Dordrecht (Holland): D. Reidel Publishing Co.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.
- McHugh, J. E., Kearney, G., Rice, H., & Newell, F. N. (2011). The sound of the crowd: Auditory information modulates the perceived emotion of a crowd based on bodily expressions. *Emotion*.
- Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science, 221*(4608), 389-391.
- Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol, 56*(3), 640-662.
- Moreland, R. L., & Zajonc, R. B. (1982). Exposure effects in person perception: Familiarity, similarity, and attraction. *Journal of Experimental Social Psychology, 18*(5), 395-415.
- Mottonen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Brain Res Cogn Brain Res, 13*(3), 417-425.
- Nath, A. R., & Beauchamp, M. S. (2011). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *Neuroimage, 59*(1):781-7.
- North, A. C., Hargreaves, D. J., Juslin, P. N., & Sloboda, J. A. (2010). Music and marketing *Handbook of music and emotion: Theory, research, applications.* (pp. 909-930). New York, NY US: Oxford University Press.
- Ohshiro, T., Angelaki, D. E., & DeAngelis, G. C. (2011). A normalization model of multisensory integration. *Nat Neurosci, 14*(6), 775-782.

- Pacchetti, C., Mancini, F., Aglieri, R., Fundaro, C., Martignoni, E., & Nappi, G. (2000). Active music therapy in Parkinson's disease: an integrative method for motor and emotional rehabilitation. *Psychosom Med*, *62*(3), 386-393.
- Pourtois, G., de Gelder, B., Bol, A., & Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex*, *41*(1), 49-59.
- Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychol Rev*, *89*(1), 60-94.
- Schmitz, T. W., De Rosa, E., & Anderson, A. K. (2009). Opposing influences of affective state valence on visual cortical encoding. *J Neurosci*, *29*(22), 7199-7207.
- Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, 'unisensory' processing. *Curr Opin Neurobiol*, *15*(4), 454-458.
- Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., & Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *J Neurophysiol*, *85*(3), 1322-1327.
- Schroger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology*, *35*(6), 755-759.
- Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neurosci Res*, *47*(3), 277-287.
- Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*(6614), 308.
- Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Curr Opin Neurobiol*, *11*(4), 505-509.
- Soto, D., Funes, M. J., Guzman-Garcia, A., Warbrick, T., Rotshtein, P., & Humphreys, G. W. (2009). Pleasant music overcomes the loss of awareness in patients with visual neglect. *Proc Natl Acad Sci U S A*, *106*(14), 6011-6016.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci*, *9*(4), 255-266.
- Sterzer, P., Kleinschmidt, A., & Rees, G. (2009). The neural bases of multistable perception. *Trends Cogn Sci*, *13*(7), 310-318.

-
- Stienen, B. M. C., Tanaka, A., & de Gelder, B. (2011). Emotional Voice and Emotional Body Postures Influence Each Other Independently of Visual Awareness. *PLoS One*, 6(10), e25517.
- Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(9), 403-409.
- ter Kuile, M. M., Both, S., & van Uden, J. (2010). The effects of experimentally-induced sad and happy mood on sexual arousal in sexually healthy women. *Journal of Sexual Medicine*, 7(3), 1177-1184.
- Thompson, W. F., Russo, F. A., & Livingstone, S. R. (2010). Facial expressions of singers influence perceived pitch relations. *Psychon Bull Rev*, 17(3), 317-322.
- Ting Siok, W., Kay, P., Wang, W. S., Chan, A. H., Chen, L., Luke, K. K., et al. (2009). Language regions of brain are operative in color perception. *Proc Natl Acad Sci U S A*, 106(20), 8140-8145.
- van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, 43(2), 271-282.
- van den Stock, J., de Jong, S. J., Hodiament, P. P., & de Gelder, B. (2011). Perceiving emotions from bodily expressions and multisensory integration of emotion cues in schizophrenia. *Soc Neurosci*, 6(5-6):537-47.
- Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: cross-modal interactions in the perception of emotions in musical performance. *Cognition*, 118(2), 157-170.
- von Glasersfeld, E. (1984). An Introduction to Radical Constructivism. In P. Watzlawick (Ed.), *The Invented Reality - How Do We Know What We Believe We Know?* (pp. 17-40). New York: W. W. Norton.

**AUDITORY CORTEX ENCODES THE PERCEPTUAL
INTERPRETATION OF AMBIGUOUS SOUND**

CORRESPONDING PUBLICATION

KILIAN-HÜTTEN, N., VALENTE, G., VROOMEN, J., AND FORMISANO, E. (2011). AUDITORY CORTEX ENCODES THE PERCEPTUAL INTERPRETATION OF AMBIGUOUS SOUND. *J. NEUROSCI.*, 31(5):1715-1720.

Abstract

The confounding of physical stimulus characteristics and the perceptual interpretations of stimuli poses a problem for most neuroscientific studies of perception. In the auditory domain, this pertains to the entanglement of acoustics and percept. Traditionally, most study designs have relied on cognitive subtraction logic, which demands the use of one or more comparisons between stimulus types. This does not allow for a differentiation between effects due to acoustic differences (i.e. sensation) and those due to conscious perception.

In order to overcome this problem, here we use functional Magnetic Resonance Imaging (fMRI) in humans and pattern recognition analysis to identify activation patterns which encode the perceptual interpretation of physically identical, ambiguous sounds. We show that it is possible to retrieve the perceptual interpretation of ambiguous phonemes – information that is fully subjective to the listener – from fMRI measurements of brain activity in auditory areas in the superior temporal cortex, most prominently on the posterior bank of the left Heschl's gyrus and sulcus and in the adjoining left planum temporale.

These findings suggest that - beyond the basic acoustic analysis of sounds - constructive perceptual processes take place in these relatively early cortical auditory networks. This disagrees with hierarchical models of auditory processing, which generally conceive of these areas as sets of feature detectors, whose task is restricted to the analysis of physical characteristics and the structure of sounds.

Introduction

Investigations of perceptual processing are usually faced with a confounding problem when attempting to separate out subjective perception from the processing of stimulus-specific characteristics. This is because usually differential percepts follow differential physical stimulus characteristics, i.e. distinct stimuli elicit distinct percepts. The traditional reliance on cognitive subtraction logic, which demands the use of one or more comparisons between stimulus types, renders dissociating sensation from conscious perception impossible.

In the visual domain, this problem has been tackled employing ambiguous stimuli and the phenomenon of multistable perception (for a review see Sterzer, Kleinschmidt, & Rees, 2009). In the auditory domain, however, the creation of stimuli and designs which permit comparing differential perceptual states while keeping the physical input constant seems a bigger challenge. In previous studies, temporal phenomena, such as auditory streaming (Cusack, 2005; Gutschalk, et al., 2005) or illusory continuity (Riecke, Esposito, Bonte, & Formisano, 2009) have been investigated similarly in human subjects. However, to date, no attempt has been made to employ ecologically-valid, ambiguous auditory stimuli to investigate the neural basis of differential perceptual interpretations of an auditory stimulus's *identity*.

Here we use functional Magnetic Resonance Imaging (fMRI) and pattern-recognition analysis to identify activation patterns which encode the perceptual interpretation of physically-identical, ambiguous phonemes. We adhere to a principle based on the McGurk effect, called cross-modal recalibration (Bertelson, Vroomen, & De Gelder, 2003), in which lip-movements are used to disambiguate ambiguous auditory phonemes. Repeated presentation of these videos increases the proportion of corresponding responses in *subsequent* audio-only forced-choice trials, thus eliciting an after-effect. This enables us to compare physically-identical, yet differentially-perceived, sounds, and thus, allows for the investigation of a purely perceptual distinction of stimulus identity.

According to popular models of auditory processing, representations become more abstract with hierarchical distance to the primary auditory cortex (A1) along two (what/where) pathways (Liebenthal, Binder, Spitzer, Possing, & Medler, 2005; Rauschecker & Scott, 2009; Scott & Johnsrude, 2003). In humans, the regions adjacent to the Heschl's gyrus - that we refer to as early auditory cortex - are supposedly restricted to the analysis of physical features and the acoustic structure of sounds. In contrast to this

notion, fMRI and pattern recognition techniques (Haxby, et al., 2001; Haynes & Rees, 2005) have recently been used to demonstrate the existence of vowel representations in these regions which were *invariant* of the specific speaker uttering them (Formisano, De Martino, Bonte, & Goebel, 2008). These findings suggest the presence of abstract perceptual sound representations in early auditory areas. The same vowels, however, even when uttered by distinct speakers, still share an abundance of acoustic similarities. Whether it is these acoustic features or the perceptual categorization of the vowels which forms the basis of classification cannot be conclusively resolved by this study. The use of acoustically identical stimuli, which are perceptually categorized as distinct auditory entities, seems to be the only way to overcome this problem. Here we trained our pattern classification algorithm on purely perceptual labels and restricted the information provided to it to voxels within the temporal lobe to investigate the role of the auditory cortex in constructive perceptual processing.

Materials and methods

Participants

Twelve healthy native Dutch students of the University Maastricht (5 male, mean age 24.83) were recruited to participate in the study. One participant was left-handed. None of the participants had a history of hearing loss or neurological abnormalities. Approval for the study was granted by the Ethical Committee of the Faculty of Psychology at the University of Maastricht. One subject was discarded on the basis of a self-reported difficulty in categorizing the phonemes (perceiving them exclusively as /asa/), which was reflected as an extreme response bias in the behavioral data.

Stimuli

The stimulation entailed digital auditory and visual recordings of a male Dutch speaker pronouncing the syllables /aba/ and /ada/. The two auditory stimuli were 640-ms stimuli with 240-ms stop closure. From these, a place-of-articulation continuum was synthesized by means of varying the F2 formant by equal steps of 39 Mel, resulting in 9 different stimuli, ranging from a clear /aba/ via seven ambiguous stimuli to a clear /ada/. The audiovisual stimuli were synthesized by pairing the visual recordings of the speaker pronouncing /ada/ and /aba/, respectively, with the most ambiguous auditory stimulus, determined as such in a pre-test.

Behavioral pre-test

Each participant underwent an auditory pre-test outside the fMRI scanner in order to individually determine the most ambiguous auditory stimulus from the /aba/-to-/ada/ continuum (Bertelson, et al., 2003). This pre-test consisted of 98 forced-choice judgments on all different stimuli from the continuum, with presentation frequency biased, so that the five central stimuli were presented on 14 trials, the second and tenth stimuli were presented on 8 trials, and the first and the eleventh stimuli were presented on 6 trials. Stimuli were presented binaurally through loudspeakers. Participants were required to press one of two buttons if they had perceived /aba/ and the other one in case they perceived /ada/. This resulted in an estimation of each participant's ambiguous auditory token (A?), which was used for the rest of the session (see figure 1).

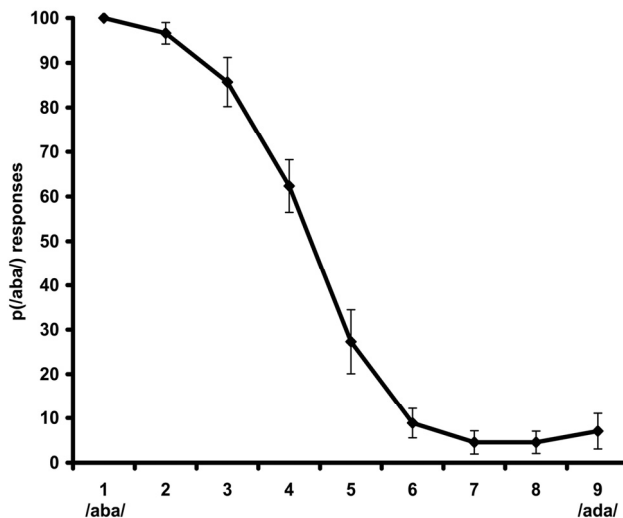


Figure 1. Results of the auditory pre-test. The mean proportions of /aba/ classifications across the 11 participants for each stimulus in the 9-sound continuum are given. Sound 4 was chosen as A? for 8 of the participants and sound 5 for the remaining 3.

Experimental procedure

The experimental procedure during scanning was based on the phenomenon of cross-modal recalibration, a McGurk after-effect (Bertelson, et al., 2003). In a typical McGurk paradigm, an auditorily presented disyllable (/aba/) is paired with an incongruent visual

disyllable (/aga/), pronounced by a speaker. The addition of this incongruent visual input changes the listener's auditory percept into an intermediate one (/ada/). Interestingly, when an ambiguous auditory component (A?, between /aba/ and /ada/) is used, exposure to this ambiguous stimulus dubbed onto a video of a face pronouncing /aba/ or /ada/ selectively increases the proportion of corresponding responses in *subsequent* audio-only forced-choice trials. This makes it possible to contrast conditions where the physical stimulus is identical, while the perceptual interpretation differs by comparing brain activation patterns in response to auditory post-test stimuli perceived as /aba/ versus /ada/. This logic, thus, allows for the investigation of a purely perceptual distinction of stimulus identity.

The design entailed two major elements (see Figure 2): blocks of multimodal exposure (the recalibration phase) and slow event-related auditory post-tests. In the recalibration phase, videos were presented, consisting of the individually determined ambiguous stimulus A?, dubbed on the visual recording of the speaker pronouncing either /aba/ (Vb) or /ada/ (Vd). Bimodal stimuli were presented in blocks of 8 identical trials, (block A?Vb or block A?Vd, respectively), with an interstimulus interval (ISI) of one TR (= 2000 ms). In each of the two runs, 5 A?Vb and 5 A?Vd blocks were run in randomized order (total: 160 trials). During these exposure trials, participants were required to press a button whenever a small white spot (12 pixels) appeared on the speaker's upper lip to ensure participants focus their attention on the speaker's lips. This occurred once per block at a random position.

Each block of bimodal exposure was followed by 6 auditory post-tests (total: 120 trials), which, like the pre-tests, consisted of forced-choice /aba/-/ada/ judgments. Unlike the pre-tests however, here only the A? token and the two tokens closest to it on the continuum were presented, twice each. Due to the use of a slow event-related design, the jittered ISI was 6 TR (= 12 s) on average (see Figure 2).

Scanning parameters

Functional MRI data were collected on a 3-Tesla fMRI scanner (head set-up, Siemens Allegra, Siemens, Erlangen, Germany) at the Maastricht Brain Imaging Center (M-BIC) in Maastricht, The Netherlands. For each participant, 2 functional runs of 665 volumes were acquired. For later overlay, also a high-resolution structural scan (voxel size = 1 x 1 x 1 mm³) was collected using a T1-weighted 3D ADNI sequence (TR = 2050 ms, TE = 2.6 ms, 192 sagittal slices). Both functional runs and the structural scan were acquired in a single

session for each participant. For functional images, a BOLD-sensitive echo-planar imaging sequence was used (matrix 64 x 64, 27 slices, slice thickness 3 mm, field of view, FOV, 192 x 192 mm³, resulting voxel size 3 x 3 x 3 mm³, TE/TA slice 30/55.5, FA = 90°). Volume acquisition was clustered in the beginning of each TR, leaving a silent delay within each TR, during which stimuli were presented in the absence of EPI noise. This was done in order to optimize stimulus audibility, an approach which has been shown to be highly efficient in auditory fMRI paradigms (Jancke, Wustenberg, Scheich, & Heinze, 2002; van Atteveldt, Formisano, Blomert, & Goebel, 2007). Hence, the effective TR was 2000 ms, including 1500 ms of sequence scanning time and a 500 ms silent delay. Stimuli were presented and synchronized with the MR pulses using the software package “Presentation” (Neurobehavioral Systems, Inc., Albany, California).

Data pre-processing

Functional and anatomical images were analyzed using BrainVoyager QX (Brain Innovation, Maastricht, The Netherlands) and customized code written in MATLAB (The MathWorks Inc., Natick, Massachusetts, 2007). Several pre-processing steps were performed; sinc interpolated slice-time correction, 3D-motion correction to correct for common small head movements by spatially aligning all volumes to the first volume by rigid body transformations, linear trend removal, and temporal high-pass filtering to remove low-frequency nonlinear drifts of 7 or less cycles per time course. Functional slices were then co-registered to the structural volume on the basis of positioning parameters from the scanner and manual adjustments to ensure optimal fit. Subsequently, they were transformed into Talairach Space. All individual brains were segmented at the gray/white matter boundary using a semi-automatic procedure based on intensity values implemented in BrainVoyager QX. For all group analyses, cortex-based alignment was used to assure optimized spatial matching of cortical locations (i.e., vertices) between participants (Goebel, Esposito, & Formisano, 2006).

Univariate data analysis

Functional runs were analyzed using voxel-wise multiple linear regression (GLM) of the BOLD-response time course. All analyses were performed at single-subject and group levels and all experimental conditions were modeled as predictors. Since the number of

these depended on the individual participant's behavior (i.e. if a participant never answered incorrectly to A?d stimuli after an A?Vd block, for instance, then this condition had zero events) and, thus, varied across runs and participants, empty predictors had to be included in all multi-run and multi-subject GLMs. In all GLMs, predictor time courses were convolved with a hemodynamic response function (where both the BOLD response and undershoot are modeled by a gamma function) to adjust for the hemodynamic delay (Friston, et al., 1998).

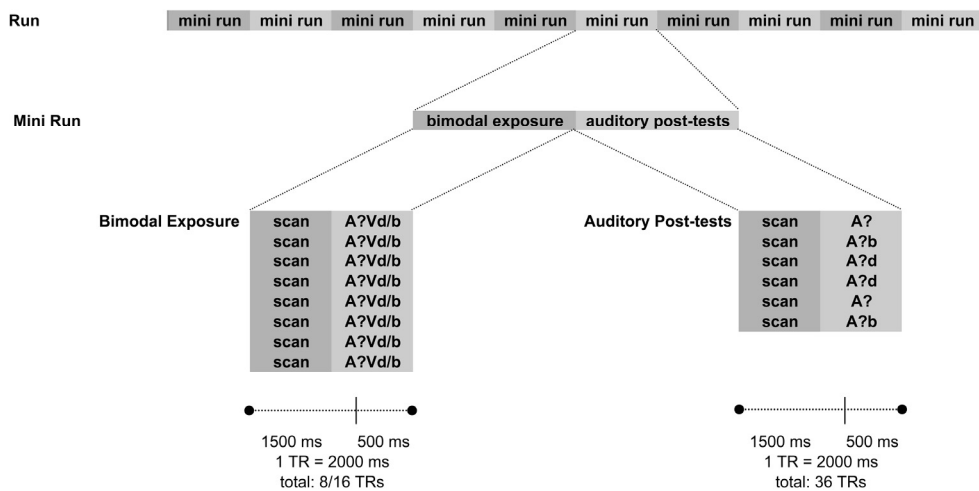


Figure 2. Schematic overview of the experimental procedure. Each run consisted of 10 mini runs, which were each composed of a bimodal exposure block (A?Vd or A?Vb, respectively) and 6 auditory post-test trials. The bimodal blocks were presented within 8 TRs, each of which entailed 1500 ms of scanning and 500 ms of video presentation. Between auditory post-test trials, fixation periods averaged 6 TRs.

Multi-voxel pattern analysis (MVPA) procedure

We decoded multivariate patterns of BOLD activation by training linear Support Vector Machine (SVM) classifiers based (solely) on perceptual labels (i.e. trials perceived as /ada/ or /aba/) and classified individual post-test auditory trials accordingly. We tested our hypothesis that abstract auditory representations can be found in early auditory cortex by confining classification anatomically to areas in the temporal lobe, excluding information from other areas (see Figure 3 (b)). An additional reason for this confinement is rooted in the necessity for the subjects to provide motor responses. The inclusion of motor areas

would arguably increase classification performance. However, since this improvement might be based on the differentiation between motor responses corresponding to the two responding fingers instead of two percepts, it would not be informative with respect to our research question.

We estimated the response of every voxel in each trial by fitting a general linear model with two predictors, one accounting for trial response, and the other one accounting for the trial mean. The hemodynamic response function was optimized on a subject basis. The pattern of activation associated with a given trial consisted of the set of beta values associated with the first predictor for all the voxels considered in the analysis. The steps of the feature extraction procedure are similar to those described and validated in (De Martino, et al., 2008).

The Interstimulus Intervals (ISI) in the experiment presentation ranged from 4 to 8 TRs (8 to 16 seconds, respectively), and we removed the ‘shortest’ trials (i.e. those that were followed by another post-test trial after 4 functional volumes) from the analysis, in order to use a larger time window to estimate betas on the remaining trials. We retained a total number of 100 trials, equally distributed among the 20 recalibration mini-blocks and two runs, from the original 120. The trials were divided in two disjoint groups, training and testing. As the two classes were generally unbalanced, we created several balanced training datasets by randomly selecting n trials from each class (with n being 80% of the total number of examples of the least represented class), leaving the remaining trials as testing dataset. We trained linear SVMs (Vapnik, 1995) on the training datasets and evaluated the generalization to new data on the test dataset. This procedure was repeated 100 times with different random trial selections.

Voxels that entered the multivariate analysis were selected using a univariate pre-selection by means of General Linear Models on the training data. We considered the design matrix obtained considering the auditory post-test used in the training data and ranked the voxels according to their activity in the two auditory post-test conditions (separately). We considered the union of the m most active voxels per condition. The value of m was optimized for each subject, in a range from 100 to 2000, with steps of 100.

We report accuracies for the classification of unseen trials. Since we used balanced subsets for training, we expect a chance level of 50% per class. However, total accuracy (on both classes) may be influenced by the unbalanced nature of the dataset. We therefore

assessed the statistical significance of our results by means of a permutation test (Nichols & Holmes, 2002) in which we performed the classification procedure after scrambling the class labels of the auditory post-test trials. In order to avoid bias from the voxel pre-selection by means of GLM, we first scrambled the auditory post-test behavioral labels, and subsequently selected the voxels on which we performed the analyses. This procedure was repeated 500 times per subject, providing an empirical distribution for the null hypotheses that the obtained accuracies derive from chance. We furthermore considered *precision* and *recall* as additional measures of performance of the classifier. Both these measures are widely used in assessing performances of classifiers on unbalanced data, and they are insensitive of relative group size. Precision for a class is defined as the ratio between true positives and the total number of elements classified as belonging to that class. Recall for a class is defined as the ratio between true positives and the total number of elements actually belonging to that class. It can be shown that the average of both Recall and Precision have a chance level of 50%, regardless of differences in class sizes (Rieger, et al., 2008).

To visualize the spatial activation patterns that were used for classification and to inspect consistency across participants, group discriminative maps, i.e. maps of the cortical locations that contribute most to the discrimination of conditions, were created after cortex-based alignment (Goebel, et al., 2006) of binarized single-subject discriminative maps (see Staeren, Renvall, De Martino, Goebel, & Formisano, 2009). As we used a *linear* support vector machine, it is meaningful - at an individual map level - to relatively rank the features (i.e. voxels) according to their contribution to the discrimination and select the “peaks” through thresholding. Therefore, for binarizing the maps, we selected the 30% threshold level empirically and assessed the spatial (anatomical) correspondence between the “peaks” of the single-subject discriminative maps. In the resulting group-level discriminative maps, a cortical location (vertex) was color-coded if it was present among the 30% of most discriminative vertices in the corresponding individual discriminative maps of at least seven of the eleven subjects. To account for the multiple tests performed in creating these maps, we calculated the proportion of expected false positives (False Discovery Rate) that corresponds to the uncorrected p-value. The *q*-value was obtained using a statistical method that ensures robust estimates also when the distribution of p-values is discrete and one-sided tests are performed (Pounds & Cheng, 2006). The classification accuracy in each subject is always calculated with respect to the whole set of features and, thus, does not depend on the threshold chosen for the creation of the maps.

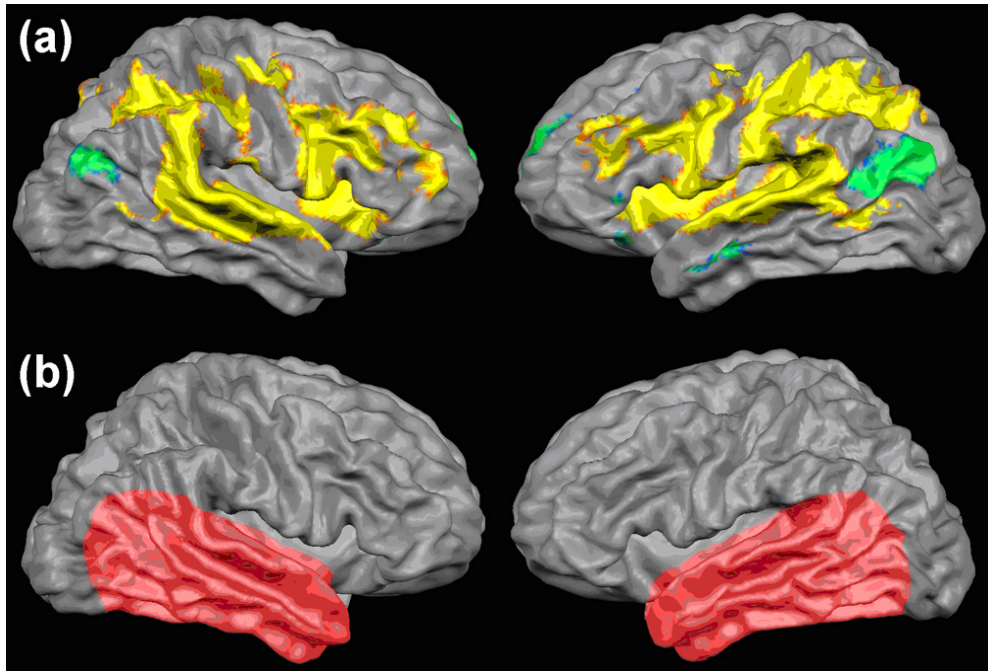


Figure 3. (a) Results of statistical univariate analysis. Overall auditory cortical activation in response to the ambiguous auditory stimuli as estimated with univariate random effects general linear model analysis (statistical parametric F map, group results of cortex-based aligned data sets, $p < 1.7e^{-11}$). The pair-wise statistical comparison between the two perceptual conditions (*/aba/* versus */ada/*) did not yield significant results. (b) Illustration of the anatomical masks used. Classification was anatomically confined to the temporal lobe. The exact boundaries of the anatomical masks were determined on a subject basis in order to account for individual differences in anatomy.

Results

Behavioral results

Participants were presented with individually-determined, ambiguous disyllables (A?, between */aba/* and */ada/*). As discussed before, it has been shown that the perceptual interpretation of these ambiguous sounds can be altered relying on crossmodal recalibration (Bertelson, et al., 2003). We were able to replicate these behavioral results in an fMRI environment. An analysis of variance (ANOVA) demonstrated a highly significant difference in response (*/aba/* vs. */ada/*) between sounds that were preceded by an A?Vb

block versus an A?Vd block ($F = 75.682, p < .001$). The overall proportion of /aba/ responses was $.31 (\pm 0.02)$ if preceded by an A?Vd block and $.53 (\pm 0.02)$ if preceded by an A?Vb block. This significant difference was observed for all three sounds individually (A?: $.25$ vs. $.60, F = 68.692, p < .001$; A?b: $.05$ vs. $.15, F = 31.546, p < .001$; A?d: $.64$ vs. $.85, F = 12.757, p < .001$). Thus, exposure to the videos significantly recalibrated later perceptual interpretation in the auditory post-test trials, such that ambiguous stimuli were more often perceived as /aba/ when preceded by an A?Vb block and vice versa.

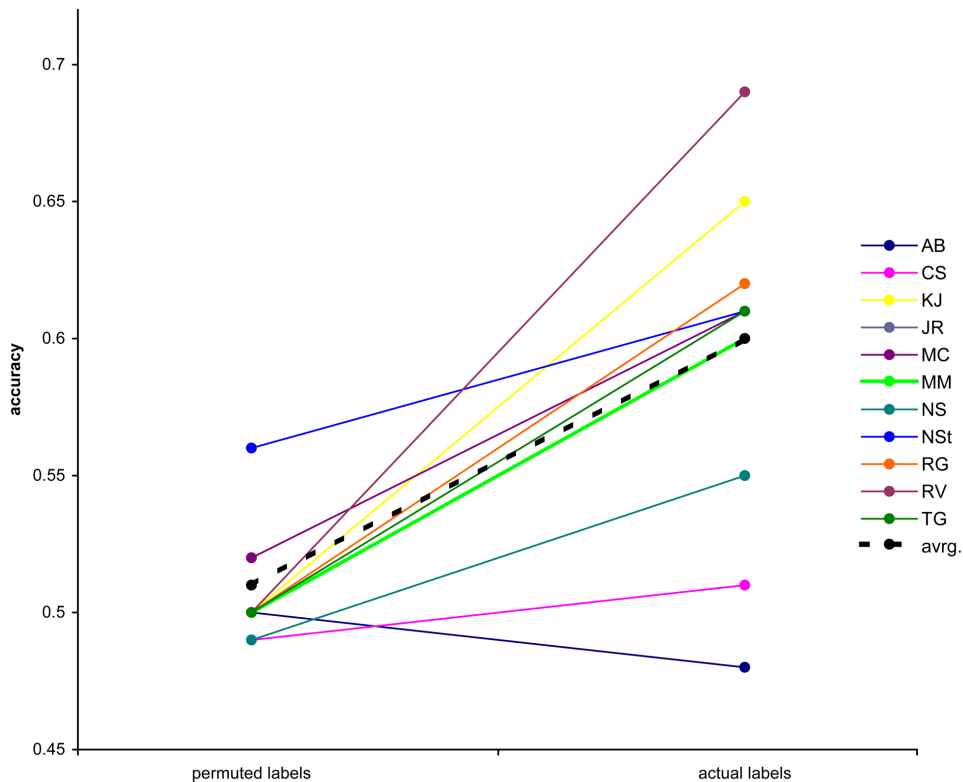


Figure 4. Classification accuracies for individual participants. Accuracies obtained with the actual labels were significantly higher than the accuracies derived from the analysis using permuted class labels ($p < 5.6 \times 10^{-4}$).

Univariate statistical analysis

These post-test trials evoked significant fMRI responses in a network of areas consistent with previous studies on phoneme processing (see Figure 3(a)), including auditory areas in the superior temporal cortex (Heschl's gyrus, multiple regions in the planum temporale

(PT), superior temporal gyrus (STG) and sulcus (STS), as well as middle temporal gyrus, insula cortex, inferior parietal cortex, inferior frontal cortex and supramarginal gyrus (Davis & Johnsrude, 2003; Desai, Liebenthal, Waldron, & Binder, 2008; Hickok & Poeppel, 2007; Liebenthal, et al., 2005). A univariate comparison between trials perceptually classified as /aba/ and those classified as /ada/ did not yield significant differences in activation ($q = 0.05$, corrected for multiple comparisons with False Discovery Rate).

Multivariate pattern recognition analysis

This homogeneity in activation between perceived /aba/ and perceived /ada/ is likely due to the fact that these stimuli differed only in perceptual category, but not in physical stimulus identity. In order to decipher subtle and supposedly distributed differences in activation and to predict the perceptual interpretation on a trial-by-trial basis, we performed multivoxel pattern analysis (MVPA). As our stimuli were physically identical, the fMRI responses to individual auditory post-test trials were labeled according to their perceptual interpretation (/aba/ versus /ada/). The algorithm succeeded in learning the relationship between perceptual labels and corresponding spatial activation patterns and in exploiting this information in order to correctly classify the patterns associated with the remaining unlabeled trials. The mean accuracy of the classifier across the 11 subjects was 0.60. Statistical significance of the original classification was assessed by means of permutation testing (Nichols & Holmes, 2002). We performed a paired t-test on the classification accuracies obtained on the 11 subjects using correct labels and using permuted labels, and we were able to reject the null hypothesis at $p < 5.6 \cdot 10^{-4}$ (see Figure 4). It should be noted that in one subject the mean of the empirical null distribution is different from the theoretical value of 50%. Nonetheless, the results of the permutation tests show that classification accuracies with correct labels are significantly greater than those obtained with scrambled labels, indicating that information on pure perception is present in the selected brain regions and it is efficiently exploited by the learning algorithm. A paired t-test between the values for precision (recall) obtained on the 11 subjects and the ones obtained from the permutation tests confirmed these results, rejecting the null hypothesis at $p < 2.4 \cdot 10^{-3}$ ($p < 2.4 \cdot 10^{-3}$).

The discriminative maps were inspected for the consistency of spatial activation patterns across participants. The main patterns of discriminative voxels revealed by the

group discriminative maps (see Figure 5, threshold corresponds to $q = 2.75 \cdot 10^{-3}$.) were left-lateralized and clustered along the posterior bank of Heschl's gyrus and Heschl's sulcus and, adjacently, in the anterior portion of PT. The exact localization of auditory cortical fields in humans is not a trivial task, since the discrepancies between cytoarchitectonic borders and macroscopical landmarks are profound and vary from brain to brain and between hemispheres (Morosan, et al., 2001). Comparing our data with data from tonotopy mapping experiments (Formisano, et al., 2003), it seems likely that our discrimination maps touch upon A1 and spread across posterior-lateral human belt regions within area PT. However, in-vivo neuroimaging experiments in humans currently do not allow the exact localization of maps in terms of cytoarchitectonic areas and/or single-cell response patterns. Said tonotopy mapping experiments as well as probabilistic cytoarchitectonic maps (Morosan, et al., 2001; Rademacher, et al., 2001) would suggest A1 itself to be centered mostly on the convexity of medial Heschl's gyrus, an area that is mostly absent from our group discriminative maps. Definite assertions remain problematic due to the aforementioned inter-subject variability in macro-anatomical morphology, cytoarchitectonic make-up, and the relation between the two, as well as due to differences in normalization. However, it seems likely that our maps identify mostly non-primary auditory areas. Additional clusters of smaller extent were found at the left temporoparietal junction (TPJ) and, bilaterally, on middle STG and STS.

Discussion

In the present study, we used cross-modal recalibration in order to influence participants' perception of ambiguous speech sounds. We then trained an SVM pattern classifier on fMRI data which was categorized using purely subjective perceptual labels. Our results show that pure perceptual interpretation of physically identical phonemes can be decoded from cortical activation patterns in early auditory areas. More specifically, our findings provide direct empirical evidence that - beyond the basic acoustic analysis of sounds - constructive perceptual information is present in regions within the anterior PT, tangent to the posterior bank of Heschl's gyrus and sulcus.

Concerning speech perception, hierarchical views of auditory processing (Rauschecker & Scott, 2009) suggest a gradient of increasing processing complexity in the anterior superior temporal gyrus (i.e. the 'what' stream), where regions show the first clear responses to abstract, linguistic information in speech. Further, it has been suggested that

“phonetic maps have some anatomical implementation in anterior temporal lobe areas” (Rauschecker & Scott, 2009). Hierarchically lower auditory regions, on the other hand, are allegedly limited to low-level acoustic feature analysis. Our finding that regions in the PT adjacent to, and touching upon, Heschl’s gyrus and sulcus discriminate stimuli on a purely perceptual level, goes beyond such a limited feature-bound processing role. It is, however, in line with single-cell recording studies in monkeys and cats (Micheyl, et al., 2007; Micheyl, Tian, Carlyon, & Rauschecker, 2005; Nelken & Bar Yosef, 2008). For instance, object representations have been suggested to be present even as early as in A1 (Bar-Yosef & Nelken, 2007). Furthermore, Micheyl and colleagues (Micheyl, et al., 2005) demonstrated a strong correspondence between psychophysical findings on auditory stream formation in humans and single-unit responses in rhesus monkeys’ A1, suggesting the possibility that auditory streaming percepts have a representation already in A1 in the absence of stimulus differences. However, comparing human behavior with primate neuronal responses is obviously problematic. Differences in behavior, perception, and neuroanatomy are unavoidable. Our results provide direct empirical evidence in humans that processing in early auditory cortex, probably corresponding to human belt areas, is not limited to low-level stimulus feature analysis.

Although across-study comparisons of the roles of human auditory areas are problematic due to differences in normalization techniques, analysis methods, and even anatomical nomenclature, our findings do seem to be in line with certain previous work. Macroscopically, area PT has been described as a computational hub, which segregates spectrotemporal patterns, compares them to stored patterns, and outputs auditory objects (Griffiths & Warren, 2002), which is compatible with our results. Location and functional role of a smaller part of our clusters in PT is accordant with Spt, a functional subdivision of PT which has been described as a sensory-motor integration region for the vocal tract motor effector (Hickok, Okada, & Serences, 2009; Hickok & Poeppel, 2007). This is in line with the suggestion that PT is part of the dorsal stream of auditory processing (revised to include language besides spatial functions) and possesses the representation of templates for “doable” articulations, effectively disambiguating phonological information (Rauschecker & Scott, 2009). In the present experiment, this sensory-motor function of the posterior PT may serve to disambiguate ambiguous phonemes on the basis of previously seen lip movements. A functional role in cross-modal integration has been suggested before for

similar regions (van Atteveldt, Formisano, Goebel, & Blomert, 2004). The output of an acoustic-phonetic analysis in PT may be probabilistic and represent pre-lexical phonemic categories, as has been suggested (Obleser & Eisner, 2009).

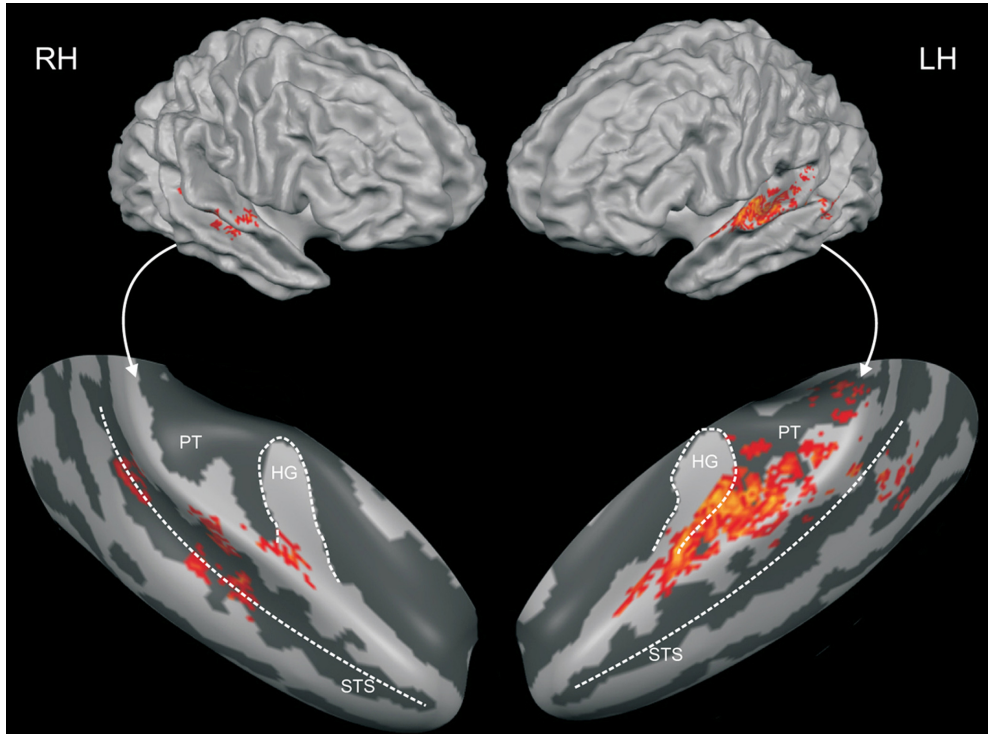


Figure 5. Discriminative map. Group map of the 30% of active voxels most discriminative for the purely perceptual difference between /aba/ and /ada/ (cortex-based aligned, smoothed). A location was color-coded if it was present on the individual maps of at least seven of the eleven subjects. This corresponds to a false discovery rate–corrected threshold of $q = 2.75e^{-3}$. Maps are overlaid on the reconstructions of the average hemispheres of the 11 subjects (upper half) and on inflated reconstructions of the right and left temporal lobes of these average hemispheres. RH = right hemisphere, LH = left hemisphere, PT = Planum Temporale, HG = Heschl's Gyrus, STS = superior temporal sulcus.

In terms of neurocomputational plausibility, our data may be accordant with reverse hierarchy theory, which states that, by default, rapid perception is based on high-level (e.g., phonological) representations alone, which are holistic and ecologically meaningful (Ahissar, Nahum, Nelken, & Hochstein, 2009; Hochstein & Ahissar, 2002).

However, when finer discrimination between similar stimuli is needed, performance relies more on lower level activity (reverse hierarchy). This concept seems to provide an interesting interpretation of our findings. Since the task to categorize ambiguous phonemes demands scrutiny, reverse hierarchy routines might recruit low-level networks. One problem here is, however, that since the stimuli are identical in both perceptual conditions, there are no low-level physical features that would help discrimination. In a way, the holistic, high-level representation is the only level that provides valuable information for this task. We propose that what is reflected in the reliance on early networks in this case is not the intensified analysis of low-level stimulus features, but a perceptual bias which is stored in these regions. This bias is responsible for the behavioral (perceptual) effect and is installed by the cross-modal recalibration mechanism. Its origin may lie within higher-order areas involved in the integration of audiovisual speech signals, such as the supramarginal gyrus (Raizada & Poldrack, 2007), the intraparietal sulcus (Cusack, 2005), or even regions involved in the motor processing of speech (D'Ausilio, et al., 2009). The information is then fed back from there to early auditory areas. Here, without eliciting overall differences in activation strength, the perceptual bias is stored and sensory input is transformed into more abstract entities or auditory objects. These abstract entities may be considered as the building blocks of further linguistic and vocal processing.

References

- Ahissar, M., Nahum, M., Nelken, I., & Hochstein, S. (2009). Reverse hierarchies and sensory learning. *Philos Trans R Soc Lond B Biol Sci*, 364(1515), 285-299.
- Bar-Yosef, O., & Nelken, I. (2007). The effects of background noise on the neural responses to natural sounds in cat primary auditory cortex. *Front Comput Neurosci*, 1, 3.
- Bertelson, P., Vroomen, J., & De Gelder, B. (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychol Sci*, 14(6), 592-597.
- Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *J Cogn Neurosci*, 17(4), 641-651.

- D'Ausilio, A., Pulvermuller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Curr Biol*, *19*(5), 381-385.
- Davis, M. H., & Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *J Neurosci*, *23*(8), 3423-3431.
- De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., & Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *Neuroimage*, *43*(1), 44-58.
- Desai, R., Liebenthal, E., Waldron, E., & Binder, J. R. (2008). Left posterior temporal regions are sensitive to auditory categorization. *J Cogn Neurosci*, *20*(7), 1174-1188.
- Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science*, *322*(5903), 970-973.
- Formisano, E., Kim, D. S., Di Salle, F., van de Moortele, P. F., Ugurbil, K., & Goebel, R. (2003). Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron*, *40*(4), 859-869.
- Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: characterizing differential responses. *Neuroimage*, *7*(1), 30-40.
- Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp*, *27*(5), 392-401.
- Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends Neurosci*, *25*(7), 348-353.
- Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., & Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *J Neurosci*, *25*(22), 5382-5388.
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*(5539), 2425-2430.
- Haynes, J. D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci*, *8*(5), 686-691.

- Hickok, G., Okada, K., & Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J Neurophysiol*, *101*(5), 2725-2732.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nat Rev Neurosci*, *8*(5), 393-402.
- Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791-804.
- Jancke, L., Wustenberg, T., Scheich, H., & Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage*, *15*(4), 733-746.
- Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cereb Cortex*, *15*(10), 1621-1631.
- Micheyl, C., Carlyon, R. P., Gutschalk, A., Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., et al. (2007). The role of auditory cortex in the formation of auditory streams. *Hear Res*, *229*(1-2), 116-131.
- Micheyl, C., Tian, B., Carlyon, R. P., & Rauschecker, J. P. (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron*, *48*(1), 139-148.
- Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage*, *13*(4), 684-701.
- Nelken, I., & Bar Yosef, O. (2008). Neurons and objects: the case of auditory cortex. *Front Neurosci*, *2*(1), 107-113.
- Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp*, *15*(1), 1-25.
- Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends Cogn Sci*, *13*(1), 14-19.
- Pounds, S., & Cheng, C. (2006). Robust estimation of the false discovery rate. *Bioinformatics*, *22*(16), 1979-1987.
- Rademacher, J., Morosan, P., Schormann, T., Schleicher, A., Werner, C., Freund, H. J., et al. (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage*, *13*(4), 669-683.

- Raizada, R. D., & Poldrack, R. A. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, *56*(4), 726-740.
- Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci*, *12*(6), 718-724.
- Riecke, L., Esposito, F., Bonte, M., & Formisano, E. (2009). Hearing illusory sounds in noise: the timing of sensory-perceptual transformations in auditory cortex. *Neuron*, *64*(4), 550-561.
- Rieger, J. W., Reichert, C., Gegenfurtner, K. R., Noesselt, T., Braun, C., Heinze, H. J., et al. (2008). Predicting the recognition of natural scenes from single trial MEG recordings of brain activity. *Neuroimage*, *42*(3), 1056-1068.
- Scott, S. K., & Johnsrude, I. S. (2003). The neuroanatomical and functional organization of speech perception. *Trends Neurosci*, *26*(2), 100-107.
- Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Curr Biol*, *19*(6), 498-502.
- Sterzer, P., Kleinschmidt, A., & Rees, G. (2009). The neural bases of multistable perception. *Trends Cogn Sci*, *13*(7), 310-318.
- van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*(2), 271-282.
- van Atteveldt, N. M., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex*, *17*(4), 962-974.
- Vapnik, V. N. (1995). *The nature of statistical learning theory*: Springer-Verlag New York, Inc.

**BRAIN ACTIVATION DURING AUDIOVISUAL EXPOSURE
ANTICIPATES FUTURE PERCEPTION OF AMBIGUOUS SPEECH**

CORRESPONDING PUBLICATION

KILIAN-HUTTEN, N., VROOMEN, J., AND FORMISANO, E. (2011). BRAIN ACTIVATION DURING AUDIOVISUAL EXPOSURE ANTICIPATES FUTURE PERCEPTION OF AUDITORY SPEECH. *NEUROIMAGE*, 57(4):1601-7.

Abstract

In modern perceptual neuroscience, the focus of interest has shifted from a restriction to individual modalities to an acknowledgement of the importance of multisensory processing. One particularly well-known example of cross-modal interaction is the McGurk illusion. It has been shown that this illusion can be modified, such that it creates an auditory perceptual bias that lasts beyond the duration of audiovisual stimulation, a process referred to as cross-modal recalibration (Bertelson et al., 2003). Recently, we have suggested that this perceptual bias is stored in auditory cortex, by demonstrating the feasibility of retrieving the subjective perceptual interpretation of recalibrated ambiguous phonemes from functional magnetic resonance imaging (fMRI) measurements in these regions (Kilian-Hütten, Valente, Vroomen, & Formisano, 2011). However, this does not explain which brain areas integrate the information from the two senses and represent the origin of the auditory perceptual bias.

Here we analyzed fMRI data from audiovisual recalibration blocks, utilizing behavioral data from perceptual classifications of ambiguous auditory phonemes that followed these blocks later in time. Adhering to this logic, we could identify a network of brain areas (bilateral inferior parietal lobe [IPL], right [and possibly left] inferior frontal sulcus [IFS], and right posterior middle temporal gyrus [MTG]), whose activation during audiovisual exposure anticipated auditory perceptual tendencies later in time.

We propose a model in which a higher-order network, including IPL and IFS, accommodates audiovisual integrative learning processes, which are responsible for the installation of a perceptual bias in auditory regions. This bias then determines constructive perceptual processing.

Introduction

Traditionally, research on human perception has focused on unimodal information processing. However, in real life we are constantly confronted with multimodal input, i.e. information coming from several/all senses. Importantly, multimodal processing is not restricted to the mere addition of a second channel. Instead, these channels interact, affecting processing in individual modalities. For instance, in face-to-face spoken conversation, lip reading provides for vital information, enhancing auditory understanding. This is especially true in suboptimal acoustic settings. Furthermore, research has consistently demonstrated intermodal effects on behavior (Shimojo & Shams, 2001). Congruent multisensory input typically lowers detection thresholds (Frassinetti, Bolognini, & Ladavas, 2002), shortens reaction times (Forster, Cavina Pratesi, Aglioti, & Berlucchi, 2002; Schroger & Widmann, 1998), and decreases saccadic eye movement latencies (Hughes, Reuter Lorenz, Nozawa, & Fendrich, 1994). Incongruent multisensory input has opposite consequences (Sekuler, Sekuler, & Lau, 1997).

One particularly interesting effect of audiovisual interaction is the McGurk illusion (McGurk & MacDonald, 1976). Here the addition of the visual modality does not only affect the quality of the auditory percept, but it actively alters its identity. Typically, an auditory disyllable (e.g., /aba/) uttered by a speaker is presented simultaneously with a video showing the speaker's lips pronouncing an incongruent visual disyllable (/aga/). The integration of this bimodal input results in a percept that is intermediate between the auditory and the visual channels (/ada/). Neuroimaging studies investigating the McGurk illusion have identified networks very similar to the ones found with other audiovisual paradigms (Beauchamp, Nath, & Pasalar, 2010; Benoit, Raij, Lin, Jaaskelainen, & Stufflebeam, 2010; Fingelkurts, Fingelkurts, Krause, Mottonen, & Sams, 2003; Hertrich, Dietrich, & Ackermann, 2011; Jones & Callan, 2003; Jones & Jarick, 2006; Mottonen, Krause, Tiippana, & Sams, 2002; Sekiyama, Kanno, Miura, & Sugita, 2003; Wiersinga Post, et al., 2010).

It has been shown that the McGurk illusion can be modified in such a way that the biasing perceptual effect lasts beyond the audiovisual stimulation itself (Bertelson, Vroomen, & De Gelder, 2003). The decisive factor for this is the use of an ambiguous auditory component (A?, between /aBa/ and /aDa/). Exposure to such an ambiguous stimulus dubbed onto a video of a face pronouncing /aBa/ or /aDa/ (A?Vb and A?Vd, respectively) selectively

increases the proportion of corresponding responses in subsequent audio-only forced-choice trials, thus eliciting an after-effect (“cross-modal recalibration”). In the original study, an audiovisual exposure block of 8 adapters (i.e., identical videos) sufficed to significantly bias the perceptual categorization of 6 subsequent auditory post-tests.

Recently, we have demonstrated that it is possible to retrieve the perceptual interpretation of the ambiguous phonemes presented in these post-tests from functional magnetic resonance imaging (fMRI) measurements of brain activity in auditory areas in the superior temporal cortex, most prominently on the posterior bank of the left Heschl’s gyrus and sulcus and in the adjoining left planum temporale (Kilian-Hütten, et al., 2011). We have proposed that the perceptual bias, which is responsible for the behavioral effect, is installed by the cross-modal recalibration mechanism and is stored within these regions. However, this leaves open the question where the top-down audiovisual recalibration effect originates, i.e. which networks are responsible for prejudicing auditory cortex in creating perceptual interpretations.

The goal of the present study was to investigate the neural origin of cross-modal recalibration, employing McGurk after-effects. We analyzed fMRI measurements from the same data set as used by Kilian-Hütten and colleagues (2011). However, instead of looking at data from the auditory post-tests, here we concentrated on the audiovisual recalibration blocks. We expected a simple comparison between the activation elicited by the audiovisual stimuli and baseline to replicate the results from prior studies on audiovisual processing and the McGurk illusion and to identify a set of areas similar to the ones found in these studies. However, our recalibration design biases perception beyond the time of audiovisual stimulation, and thus provided us with the unique possibility to go beyond the mere identification of active areas. In order to determine which subset of active areas drives the lasting biasing effect on auditory perception, we applied a behaviorally-defined contrast. The strength of the recalibration effect is variable from one particular exposure block to another and can be quantified in terms of the number of auditory post-tests perceived in line with the type of exposure block (A?Vb or A?Vd). We could then use these values to identify brain regions whose activation during the exposure blocks varied with the strength of the recalibration effect. The hemodynamic response in these areas, thus, ‘predicted’ the perceptual tendency in a separate, later time window. Hence, this approach provides the possibility to go beyond the basic identification of responsive regions in an audiovisual paradigm and allows identifying those regions which are responsive to audiovisual

stimulation and which, further, are likely functionally relevant in driving the effect of these stimuli in biasing subsequent auditory perception. In order to further investigate whether the ROIs identified with the behaviorally-weighted contrast exert a task-related influence on auditory regions, which are involved in the perceptual interpretation of the auditory post-tests (Kilian-Hütten, et al., 2011), we carried out a psychophysiological interaction analysis (PPI; Friston, et al., 1997). PPI is a method to investigate how the influence of the activity in one region of the brain on the activity of spatially distant regions changes depending on the task context.

Methods

Participants

Twelve healthy native Dutch students of the University Maastricht (5 male, mean age 24.83 years) were recruited to participate in the study. All but one participant were right-handed. None of the participants had a history of hearing loss or neurological abnormalities. Approval for the study was granted by the Ethical Committee of the Faculty of Psychology at the University of Maastricht.

Stimuli

Based on the psychophysical study by Bertelson, Vroomen, & De Gelder (2003), the stimulation entailed digital auditory and visual recordings of a male Dutch speaker pronouncing the syllables /aBa/ and /aDa/. The two auditory stimuli were 640-ms stimuli with 240-ms stop closure. From these, a place-of-articulation continuum was synthesized by means of varying the F2 formant by equal steps of 39 Mel, resulting in 9 different stimuli, ranging from a clear /aba/ via seven ambiguous stimuli to a clear /ada/. The audiovisual stimuli were synthesized by pairing the visual recordings of the speaker pronouncing /ada/ and /aba/, respectively, with the most ambiguous auditory stimulus, determined as such in a pre-test (please see the online supplementary material for example stimuli).

Behavioral pre-test

In an auditory pre-test outside the fMRI scanner the most ambiguous auditory stimulus from the /aBa/-to-/aDa/ continuum was determined for each participant individually

(Bertelson, et al., 2003). All stimuli on the continuum were presented (98 trials in total). The presentation frequency was biased, so that the five central stimuli were presented on 14 trials, the second and tenth stimuli were presented on 8 trials, and the first and the eleventh stimuli were presented on 6 trials. Stimuli were presented binaurally through loudspeakers. Participants indicated whether they had perceived /ada/ or /aba/ by means of a button press. We estimated each participant's most ambiguous auditory token ($A?$), which was then used in the actual experiment.

Scanning parameters

We collected functional brain data on a 3-Tesla fMRI scanner (head set-up, Siemens Allegra, Siemens, Erlangen, Germany) at the Maastricht Brain Imaging Center (M-BIC) in Maastricht, The Netherlands. For each participant, 2 functional runs of 665 volumes were acquired. A high-resolution structural scan (voxel size = $1 \times 1 \times 1 \text{ mm}^3$) was collected using a T1-weighted 3D ADNI sequence (TR = 2050 ms, TE = 2.6 ms, 192 sagittal slices) for later overlay of the functional data. All scans were carried out in a single session for each participant.

For functional images, a BOLD-sensitive echo-planar imaging (EPI) sequence was used (matrix 64×64 , 27 slices, slice thickness 3 mm, field of view, FOV, $192 \times 192 \text{ mm}^2$, resulting voxel size $3 \times 3 \times 3 \text{ mm}^3$, TE/TR slice 30/55.5, FA = 90°). Volume acquisition was clustered in the beginning of each TR, leaving a silent delay within each TR, during which stimuli were presented in the absence of EPI noise. This was done in order to optimize stimulus audibility, an approach which has been shown to be highly efficient in auditory fMRI paradigms (Jancke, Wustenberg, Scheich, & Heinze, 2002; N. van Atteveldt, Formisano, Goebel, & Blomert, 2004; N. M. van Atteveldt, Formisano, Blomert, & Goebel, 2007). Hence, the effective TR was 2000 ms, including 1500 ms of sequence scanning time and a 500 ms silent delay. Stimuli were presented and synchronized with the MR pulses using the software package "Presentation" (Neurobehavioral Systems).

Experimental procedure

Each functional run consisted of 10 mini runs, which were composed of bimodal exposure blocks and slow event-related auditory post-tests. The bimodal (recalibration) blocks entailed the repeated presentation of videos, which consisted of the participant's $A?$ paired with the visual recording of the speaker's lips pronouncing either /aba/ (Vb) or /ada/ (Vd).

Each block was composed of 8 identical videos (block A?Vb or block A?Vd). The interstimulus interval (ISI) was one TR (= 2000 ms). For each run, 5 A?Vb and 5 A?Vd blocks were presented in random order (160 trials in total). In order to warrant attentional focus, participants were instructed to press a button whenever a small white dot (12 pixels) appeared on the speaker's upper lip. This was the case once per block at a random position. Each bimodal recalibration block was followed by 6 auditory-only post-tests (total: 120 trials), which consisted of forced-choice /aba/-/ada/ judgments. The A? token and the two tokens closest to it on the continuum were presented, twice each. Owing to the use of a slow event-related design, the jittered ISI was 6 TR (= 12 s) on average.



Figure 1. Schematic overview of one example mini run (A?Vb). Each run consisted of 10 such mini runs, which each entailed a bimodal exposure block (A?Vd or A?Vb) and 6 auditory post-test trials. The bimodal blocks were presented within 8 TRs, each of which entailed 1500 ms of scanning and 500 ms of video presentation. Between auditory post-test trials, fixation periods averaged 6 TRs.

Data pre-processing

Functional and anatomical images were analyzed using BrainVoyager QX (Brain Innovation, Maastricht, The Netherlands). Several pre-processing steps were performed: sinc interpolated slice-time correction, 3D-motion correction to correct for common small head movements by spatially aligning all volumes to the first volume by rigid body transformations, linear trend removal, and temporal high-pass filtering to remove low-frequency nonlinear drifts of 7 or less cycles per time course. Functional slices were then co-registered to the structural volume on the basis of positioning parameters from the scanner and manual adjustments to ensure optimal fit. Subsequently, they were transformed into Talairach Space. All individual brains were segmented at the gray/white matter boundary using a semi-automatic procedure based on intensity values implemented in BrainVoyager QX. For all group analyses, cortex-based alignment was used to assure optimized spatial matching of cortical locations (i.e., vertices) between participants (Goebel, Esposito, & Formisano, 2006).

Statistical analysis

Functional runs were analyzed using voxel-wise multiple linear regression (GLM) of the BOLD-response time course. All analyses were performed at single-subject and group levels and all experimental conditions were modeled as predictors. In all GLMs, predictor time courses were convolved with a hemodynamic response function (where both the BOLD response and undershoot are modeled by a gamma function) to adjust for the hemodynamic delay (Friston, et al., 1998).

Subject	Talairach (center of gravity)		
	x	y	z
AB	-38	-34	15
AF	-53	-26	14
CS	-43	-29	10
JR	-44	-23	5
KJ	-39	-31	14
MC	-51	-24	5
MM	-37	-33	15
NS	-44	-27	10
NSt	-43	-21	8
RG	-49	-23	6
RV	-39	-32	12
TG	-43	-30	5

Table 1. Seed regions for the PPI analysis. ROIs were identified individually in each subject on the basis of overall audiovisual block activation. The average number of voxels per ROI was 443 (SD = 134).

Linear contrasts served to implement comparisons of interest. To identify brain regions involved in the processing of the audiovisual (recalibration block) stimuli, a contrast of audiovisual stimulation (all recalibration blocks) versus baseline was applied. In order to find a subset of these areas whose activation during the recalibration blocks would predict post-test performance / perception, a contrast based on behavioral performance was then utilized. To achieve this, we implemented a predictor, in which each recalibration block was coded with an individual, behaviorally defined value. These values were obtained by identifying the number of responses in line with recalibration from the six auditory post-tests belonging to a specific block (i.e., /ada/ after a A?Vd block and /aba/ after an A?Vb block). These values between zero and six were then normalized across all blocks of one type for a given participant and orthogonalized with respect to the overall activation

contrast. We then applied this behaviorally defined contrast in order to identify ROIs whose activation pattern was correlated with post-test performance. ROIs exhibiting the strongest significance for this contrast, thus, demonstrated a BOLD response pattern, whose modulation on a block level best predicted *subsequent* behavioral / perceptual performance. These statistical analyses were performed following a two-step procedure; first, contrasts were implemented on the individual level. These contrasts were then entered into a second-level, random effects model in order to allow for inferences at the population level (Friston, Holmes, Price, Buchel, & Worsley, 1999). In order to correct for false positives, a cluster-size threshold procedure was employed (Forman, et al., 1995; Goebel, et al., 2006). For this, cluster-level false-positive rates were estimated iteratively (500 iterations; Monte Carlo simulation). The minimum cluster-size threshold (based on its observed relative frequency) which yielded a cluster-level false-positive rate (alpha) of 5% was then applied to the statistical maps

Psychophysiological Interaction

For the PPI analysis (Friston, et al., 1997), we identified a seed region in the left auditory cortex (IAC) individually in all twelve subjects based on the [audiovisual blocks > baseline] contrast (see Table 1). This was done following Kilian-Hütten and colleagues, 2011, who identified a distributed pattern of voxels that encoded the perceptual interpretation of the auditory post-test and was located predominantly in left auditory cortex. We then investigated the functional connectivity of the remaining voxels with IAC during the audiovisual exposure blocks compared to baseline. This was done by employing a GLM in which we implemented a predictor for the activity time course in the seed region and one for the interaction between the task (audiovisual blocks) and the time course of the seed (PPI). This latter predictor, thus, served to identify regions which exhibited increased functional connectivity with the seed region during the audiovisual blocks relative to baseline. We also included all other predictors (auditory post-tests) in the analysis, in order to account for common effects and only identify those effects that were unique to the PPI. Group maps were created by running this analysis as a random-effects GLM on the cortex-based aligned data and thresholding the results using a cluster-size thresholding approach such as described above. Additionally, the same analyses were implemented as random-effects ROI GLMs in the ROIs identified by the behaviorally-weighted contrast in order to

investigate whether these regions exhibit increased connectivity during the recalibration blocks.

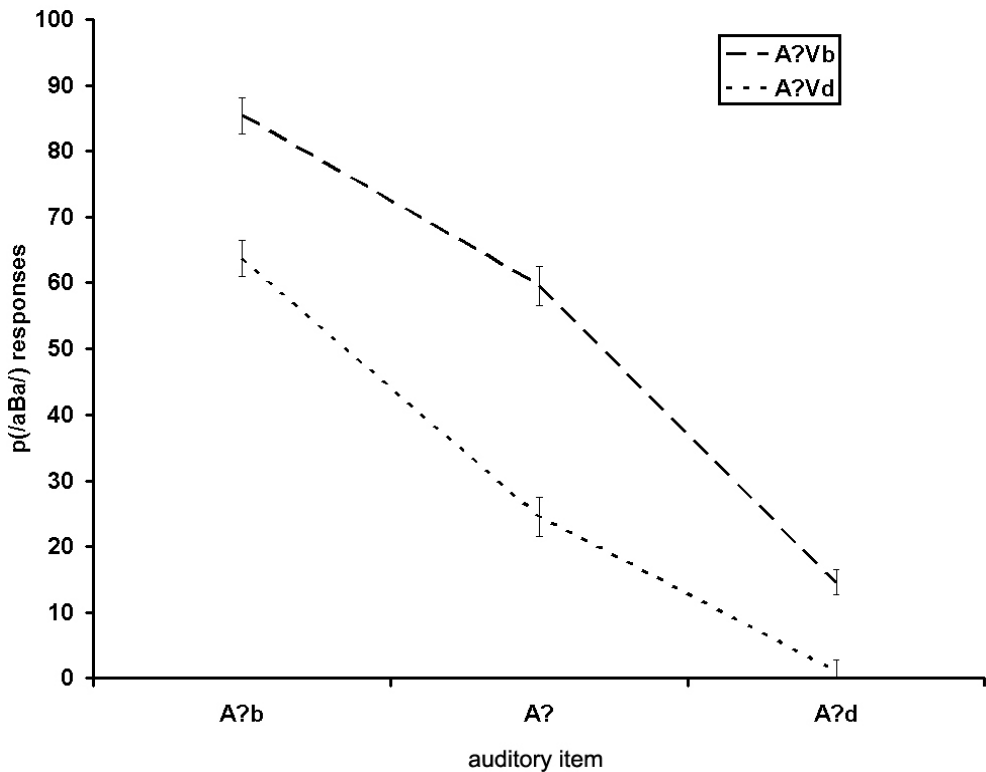


Figure 2. Behavioral results of the auditory post-tests. For the participant’s ambiguous auditory item and its two neighbors on the continuum, the graph shows the proportion of /aba/ classifications after exposure to an audiovisual adapter comprised of the ambiguous item paired with either visual /aba/ (A?Vb) or with visual /ada/ (A?Vd). For all three auditory items, the difference in proportion /aba/ responses after exposure to the A?Vb adapter vs. the A?Vd adapter was significant.

Results

Behavioral results

Individual ambiguous sounds chosen on the basis of the pre-tests were item 4 for eight participants and item 5 for the four others. Exposure to the audiovisual recalibration blocks significantly influenced the perception of the auditory post-test stimuli (see figure 2 and Kilian-Hütten et al., 2011). An analysis of variance (ANOVA) over all sounds (A?, A?d,

A?b) demonstrated a highly significant difference in proportion of /aBa/ responses between sounds that were preceded by an A?Vd block versus an A?Vb block (.53 following A?Vb versus .31 following A?Vd, $F = 75.682$, $p < .001$). This significant difference could be shown for all three sounds separately (A?: .60 following A?Vb versus .25 following A?Vd, $F = 68.692$, $p < .001$; A?b: .85 following A?Vb versus .64 following A?Vd, $F = 31.546$, $p < .001$; A?d: .15 following A?Vb versus .05 following A?Vd, $F = 12.757$, $p < .001$). These results demonstrate the effectiveness of the cross-modal recalibration design in biasing subsequent auditory perception and replicate the results from Bertelson, Vroomen, and De Gelder (2003).

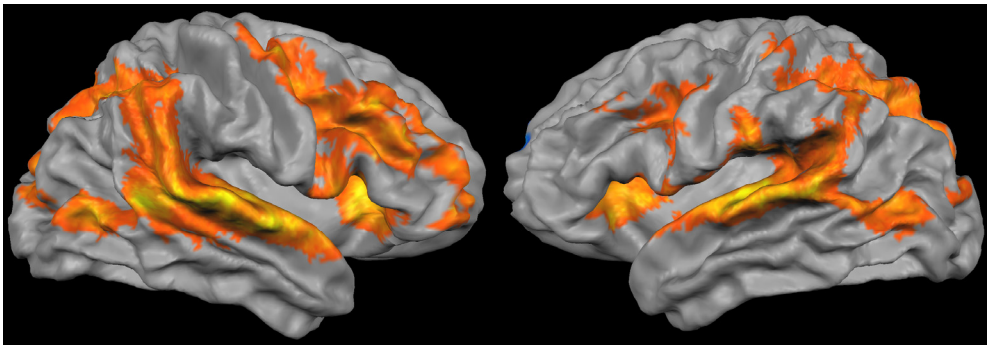


Figure 3. Group results for the random-effects GLM and the contrast [AV > baseline] overlaid on the average hemispheres obtained from the cortex-based alignment procedure (cluster-size corrected to include only clusters of 168 mm² or more, corresponding to a cluster-level false-positive rate (alpha) of 5%). Shown are early auditory and visual areas, STG/STS, insula, IFS and IFG, IPL and premotor cortex.

FMRI results

The first contrast of interest was a comparison of activity in response to the audiovisual blocks versus baseline ([AV > baseline]). Figure 3 shows an overview of brain areas active in response to this audiovisual stimulation. Among these are early unimodal areas (primary and extrastriate visual areas and primary and early auditory areas), as well as areas which have been suggested to be involved in audiovisual integration (STG/STS, inferior parietal lobe and insula cortex), and higher-order areas, possibly involved in attention, action understanding and intellectualization processes (inferior frontal gyrus [IFG], inferior frontal sulcus [IFS] and premotor cortex). The activation pattern is very symmetrical with respect

to hemisphere, except for the frontal activation, which seems to be stronger in the right hemisphere. The activation map obtained from the behaviorally-weighted contrast is shown in figure 4 superimposed onto the average hemispheres obtained from the cortex-based alignment procedure. Activation peaks are listed in Table 2 along with their Talairach coordinates and their z and p values for the behaviorally-weighted contrast. This analysis highlights very symmetric areas of robust activation in the right and left inferior parietal lobe (IPL), approximately corresponding to Brodmann's area (BA) 40. Additionally, we found ROIs in the inferior frontal sulcus (IFS, BA 44) and the posterior middle temporal gyrus (MTG, BA 21), in both hemispheres.

Brain region	Brodmann area	Side	Talairach (center of gravity)			behaviorally weighted contrast	
			x	y	z	z	p
Inferior parietal lobe	39/40	right	41	-59	36	2.627	< .01
Inferior parietal lobe	39/40	left	-38	-60	36	2.916	< .01
Middle temporal gyrus	21	right	57	-29	-6	2.935	< .01
Middle temporal gyrus	21	left	-55	-34	-8	2.921	< .01
Inferior frontal sulcus	09/46	right	42	12	30	2.390	< .05
Inferior frontal sulcus	09/46	left	-43	32	12	2.834	< .01

Table 2. ROIs obtained from the behaviorally-weighted contrast in the cortex-based aligned group analysis, along with the spatially closest Brodmann area, Talairach coordinates (center of gravity) and the associated z and p values for the behaviorally-weighted contrast (random-effects GLM).

The relation between activation level and strength of the recalibration effect is further illustrated in figure 5. Percent BOLD signal changes in blocks that strongly recalibrated later perception (4-6 auditory post-tests in line with the direction of recalibration) were higher than in those blocks that did not (0-3 auditory post-tests in line with the direction of recalibration) for the six ROIs, identified in individual participants.

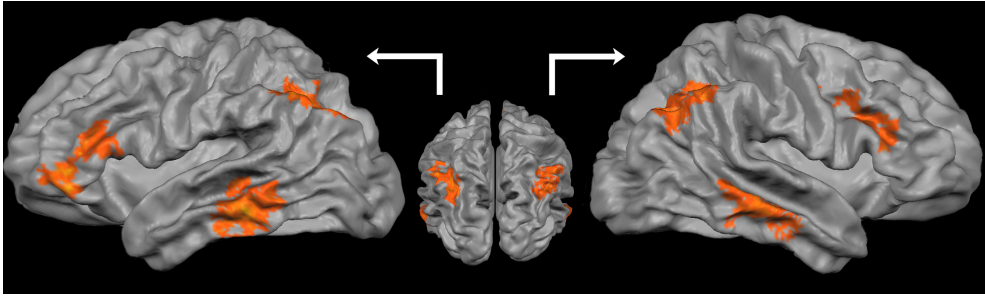


Figure 4. Group results for the behaviorally-weighted contrast (random-effects GLM of cortex-based aligned data) overlaid on the average hemispheres obtained from the cortex-based alignment procedure (cluster-size corrected to include only clusters of 174 mm² or more, corresponding to a cluster-level false-positive rate (alpha) of 5%). Shown are bilateral IPL, IFS, and posterior middle temporal gyrus.

Brain region	Side	psychophysiological interaction	
		z	p
Inferior parietal lobe	right	3.146	< .01
Inferior parietal lobe	left	3.814	< .001
Middle temporal gyrus	right	2.762	< .01
Middle temporal gyrus	left	2.672	< .01
Inferior frontal sulcus	right	3.268	< .01
Inferior frontal sulcus	left	1.949	< .052

Table 3. Results from the psychophysiological interaction ROI GLMs for ROIs obtained from the behaviorally-weighted contrast. Z and p values are shown.

PPI results

The group map from the PPI analysis is shown in figure 6. Although this map includes a broader network than the one obtained with the behaviorally-weighted contrast, it bears quite a clear resemblance with it, including bilateral IPL, IFS and MTG (although there is no perfect overlap between ROIs). Additionally, it includes dorsal and ventral stream visual

areas in the occipital lobe. The ROI GLM analyses corroborate the impression of resemblance and demonstrate a significant increase of connectivity during the recalibration blocks for all ROIs obtained with the behaviorally-weighted contrast, except for left IFS, which just did not reach significance (Table 3).

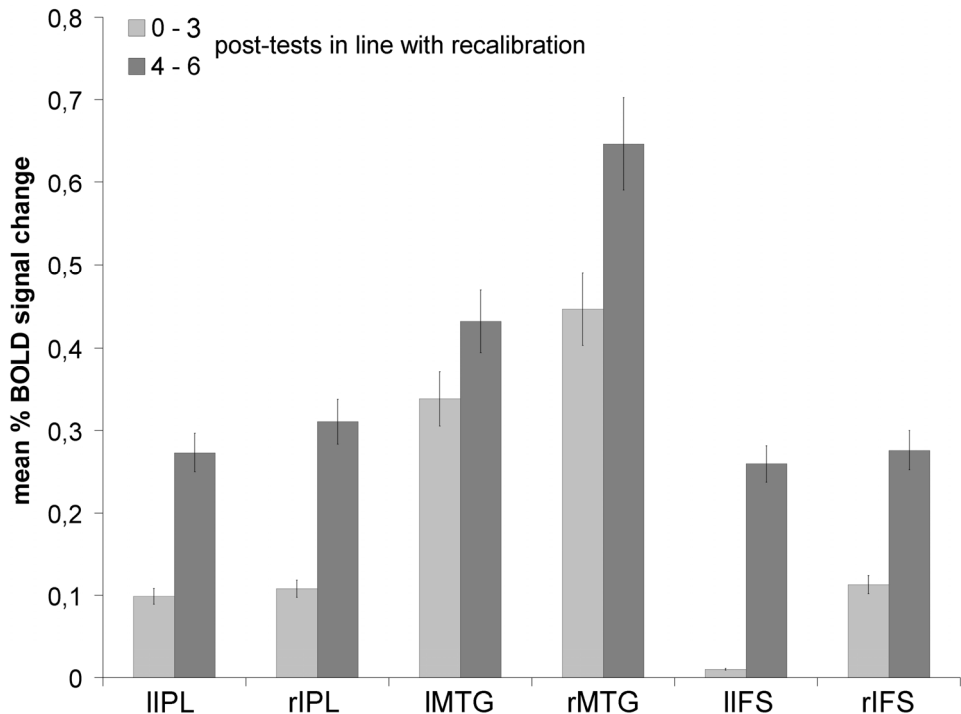


Figure 5. BOLD activation is shown for weak versus strong recalibration blocks in several ROIs (IPL, IFS and MTG; in the left and right hemisphere, respectively) identified individually in each participant on the basis of the behaviorally-weighted GLM contrast. Shown is the mean percentage of BOLD signal change in blocks that strongly recalibrated later auditory perception (4-6 post-tests in line with the direction of recalibration) versus those that did not (0-3 post-tests in line with the direction of recalibration). Error bars denote ± 1 standard error.

Discussion

The goal of the present study was to investigate the neural mechanisms underlying audiovisual recalibration of auditory phoneme perception. In order to identify brain regions which were involved in the processing of the audio-visual stimuli employed in this study,

we contrasted blocks of videos against baseline. A broad, rather symmetric network of brain areas showed a significant increase in BOLD response during these blocks. ROIs included primary and extrastriate visual areas, primary and early auditory areas, STG/STS, inferior frontal sulcus (IFS), premotor cortex, and inferior parietal lobe, spreading out to touch angular and supermarginal gyri and the intraparietal sulcus (jointly referred to as IPL from now on). These results replicate earlier neuroimaging studies on cross-modal processing and the McGurk effect (Beauchamp, et al., 2010; Benoit, et al., 2010; Calvert, 2001; Calvert & Campbell, 2003; Calvert, Campbell, & Brammer, 2000; Calvert & Thesen, 2004; Fingelkurts, et al., 2003; Hertrich, et al., 2011; Jones & Callan, 2003; Jones & Jarick, 2006; Meienbrock, Naumer, Doehrmann, Singer, & Muckli, 2007; Mottonen, et al., 2002; Nishitani & Hari, 2002; Ojanen, et al., 2005; Olson, Gatenby, & Gore, 2002; Saito, et al., 2005; Sekiyama, et al., 2003; Wiersinga Post, et al., 2010; Wright, Pelphrey, Allison, McKeown, & McCarthy, 2003). The activation in early unimodal regions (due to auditory and visual stimulation, respectively) and in premotor cortex related to the button press response) is in line with expectations, as is the activation in the superior temporal lobe. These regions (STS/STG) have consistently been found to be implicated in the integration of audiovisual, and more specifically, face-voice stimuli (Calvert, 2001; Calvert & Campbell, 2003; Calvert, et al., 2000; Calvert & Thesen, 2004; Jancke, et al., 2002; Olson, et al., 2002; Sekiyama, et al., 2003; Wright, et al., 2003). At least two recent studies corroborate this view by demonstrating that these regions follow the principle of inverse effectiveness in humans (Stevenson & James, 2009; Werner & Noppeney, 2010) and one study demonstrates that transcranial magnetic stimulation over STS disrupts the perception of the McGurk illusion (Beauchamp, et al., 2010). Although we did not specifically test for this, our results are compatible with the STS/STG area's suggested role in the integration of audiovisual stimuli.

The remaining regions are located in higher-order association cortex, namely IPL and IFS. These are the areas that seem to be functionally relevant in driving the effect of audiovisual recalibration of auditory phoneme perception. In order to identify the subset of active regions for which this is true, we made use of participants' online behavior (i.e., the efficiency of the individual blocks of audiovisual stimulation in eliciting the recalibration effect). This allowed identifying those areas whose hemodynamic response profile predicted the strength of the recalibration effect on a block-to-block basis. Since this

reasoning relies on a correlation between hemodynamic response and participants' behavior, it does not permit the formulation of definite causal claims. However, the fact that the modulation of the hemodynamic response in these areas considerably preceded the auditory stimuli and forced-choice perceptual judgments in time, does strongly suggest a functional role in the audiovisual recalibration of auditory phoneme perception. This is further corroborated by the results from the psychophysiological interaction analysis, which show that this network of areas exhibits increased functional/effective connectivity with the left auditory cortex during recalibration blocks relative to baseline. It has to be noted that the present paradigm does not lend itself easily to more complex connectivity analysis approaches such as dynamic causal modeling or Granger causality mapping. This is mainly due to the fact that the encoding of the perceptual bias is represented in distributed patterns (which we showed before using multivoxel pattern analysis (Kilian-Hütten, et al., 2011)), while the ROIs in the present analysis were identified with a univariate approach. We are not aware of any past demonstrations of complex connectivity analyses in this situation and have, hence, restricted ourselves to the more straightforward PPI analysis. It is a crucial challenge for future research to devise functional and effective connectivity analysis techniques that integrate univariate and multivariate results. The following section explores our interpretation of what the functional role of the identified network might be. Subsequently, alternative explanations will be discussed.

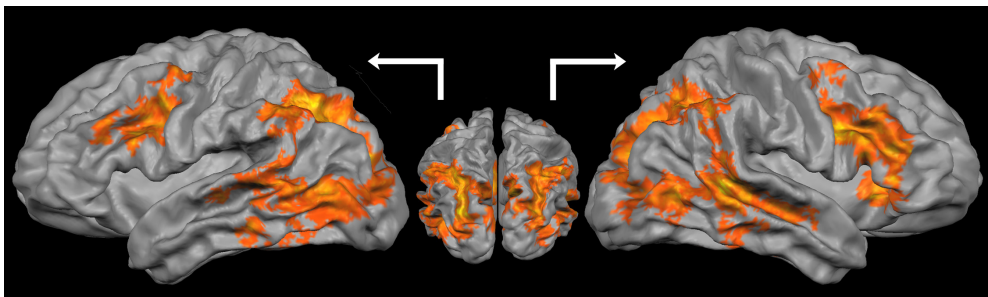


Figure 6. Group results for the psychophysiological interaction analysis (random-effects GLM of cortex-based aligned data) overlaid on the average hemispheres obtained from the cortex-based alignment procedure (cluster-size corrected to include only clusters of 430 mm² or more, corresponding to a cluster-level false-positive rate (alpha) of 5%). Shown are bilateral IPL, IFS, posterior middle temporal gyrus, and dorsal and ventral stream visual areas in the occipital lobe.

Integrative learning effects on constructive perceptual processes

The very symmetrical ROIs in the IPL are the most robust regions obtained from the behaviorally-weighted contrast. One possible functional explanation relates to the brain's pervasive ability and need to categorize incoming information. A recent study (Raizada & Poldrack, 2007) investigated the selective amplification of phonetic stimulus differences. Stimuli from a phonetic continuum were presented in pairs. In some brain regions the activation for pairs which crossed a category boundary was amplified, while activation in response to pairs which did not to cross the boundary was suppressed. It could be shown that activation in the left supramarginal gyrus most strongly reflected such categorization. This suggests a role for the supramarginal gyrus (which is part of our ROI in the IPL) in constructive perceptual processes and is in line with the present study. It can be speculated that the presentation of the audiovisual stimuli provides information that is used by the supramarginal gyrus in order to then influence regions earlier in the hierarchy in building the percept, i.e. in categorizing the auditory stimuli. Interestingly, another study investigating phonetic invariance observed a region in the left IFS which was insensitive to acoustic changes within a phonetic category while demonstrating sensitivity to changes between phonetic categories (Myers, Blumstein, Walsh, & Eliassen, 2009). We also found ROIs in this region bilaterally for the behaviorally-weighted contrast.

A recent study (Naumer, et al., 2009) may expand this general view. In order to study the effect of learning associations between novel sounds and images on audiovisual object representation, the authors compared the integration of such stimuli before and after training of specific pairings. Intriguingly, the network of regions they found for post-training integration is very similar to the network we found here. More specifically, while pre-training activation was largely confined to the right inferior frontal cortex (coinciding with our ROI in the IFS), this activation became bilateral after training. Additionally, activation was observed bilaterally in the inferior parietal lobe, seemingly in agreement with the location of our IPL ROIs. Even our ROIs in posterior middle temporal gyrus can be found back in these data, namely when comparing post-training activation in response to learned associations versus mismatched audiovisual stimuli. The conspicuous agreement of this network with the network of regions we found with our behaviorally-weighted contrast suggests that there should be a role of these areas that is shared by both paradigms. What the paradigms have in common is learning of an audiovisual association. While learning is

explicitly required in the study by Naumer and colleagues (2009), it is of implicit nature in the present paradigm. This fits the definition put forward by Gilbert et al. (2001, p. 681), who refer to perceptual learning as “improving one's ability, with practice, to discriminate differences in the attributes of simple stimuli”. Improvement has to be viewed in relative and contextual terms here. In the case of recalibration, the disambiguating information from audiovisual exposure biases auditory perception such that it can be regarded as improved in reference to the (momentary) demands of sensory reality. It, thus, qualifies as a case of perceptual learning. Cross-modal learning, in this case, does not only induce plastic changes of the neural substrates of audiovisual processing, but these also have a functional role. This is constructive in nature. Every time an association between the ambiguous auditory stimuli and a certain lip movement is learned, this updates our interpretation of the world, i.e. biases subsequent perception. This is the same role that can be assigned to the supramarginal gyrus on the basis of the study by Raizada and Poldrack (2009). The major difference lies only in the information that is used as a basis for the constructive process. While in our paradigm it is associations with lip movements, in their study it is the contrast with phonemes from the other side of the category boundary that forms this basis. Extreme cases of such multisensory plastic functional changes are apparent in phenomena such as acquired synesthesia (Beauchamp & Ro, 2008; Naumer & van den Bosch, 2009).

Alternative explanations: verbal working memory and cognitive control

There are some alternative interpretations of our data which demand some consideration. Both inferior parietal and inferior frontal cortices have been linked to different verbal working memory processes (Baldo & Dronkers, 2006; Hertrich, et al., 2011). It is conceivable that stronger activation of such a verbal working memory network (possibly due to block-to-block variability of attention) would reflect better retention of the phonemes presented during the audiovisual blocks, which then would exert a bigger influence on the perceptual interpretation of the auditory stimuli presented afterwards. If this was the decisive factor behind audiovisual recalibration, it would be hard to assess whether this effect was really due to actual constructive perceptual changes. Alternatively, a more cognitive response bias could have an impact here. Support for this notion comes from a meta analysis that linked activation in a region within the IFS to cognitive control in a number of switching paradigms (Derrfuss, Brass, Neumann, & von Cramon, 2005). In such studies (task-switching, set-shifting, stimulus-response reversal studies, and color-

word Stroop studies) the participant is required to select between two competing responses. These studies would suggest then that part of the recalibration effect was due to a response bias as opposed to a pure perceptual shift. In view of the automaticity of the immediate effect that the lip movement information exerts on auditory perception during audiovisual exposure and the strength of the recalibration effect, we do, however, consider this possibility unlikely. Evidence in support of this view is threefold. First, when no ambiguous, but an unambiguous auditory component is used, instead of recalibration effects, selective speech adaptation is found (Bertelson, et al., 2003; Vroomen, van Linden, de Gelder, & Bertelson, 2007). In other words, only pairing an ambiguous auditory component with a disambiguating lip movement increases the probability of similar percepts later in time. Stimuli that are unambiguous in both modalities have the opposite effect, i.e. they decrease the probability of similar percepts. This discrepancy is not explainable on the basis of memory-related response biases, which should be the same in both cases. Secondly, engaging participants in a visuospatial or verbal working memory task under different memory load conditions does not affect phonetic recalibration (Baart & Vroomen, 2010), which would be expected in case of a working-memory-based response bias. Thirdly, we have shown that it is possible to predict the perceptual interpretation of the auditory post-tests on a single-trial basis from activation in the auditory cortex (Kilian-Hütten, et al., 2011). Again, this speaks against a cognitive response bias and suggests a perceptual effect, which is stored in hierarchically lower regions.

Conclusion

We showed that ambiguous sounds paired with lip movements, which effectively disambiguated the sounds, strongly influenced the dichotic categorization of these ambiguous sounds presented in isolation later in time. The hemodynamic response during audiovisual exposure within a network of brain areas, namely bilateral and very symmetric regions in the IPL, IFS, and posterior middle temporal gyrus, predicted later speech perception. Drawing upon published empirical work, we interpret the activation in this network to reflect integrative audiovisual learning effects, which in turn affect constructive (auditory) perceptual processes.

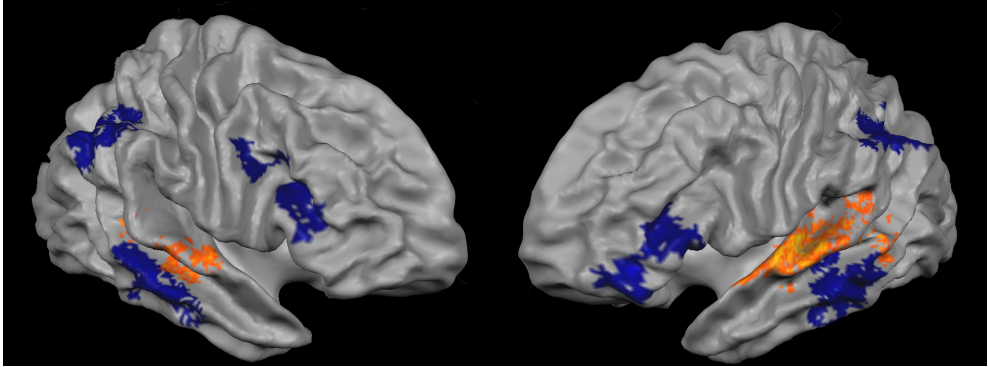


Figure 7. The proposed model. A higher order network (in blue) including IPL, IFS, and MTG is suggested to process integrative learning effects, and consequently install a perceptual bias in auditory regions (in red and orange), most prominently the left Heschl's sulcus and the planum temporale, influencing future constructive auditory perception.

Elsewhere (Kilian-Hütten, et al., 2011), we have provided evidence that suggests that these constructive perceptual processes themselves are located in hierarchically lower, auditory regions (the posterior bank of the left Heschl's gyrus and sulcus and the adjoining left planum temporale). Thus, together with the results from this article, a complete model of cross-modal recalibration emerges, in which a higher-order network of brain areas (including IPL and IFS) is involved in audiovisual integrative learning processes, which lead to the installation of a perceptual bias in auditory regions (figure 7). This bias in turn determines constructive perceptual processes.

References

- Baart, M., & Vroomen, J. (2010). Phonetic recalibration does not depend on working memory. *Exp Brain Res*, 203(3), 575-582.
- Baldo, J. V., & Dronkers, N. F. (2006). The role of inferior parietal and inferior frontal cortex in working memory. *Neuropsychology*, 20(5), 529-538.
- Beauchamp, M. S., Nath, A. R., & Pasalar, S. (2010). fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *J Neurosci*, 30(7), 2414-2417.

- Beauchamp, M. S., & Ro, T. (2008). Neural substrates of sound-touch synesthesia after a thalamic lesion. *J Neurosci*, *28*(50), 13696-13702.
- Benoit, M. M., Raij, T., Lin, F. H., Jaaskelainen, I. P., & Stufflebeam, S. (2010). Primary and multisensory cortical activity is correlated with audiovisual percepts. *Hum Brain Mapp*, *31*(4), 526-538.
- Bertelson, P., Vroomen, J., & De Gelder, B. (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychol Sci*, *14*(6), 592-597.
- Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex*, *11*(12), 1110-1123.
- Calvert, G. A., & Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. *J Cogn Neurosci*, *15*(1), 57-70.
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol*, *10*(11), 649-657.
- Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *J Physiol Paris*, *98*(1-3), 191-205.
- Derrfuss, J., Brass, M., Neumann, J., & von Cramon, D. Y. (2005). Involvement of the inferior frontal junction in cognitive control: meta-analyses of switching and Stroop studies. *Hum Brain Mapp*, *25*(1), 22-34.
- Fingelkurts, A. A., Fingelkurts, A. A., Krause, C. M., Mottonen, R., & Sams, M. (2003). Cortical operational synchrony during audio-visual speech integration. *Brain Lang*, *85*(2), 297-312.
- Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med*, *33*(5), 636-647.
- Forster, B., Cavina Pratesi, C., Aglioti, S. M., & Berlucchi, G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Exp Brain Res*, *143*(4), 480-487.
- Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Exp Brain Res*, *147*(3), 332-343.

- Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, 6(3), 218-229.
- Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: characterizing differential responses. *Neuroimage*, 7(1), 30-40.
- Friston, K. J., Holmes, A. P., Price, C. J., Buchel, C., & Worsley, K. J. (1999). Multisubject fMRI studies and conjunction analyses. *Neuroimage*, 10(4), 385-396.
- Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, 31(5), 681-697.
- Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp*, 27(5), 392-401.
- Hertrich, I., Dietrich, S., & Ackermann, H. (2011). Cross-modal interactions during perception of audiovisual speech and nonspeech signals: an fMRI study. *J Cogn Neurosci*, 23(1), 221-237.
- Hughes, H. C., Reuter Lorenz, P. A., Nozawa, G., & Fendrich, R. (1994). Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. *J Exp Psychol Hum Percept Perform*, 20(1), 131-153.
- Jancke, L., Wustenberg, T., Scheich, H., & Heinze, H. J. (2002). Phonetic perception and the temporal cortex. *Neuroimage*, 15(4), 733-746.
- Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport*, 14(8), 1129-1133.
- Jones, J. A., & Jarick, M. (2006). Multisensory integration of speech signals: the relationship between space and time. *Exp Brain Res*, 174(3), 588-594.
- Kilian-Hütten, N., Valente, G., Vroomen, J., & Formisano, E. (2011). Auditory cortex encodes the perceptual interpretation of ambiguous sound. *J Neurosci*, 31(5), 1715-1720.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746-748.

- Meienbrock, A., Naumer, M. J., Doehrmann, O., Singer, W., & Muckli, L. (2007). Retinotopic effects during spatial audio-visual integration. *Neuropsychologia*, *45*(3), 531-539.
- Mottonen, R., Krause, C. M., Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Brain Res Cogn Brain Res*, *13*(3), 417-425.
- Myers, E. B., Blumstein, S. E., Walsh, E., & Eliassen, J. (2009). Inferior frontal regions underlie the perception of phonetic category invariance. *Psychol Sci*, *20*(7), 895-903.
- Naumer, M. J., Doehrmann, O., Muller, N. G., Muckli, L., Kaiser, J., & Hein, G. (2009). Cortical plasticity of audio-visual object representations. *Cereb Cortex*, *19*(7), 1641-1653.
- Naumer, M. J., & van den Bosch, J. J. (2009). Touching sounds: thalamocortical plasticity and the neural basis of multisensory integration. *J Neurophysiol*, *102*(1), 7-8.
- Nishitani, N., & Hari, R. (2002). Viewing lip forms: cortical dynamics. *Neuron*, *36*(6), 1211-1220.
- Ojanen, V., Mottonen, R., Pekkola, J., Jaaskelainen, I. P., Joensuu, R., Autti, T., et al. (2005). Processing of audiovisual speech in Broca's area. *Neuroimage*, *25*(2), 333-338.
- Olson, I. R., Gatenby, J. C., & Gore, J. C. (2002). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Brain Res Cogn Brain Res*, *14*(1), 129-138.
- Raizada, R. D., & Poldrack, R. A. (2007). Selective amplification of stimulus differences during categorical processing of speech. *Neuron*, *56*(4), 726-740.
- Saito, D. N., Yoshimura, K., Kochiyama, T., Okada, T., Honda, M., & Sadato, N. (2005). Cross-modal binding and activated attentional networks during audio-visual speech integration: a functional MRI study. *Cereb Cortex*, *15*(11), 1750-1760.
- Schroger, E., & Widmann, A. (1998). Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology*, *35*(6), 755-759.
- Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neurosci Res*, *47*(3), 277-287.

- Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, 385(6614), 308.
- Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Curr Opin Neurobiol*, 11(4), 505-509.
- Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *Neuroimage*, 44(3), 1210-1223.
- van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, 43(2), 271-282.
- van Atteveldt, N. M., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex*, 17(4), 962-974.
- Vroomen, J., van Linden, S., de Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Contrasting build-up courses. *Neuropsychologia*, 45(3), 572-577.
- Werner, S., & Noppeney, U. (2010). Superadditive Responses in Superior Temporal Sulcus Predict Audiovisual Benefits in Object Categorization. *Cereb Cortex*, 20 (8): 1829-1842.
- Wiersinga Post, E., Tomaskovic, S., Slabu, L., Renken, R., de Smit, F., & Duifhuis, H. (2010). Decreased BOLD responses in audiovisual processing. *Neuroreport*, 21(18), 1146-1151.
- Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb Cortex*, 13(10), 1034-1043.

**CORRESPONDENCE BETWEEN EMOTION PERCEPTION BIASES IN
CLINICAL DEPRESSION AND MUSICALLY INDUCED SAD MOOD**

CORRESPONDING PUBLICATION

KILIAN-HUTTEN, N., SARKHEIL, P., ESPOSITO, F., AND FORMISANO, E. (2011). CORRESPONDENCE BETWEEN EMOTION PERCEPTION BIASES IN CLINICAL DEPRESSION AND MUSICALLY INDUCED SAD MOOD. *IN PREPARATION.*

Abstract

Besides heavily affecting mood and emotional experience, major depressive disorder (MDD) is characterized by a number of cognitive symptoms. One of the most crucial ones is biased emotion perception, which results in preferential processing of emotionally negative stimuli, for instance in the domains of attention and memory. Depressed patients have also been suggested to perceive facial expressions as sadder than controls. However, these assertions rely mainly on explicit ratings by the patients, which may be problematic due to experimental demand characteristics.

In two behavioral studies we show here that depressed patients do indeed perceive sad emotions to persist longer on changing faces than controls using an implicit measure (study 1) and that musically induced mood can be employed in order to mimic a similar perceptual bias in healthy participants (study 2). To this end, videos were synthesized of facial expression morphing from sad or happy expressions, respectively, to neutral ones. Participants were required to stop the videos when the original expression was not perceived anymore. Depressed patients and healthy participants under sad mood induction both demonstrated delayed response times for sad, but not happy videos.

These results suggest that acute and pathological sad mood can induce similar cognitive symptoms, a notion which is compatible with a dimensional conceptualization of depression. Furthermore, they emphasize the impact that music can have on the way we perceive the world, calling for future research into potential benefits of music in therapeutic settings.

Introduction

Although mainly diagnosed according to affective and motivational criteria, major depressive disorder (MDD) has been suggested to substantially affect cognitive processing as well (Beck, 1967; Beck, 2008; Disner, Beevers, Haigh, & Beck, 2011). More specifically, a growing body of literature links MDD to biases in emotion perception (Bourke, Douglas, & Porter, 2010; J. M. Leppanen, 2006; Mathews & MacLeod, 2005). These biases usually take the form of preferentially processing sad, as compared to happy, information. Individuals afflicted with MDD have, for instance, been shown to demonstrate increased memory performance for sad stimuli (Barry, Naus, & Rehm, 2004; Matt, Vázquez, & Campbell, 1992; Watkins, 2002), to devote more attention to negative emotional stimuli (Desseilles, et al., 2011; Gotlib, Krasnoperova, Yue, & Joormann, 2004; Peckham, McHugh, & Otto, 2010; Schock, Schwenzer, Sturm, & Mathiak, 2011) and to perceive emotional facial expressions, including ambiguous ones, as sadder (Beck, 1967; Bouhuys, Geerts, & Gordijn, 1999; Bouhuys, Geerts, Mersch, & Jenner, 1996; Fu, et al., 2008; Geerts & Bouhuys, 1998; George, et al., 1998; Gur, Erwin, Gur, & Zvil, 1992; Hale, 1998; Hale, Jansen, Bouhuys, & van den Hoofdakker, 1998; Leppanen, Milders, Bell, Terriere, & Hietanen, 2004; Levkovitz, Lamy, Ternochiano, Treves, & Fennig, 2003; Surguladze, et al., 2004).

The recognition of emotional facial expressions is a crucial, ancient feat of human social interaction. The explanation of its evolutionary importance is straightforward in that the ability to correctly read somebody else's emotional state endorses empathy and helping behavior and aids interpretations of social situations as well as the identification of potential threats (Bourke, et al., 2010; Phillips, Drevets, Rauch, & Lane, 2003). Mood-congruent processing biases in the emotion perception of individuals diagnosed with MDD pose a severe threat for the progression and self-preservation of the disorder. This is because perceiving more sadness on the faces of other people arguably has the potential to corroborate a patient's own sad mood in the sense of a self-fulfilling prophecy. The studies cited above typically examined emotion recognition by explicitly asking participants to categorize pictures of facial expressions according to their emotional content or to rate the intensity of the expressed emotion. Potentially problematic for this rationale are the demand characteristics inherent in such a set up. In other words, patients diagnosed with MDD may rate images sadder due to their expectancy of the experimenter's desired results. In the first

of the two studies presented here, we therefore employed a more implicit test of the emotion perception of depressed patients. To this end we synthesized videos of faces morphed from sad and happy expressions, respectively, to neutral ones. The participants' task was to indicate the time point at which they did not recognize the original emotional expression anymore. We expected patients to perceive sad expressions to persist longer on the faces than matched control subjects, and response times to happy expressions to either show no difference between the groups or to be shorter for depressed patients.

The second study presented here employs classical music in order to differentially induce sad and happy mood states, respectively. Music is an extremely powerful stimulus and probably the most consumed art form in terms of time spent on it. From personal experience we know that music can influence our perception of social scenes, which is why it plays such a prominent role in the movie industry (Hoeckner, Wyatt, Decety, & Nusbaum, 2011). Furthermore, it has been demonstrated that the emotional valences transported via faces and music, respectively, interact and modify cortical multisensory activation depending on their congruency (Jeong, et al., 2011). Furthermore, it has been found that emotional music can effectively influence the perception of emotional faces (Bouhuys, Bloem, & Groothuis, 1995; Jolij & Meurs, 2011; Logeswaran & Bhattacharya, 2009) and modulate the strength of, and event-related responses to, an emotional negativity bias in face perception (Chen, Yuan, Huang, Chen, & Li, 2008). In the second study presented here, we differentially induced sad and happy mood states in the same, healthy subjects using classical music pieces that were previously rated for emotional valence by independent subjects. The same implicit emotion perception task as in study 1 was employed. We hypothesized that musically induced mood can create an emotion perception bias in healthy participants that mirrors that found in clinically depressed patients. Besides being relevant for the topic of cross-modal interaction in emotional processing, these investigations are also of interest in the discussion of a categorical versus dimensional conceptualization of depression (Widiger & Samuel, 2005).

Methods

Study 1

Participants

Twenty depressed patients (11 female; age 46.95 ± 13.19 years [mean \pm s.d.]; range 21-72) were recruited from the in-patient population of the University Hospital RWTH Aachen,

Germany. Twenty healthy controls (age 47.60 ± 13.74 years [mean \pm s.d.]; range 21-73) without psychiatric or neurological disorders were recruited and matched for age and sex. All patients met the ICD-10 (International Classification of Diseases) criteria for a diagnosis of major depressive disorder. Participants were excluded in case of any neuropsychiatric diagnosis (for patients: other than depression), head trauma within the last 6 months, motor problems, or a history of substance and/or alcohol abuse (acute or in the past six months). Patients underwent treatment with standard antidepressant medication (selective serotonin re-uptake inhibitors, tricyclic antidepressants, selective noradrenalin re-uptake inhibitors, or 5-HT₂ antagonists). Patients taking benzodiazepines like diazepam were excluded from the study while patients taking lorazepam would only be tested at least 6 hours after medical treatment with lorazepam. Two patients were in treatment with electroconvulsive therapy at the time of assessment. The study was approved by the local Ethics Committee. Participants were instructed verbally, received an information sheet explaining the procedure and gave written informed consent. They were paid for their participation. Two patients were excluded from the analysis due to corrupted data sets.

Stimuli

Based on the NimStim set of facial expressions (Tottenham, 2009), videos were synthesized for the faces of 30 actors of different ethnicities by morphing between images classified as sad or happy, respectively, and those classified as neutral (see figure 1). This resulted in a set of 30 sad-to-neutral and 30 happy-to-neutral videos, which were each 8 seconds long and consisted of 120 frames (15 frames per second). Additionally, the corresponding videos morphed from a neutral expression to sad or happy, respectively, were created. Videos were morphed using MorphMan 4.0 (Stoik, Moscow) software. The videos were 640 x 500 pixels in size.

Procedure

Participants were asked to fill out the German versions of the Beck Depression Inventory (BDI) II (Hautzinger, 1991) and the Positive and Negative Affect Schedule (PANAS; Krohne, Egloff, Kohlmann, & Tausch, 1996). During the actual experiment, participants watched a subset of the videos (sad-to-neutral, happy-to-neutral, neutral-to-sad, neutral-to-happy for 15 actors = 60 videos) presented in random order. Participants were instructed to

indicate via button press when the facial expression changed, either from an emotional to a neutral expression or from a neutral to an emotional one. The inter-stimulus intervals were of varying length (8000 ms to 12000 ms) to prevent automatic response patterns. There was a short break after the first 30 trials. Before the experiment, participants were allowed to familiarize themselves with the task by performing a practice session that consisted of 8 trials with videos from two additional actors not presented in the actual experiment. The whole experimental session lasted approximately 45 minutes.

Analysis

Data were analyzed using SPSS (SPSS Inc., Chicago). The data set was split according to morphing direction (emotional-to-neutral versus neutral-to-emotional) for the analyses. In order to account for the mixed within-/between-subjects nature of the study, repeated-measures general linear models (mixed model analysis of variance, ANOVA) were run for each direction. Included were the within-subjects factor Video Valence (sad versus happy) and the between-subjects factor Depression (depressed versus control). The dependent variable was response time (RT), measured in milliseconds from the beginning of the respective video. Main effects and simple main effects were analyzed using post-hoc independent-samples t-tests.

Study 2

Participants

Thirty-nine adults, who were neither currently, nor in the past, afflicted with a mood disorder, participated in the experiment (age 21.6 ± 2.8 years [mean \pm s.d.]; range 18–32 years), twenty-eight of them female. None of the participants had a history of hearing loss or neurological abnormalities. After complete description of the study to the subjects, all gave written informed consent. The ethics commission of the Faculty of Psychology and Neuroscience at Maastricht University approved the study.

Stimuli

Music. Thirty 120-second excerpts were cut out from classical music pieces, subjectively chosen for sad and happy valence. Group waveform normalization to the average volume level of the source files was applied. The ten happiest and ten saddest of these sequences were then selected based on emotional valence ratings by twenty participants not included in the actual experiment. Ratings were obtained using Self-Assessment Manikins (SAM;

Bradley & Lang, 1994). SAM scores can be analyzed numerically on a scale from 1 to 9. The ten saddest rated sequences had a mean valence rating of 3.29 and the ten happiest ones had a mean valence rating of 6.80. All mean ratings of individual pieces were significantly different from the overall mean valence (5.24). All sad pieces were rated significantly sadder than the overall mean ($-3.07 \geq t \geq -8.26$; all p -values $< .01$). All happy pieces were rated significantly happier than the overall mean ($4.09 < t < 12.07$; all p -values $< .01$).

Videos. The videos used in study 2 were identical to those used in study 1. However, only videos morphing from sad or happy expressions, respectively, to neutral were employed.

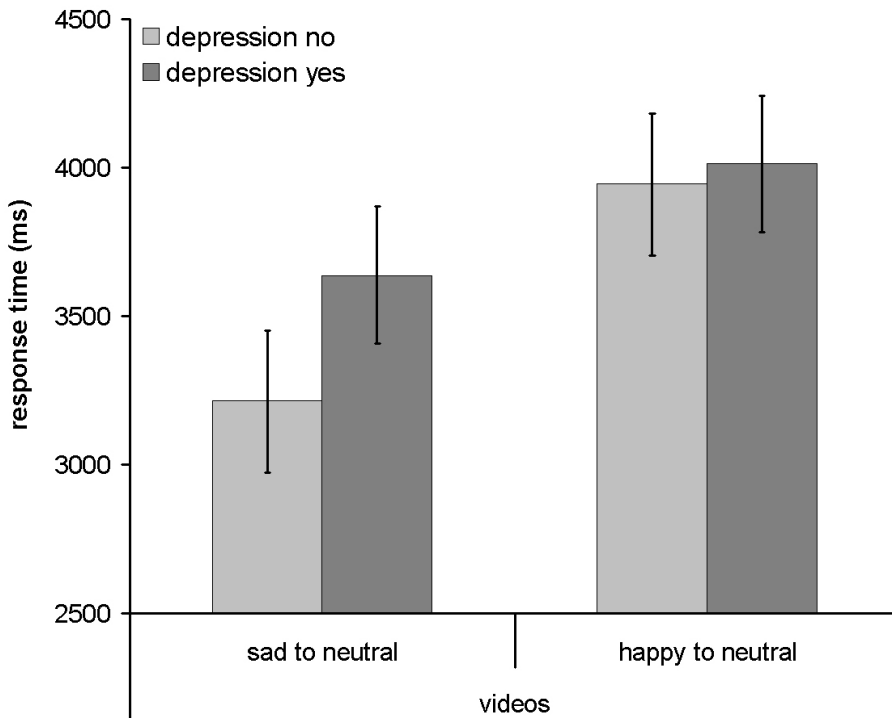


Figure 1. Behavioral results of study 1. For depressed versus healthy participants, the graph depicts average response times for sad-to-neutral and happy-to-neutral videos. Participants were required to respond as soon as they did not recognize the original emotional expression anymore. Error bars indicate \pm one SEM.

Procedure

Participants listened to the happy and sad music pieces in random order. During each piece, six videos were presented showing a face gradually morphed from a happy or a sad expression, respectively, to a neutral one. Three of the six videos were sad-to-neutral and the other three were happy-to-neutral. Participants were asked to press a button with their right index finger as soon as they could not recognize the initial emotional expression on the face anymore. Participants were deceived about the real purpose of the experiment and were told that its aim was the investigation of musical memory. In order to emphasize this deception, six 1.5 second excerpts of music were presented after the videos and participants were asked to identify those excerpts that belonged to the previously heard music sequence by means of a button press. Three of the six excerpts were actually taken from the respective previous pieces and three were taken from others. Stimuli were presented binaurally through loudspeakers and on a standard PC screen with a 60-Hz refresh rate in the behavioral experiment. Stimuli were presented using the software package Presentation (Neurobehavioral Systems). After participation, subjects were asked to indicate their affinity toward classical music on a five point scale ('I hate it' / 'I don't like it' / 'I don't mind it' / 'I like it' / 'I love it').

Analysis

A within-subjects analysis of variance (ANOVA) was carried out including three factors; Music Valence, Video Valence and Classical Music Affinity. The dependent variable was RT as defined in study 1. The variable Classical Music Affinity was recoded in order to obtain a dichotomous measure ("I hate it" / "I don't like it" / "I don't mind it" versus "I like it" / "I love it"). Post-hoc analyses were carried out using lower-order ANOVAs.

Results

Study 1

Questionnaires

As expected, patients differed significantly from controls on the BDI-II (24.5 ± 13.19 versus 3.55 ± 3.58 [mean \pm s.d.]; $t(21.78) = 6.86$, $p < 0.001$). On the PANAS, patients scored significantly lower on the positive items (27.6 ± 9.72 [mean \pm s.d.]) than controls (33.4 ± 6.49 [mean \pm s.d.]; $t(33.14) = -2.22$, $p < 0.05$). On negative items, patients scored

significantly higher than controls (19.75 ± 8.78 versus 10.95 ± 1.43 [mean \pm s.d.]; $t(20.01) = 4.423$, $p < 0.001$).

Neutral-to-emotional direction

For videos morphed from neutral expressions toward emotional ones, the interaction between Video Valence and Depression was not significant ($F(1,36) = 1.315$, $p > .25$), neither was the main effect of Depression ($F(1,36) = .359$, $p > .5$). The RTs to videos morphed toward sad were significantly shorter than the RTs to videos morphed toward happy ($F(1,36) = 93.385$, $p < .001$).

Emotional-to-neutral morphing direction

For videos morphed from emotional expressions toward neutral ones, the interaction between Video Valence and Depression was marginally significant ($F(1,36) = 2.963$, $p < .1$). The inspection of simple main effects showed that for happy-to-neutral videos, there was no significant difference between patients and controls ($t(568) = -.607$, $p > .5$). However, for sad-to-neutral videos, patients were significantly slower in responding than controls (3212 ms versus 3635 ms, $t(567) = -3.667$, $p < .001$). In other words, depressed patients perceived sad expressions to persist longer on the faces than healthy controls (see Figure 1).

Study 2

Concerning their affinity toward classical music, 24 participants answered 'I like it' or 'I love it', 15 participants answered with one of the other options. The three-way interaction Music Valence x Video Valence x Classical Music Affinity was significant ($F(1,4551) = 4.134$, $p < .05$). Consequently, we analyzed the two-way interactions Music Valence x Video Valence on both levels of the third variable (Classical Music Affinity) individually. For participants who indicated low affinity for classical music, this interaction was not significant ($F(1,1596) = .896$, $p > .3$). However, for participants who indicated high affinity for classical music, it did reach significance ($F(1, 2955) = 4.618$, $p < .05$). More specifically, the simple main effects of Musical Valence indicated that, among participants with a high affinity for classical music, sad expressions were perceived to persist longer on the faces during sad musical mood induction than during happy musical mood induction

(6682 ± 1449 ms vs. 6410 ± 1467 ms [mean ± s.d.]; $F(1,1462) = 12.777$, $p < .001$). The response time to happy-to-neutral videos was not affected by sad versus happy musical mood induction (5672 ± 1490 ms vs. 5635 ± 1542 ms [mean ± s.d.]; $F(1,1495) = .229$, $p > .6$). In other words, for people who like classical music, musically induced mood generated a bias in emotion perception that exclusively affected the perception of sad facial expressions, delaying the time point at which participants ceased to see these expressions (see figure 2).

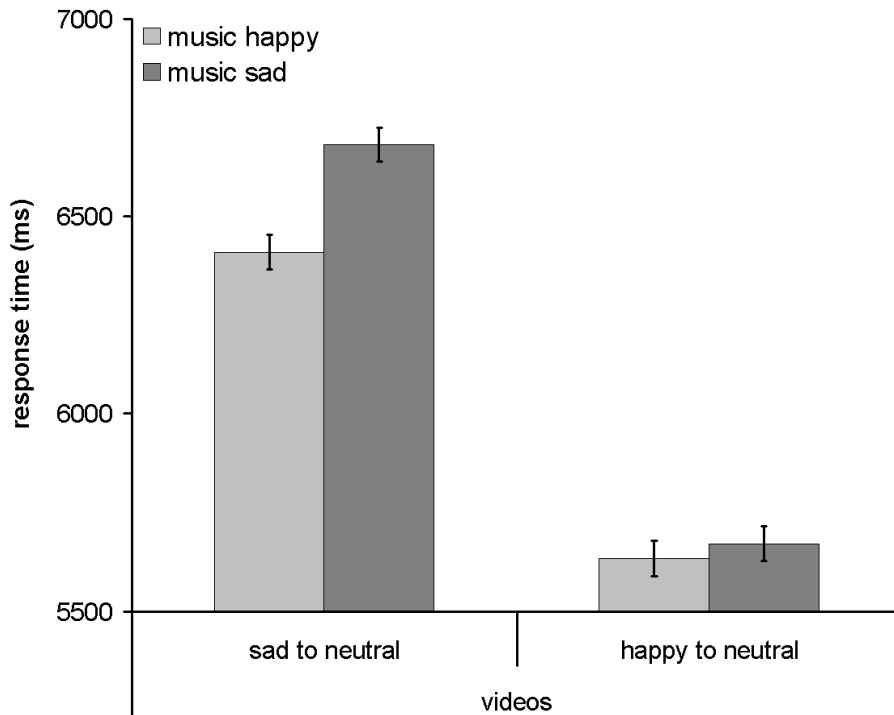


Figure 2. Behavioral results. For participants with an affinity for classical music, the graph depicts average response times for sad-to-neutral and happy-to-neutral videos, during the presentation of sad versus happy music. Participants were required to respond as soon as they did not recognize the original emotional expression anymore. Error bars indicate ± one SEM.

Discussion

In the work presented here, we have shown that depressed patients exhibit a bias in emotion perception, in that they perceive sad expressions to persist longer on faces changing from

sad to neutral expressions than controls. This effect corroborates earlier findings (Beck, 1967; Bouhuys, et al., 1999; Bouhuys, et al., 1996; Fu, et al., 2008; Geerts & Bouhuys, 1998; George, et al., 1998; Gur, et al., 1992; Hale, 1998; Hale, et al., 1998; Leppanen, et al., 2004; Levkovitz, et al., 2003; Surguladze, et al., 2004), and adds the insight that a depressive perceptual bias can even be demonstrated when demand characteristics can be ruled out. Furthermore, we show that a similar bias can be mirrored in healthy participants by means of musically induced mood. More specifically, sad classical music lets people with an affinity to classical music perceive sad expressions to persist longer on the changing faces than happy classical music.

Biased emotion perception is a serious aspect of clinical depression. While most investigations, as well as treatment approaches, of depression concentrate on the affective aspects of the disorder, the importance of its cognitive symptoms is often overlooked. A bias in emotion perception, for instance, may actually play a crucial role in the self-perpetuation and corroboration of depression (Hill, Lewicki, & Neubauer, 1991; Wray, Freund, & Dougher, 2009). It is easily conceivable that seeing more negative affective facial expressions around one would further depress one's mood, which would in turn increase the perceptual bias, effectively creating a vicious circle. Self-fulfilling prophecies easily come into play when the distorted perception of the people around one confirms one's grim expectations.

In the second study, we have shown that such an emotion perception bias is not unique to clinical depression. In fact, classical music could be employed to differentially induce sad and happy mood states within healthy participants. Sad mood states manifested themselves in distorted perception of facial expressions, mirroring the bias found in depressed patients. This finding suggests that acute and pathological sad mood can result in similar cognitive symptoms. This idea is compatible with a dimensional conceptualization of depression, a notion that has become increasingly popular in modern psychiatry (Bjelland, et al., 2009; Krueger & Bezdjian, 2009; Prisciandaro & Roberts, 2009; Shankman & Klein, 2002), and has been discussed animatedly in the development of the fifth edition of Diagnostic and Statistical Manual of Mental (DSM-V; Widiger & Samuel, 2005). The fact that we could demonstrate the perceptual distortion effect only in people who indicated an affinity toward classical music can be explained rather straightforwardly.

It is not surprising that only a stimulus one is fond of has the potential to alter one's mood state to such a profound effect that it affects cognition.

In the first study, not only videos of faces morphing from an emotion toward neutral were presented, but also videos following the opposite trajectory- from neutral toward an emotion. However, neither the interaction between Depression and Video Valence, nor the main effect of Depression was significant. The reasons for this can only remain speculative, but it is conceivable that the task for neutral-to-emotional videos is merely based on detection of change, which would, thus, not necessarily privilege mood-congruent information. In contrast, for videos morphed from an emotional state toward neutral, sad mood can be expected to interact with the sadness of the facial expressions presented, effectively delaying the time point at which it is perceived to dissolve.

In the analyses, main effects of Video Valence could be observed. These effects can probably be attributed to a differential amount of facial muscle activity involved, making changes more or less detectable, respectively. One more observation is worth remarking, which is a large difference between average response times in studies 1 and 2. Relatedly, while in study 1 response times were larger for sad-to-neutral videos, it was the other way around in study 2. Several factors might be at play here. First, the subject populations differed considerably. While study 1 was carried out in a hospital setting, including patients and controls with an average age of approximately 47 years, participants in study 2 were college students with an average age of approximately 22 years. Which subject characteristics would be responsible for these effects remains difficult to pinpoint, though. One alternative explanation is that music might have a general effect of delaying responses. Unfortunately, we cannot test this possible explanation statistically for the time being, since no direct empirical comparison between music and no-music conditions was carried out. Importantly, though, the fact that we find our experimental effects of interest within-subjects in study 2 and between subjects that were closely matched according to age and gender in study 1, makes the observed differences less of a concern. The interactions Depression x Video Valence and Music Valence x Video Valence were unaffected by the overall differences in response time.

Our results emphasize the impact that music has on our emotional experience, as well as on our perception of the world around us. This potential is routinely employed in our society to entertain us or to affect consumer behavior. However, the present results stress music's potential importance in a therapeutic context. As a future direction, it will be

worthwhile to test directly whether happy music can be employed in order to counteract emotion perception biases in the clinically depressed population.

References

- Barry, E. S., Naus, M. J., & Rehm, L. P. (2004). Depression and Implicit Memory: Understanding Mood Congruent Memory Bias. *Cognitive Therapy and Research*, 28(3), 387-414.
- Beck, A. T. (1967). *Depression: Clinical, Experimental, and Theoretical Aspects*. New York: Harper & Row.
- Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates. *Am J Psychiatry*, 165(8), 969-977.
- Bjelland, I., Lie, S. A., Dahl, A. A., Mykletun, A., Stordal, E., & Kraemer, H. C. (2009). A dimensional versus a categorical approach to diagnosis: Anxiety and depression in the HUNT 2 study. *International Journal of Methods in Psychiatric Research*, 18(2).
- Bouhuys, A. L., Bloem, G. M., & Groothuis, T. G. (1995). Induction of depressed and elated mood by music influences the perception of facial emotional expressions in healthy subjects. *J Affect Disord*, 33(4), 215-226.
- Bouhuys, A. L., Geerts, E., & Gordijn, M. C. M. (1999). Depressed patients' perceptions of facial emotions in depressed and remitted states are associated with relapse: A longitudinal study. *Journal of Nervous and Mental Disease*, 187(10), 595-602.
- Bouhuys, A. L., Geerts, E., Mersch, P. P. A., & Jenner, J. A. (1996). Nonverbal interpersonal sensitivity and persistence of depression: Perception of emotions in schematic faces. *Psychiatry Research*, 64(3), 193-203.
- Bourke, C., Douglas, K., & Porter, R. (2010). Processing of facial emotion expression in major depression: a review. *Aust N Z J Psychiatry*, 44(8), 681-696.
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *J Behav Ther Exp Psychiatry*, 25(1), 49-59.

- Chen, J., Yuan, J., Huang, H., Chen, C., & Li, H. (2008). Music-induced mood modulates the strength of emotional negativity bias: An ERP study. *Neuroscience Letters*, *445*(2), 135-139.
- Desseilles, M., Schwartz, S., Dang-Vu, T. T., Sterpenich, V., Anseau, M., Maquet, P., et al. (2011). Depression alters "top-down" visual attention: A dynamic causal modeling comparison between depressed and healthy subjects. *NeuroImage*, *54*(2), 1662-1668.
- Disner, S. G., Beevers, C. G., Haigh, E. A., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nat Rev Neurosci*, *12*(8), 467-477.
- Fu, C. H. Y., Williams, S. C. R., Cleare, A. J., Scott, J., Mitterschiffthaler, M. T., Walsh, N. D., et al. (2008). Neural responses to sad facial expressions in major depression following cognitive behavioral therapy. *Biological Psychiatry*, *64*(6), 505-512.
- Geerts, E., & Bouhuys, N. (1998). Multi-level prediction of short-term outcome of depression: Non-verbal interpersonal processes, cognitions and personality traits. *Psychiatry Research*, *79*(1), 59-72.
- George, M. S., Huggins, T., McDermut, W., Parekh, P. I., Rubinow, D., & Post, R. M. (1998). Abnormal facial emotion recognition in depression: Serial testing in an ultra-rapid-cycling patient. *Behavior Modification*, *22*(2), 192-204.
- Gotlib, I. H., Krasnoperova, E., Yue, D. N., & Joormann, J. (2004). Attentional Biases for Negative Interpersonal Stimuli in Clinical Depression. *Journal of Abnormal Psychology*, *113*(1), 127-135.
- Gur, R. C., Erwin, R. J., Gur, R. E., & Zwil, A. S. (1992). Facial emotion discrimination: II. Behavioral findings in depression. *Psychiatry Research*, *42*(3), 241-251.
- Hale, W. W., III. (1998). Judgment of facial expressions and depression persistence. *Psychiatry Research*, *80*(3), 265-274.
- Hale, W. W., III, Jansen, J. H. C., Bouhuys, A. L., & van den Hoofdakker, R. H. (1998). The judgment of facial expressions by depressed patients, their partners and controls. *Journal of Affective Disorders*, *47*(1-3), 63-70.
- Hautzinger, M. (1991). Das Beck-Depressionsinventar (BDI) in der Klinik. *Nervenarzt*, *62*(11), 689-696.
- Hill, T., Lewicki, P., & Neubauer, R. M. (1991). The development of depressive encoding dispositions - A case of self-perpetuation of biases. *Journal of Experimental Social Psychology*, *27*(4), 392-409.

- Hoeckner, B., Wyatt, E. W., Decety, J., & Nusbaum, H. (2011). Film music influences how viewers relate to movie characters. *Psychology of Aesthetics, Creativity, and the Arts*, 5(2), 146-153.
- Jeong, J. W., Diwadkar, V. A., Chugani, C. D., Sinsoongsud, P., Muzik, O., Behen, M. E., et al. (2011). Congruence of happy and sad emotion in music and faces modifies cortical audiovisual activation. *Neuroimage*, 54(4), 2973-2982.
- Jolij, J., & Meurs, M. (2011). Music alters visual perception. *PLoS One*, 6(4), e18861.
- Krohne, H. W., Egloff, B., Kohlmann, C. W., & Tausch, A. (1996). Investigations with a German version of the positive and negative affect schedule (PANAS). *Diagnostica*, 42(2), 139-156.
- Krueger, R. F., & Bezdjian, S. (2009). Enhancing research and treatment of mental disorders with dimensional concepts: toward DSM-V and ICD-11. *World Psychiatry*, 8(1), 3-6.
- Leppanen, J. M. (2006). Emotional information processing in mood disorders: a review of behavioral and neuroimaging findings. *Curr Opin Psychiatry*, 19(1), 34-39.
- Leppanen, J. M., Milders, M., Bell, J. S., Terriere, E., & Hietanen, J. K. (2004). Depression biases the recognition of emotionally neutral faces. *Psychiatry Research*, 128(2), 123-133.
- Levkovitz, Y., Lamy, D., Ternochiano, P., Treves, I., & Fennig, S. (2003). Perception of dyadic relationship and emotional states in patients with affective disorder. *Journal of Affective Disorders*, 75(1), 19-28.
- Logeswaran, N., & Bhattacharya, J. (2009). Crossmodal transfer of emotion by music. *Neurosci Lett*, 455(2), 129-133.
- Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annu Rev Clin Psychol*, 1, 167-195.
- Matt, G. E., Vázquez, C., & Campbell, W. K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review*, 12(2), 227-255.
- Peckham, A. D., McHugh, R. K., & Otto, M. W. (2010). A meta-analysis of the magnitude of biased attention in depression. *Depression and Anxiety*, 27(12), 1135-1142.

- Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003). Neurobiology of emotion perception I: The neural basis of normal emotion perception. *Biol Psychiatry*, 54(5), 504-514.
- Prisciandaro, J. J., & Roberts, J. E. (2009). A comparison of the predictive abilities of dimensional and categorical models of unipolar depression in the National Comorbidity Survey. *Psychological Medicine*, 39(7).
- Schock, L., Schwenzer, M., Sturm, W., & Mathiak, K. (2011). Alertness and visuospatial attention in clinical depression. *BMC Psychiatry*, 11.
- Shankman, S. A., & Klein, D. N. (2002). Dimensional diagnosis of depression: adding the dimension of course to severity, and comparison to the DSM. *Compr Psychiatry*, 43(6), 420-426.
- Surguladze, S. A., Young, A. W., Senior, C., Brébion, G., Travis, M. J., & Phillips, M. L. (2004). Recognition Accuracy and Response Bias to Happy and Sad Facial Expressions in Patients With Major Depression. *Neuropsychology*, 18(2), 212-218.
- Tottenham, N., Tanaka, J. W., Leon, A. C., Mccarry, T., Nurse, M., Hare, T. A., Marcus, D. J., Westerlund, A., Casey, B. J., & Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research*, 168(3):242-249.
- Watkins, P. C. (2002). Implicit memory bias in depression. *Cognition and Emotion*, 16(3), 381-402.
- Widiger, T. A., & Samuel, D. B. (2005). Diagnostic categories or dimensions? A question for the Diagnostic And Statistical Manual Of Mental Disorders--fifth edition. *J Abnorm Psychol*, 114(4), 494-504.
- Wray, A. M., Freund, R. A., & Dougher, M. J. (2009). A Behavior-Analytic Account of Cognitive Bias in Clinical Populations. *Behavior Analyst*, 32(1), 29-49.

**NEURAL BINDING OF SEPARATE COGNITIVE ASPECTS OF
DEPRESSION THROUGH MUSICALLY INDUCED SAD MOOD**

CORRESPONDING PUBLICATION

KILIAN-HÜTTEN, N., ESPOSITO, F., AND FORMISANO, E. (2011). NEURAL BINDING OF SEPARATE COGNITIVE ASPECTS OF DEPRESSION THROUGH MUSICALLY INDUCED SAD MOOD. *SUBMITTED*.

Abstract

Contrary to popular belief, mood disorders, such as depression, are not only characterized by emotional symptoms, but also heavily affect cognition. Two most crucial cognitive symptoms of depression are biased emotion perception and rumination. This study aims to link these aspects of depression and their neural substrates. The neural interaction between rumination and biased emotion perception in non-pathological sad mood is investigated.

Twelve healthy subjects participated in this functional magnetic resonance imaging (fMRI) experiment. Sequences of classical music, pre-rated for emotional valence, were employed to induce sad and happy mood states, respectively. We have shown before that these music sequences can effectively induce an emotion perception bias that mirrors that in clinically depressed patients (chapter 3). FMRI data were analyzed with a novel combination of functional-connectivity analysis techniques.

We demonstrate that the neural substrates underlying rumination and biased emotion perception are linked selectively during sad mood via the posterior superior temporal sulcus. Our results indicate that separate cognitive symptoms of depression interact neurally, which may result in a vicious cognitive circle, and that similar neural substrates may underlie acute and pathological sad mood, suggesting a dimensional concept of depression. Furthermore, they emphasize the impact that music has on our cognitive and behavioral experience, encouraging further investigations of its therapeutic potential.

Introduction

Few external stimuli equal music in terms of the immediacy and strength with which it takes hold of our emotional experience. Music can influence consumer behavior, pain experience, sexual arousal, emotional and cognitive processing, and even aid therapeutical efforts in such diverse disorders as Parkinson's (Pacchetti, et al., 2000) or visual neglect (Soto, et al., 2009). Notably, music even changes our semantic interpretation of social scenes, for instance when watching movies. This is in line with the emotion congruence effect, which describes the general tendency to perceive the same emotion on the faces of other people that is currently being felt (Bouhuys, Bloem, & Groothuis, 1995; Schiffenbauer, 1974). This effect has far-reaching implications for social interaction, most fatally manifested in mood-congruent processing biases in clinically depressed individuals (Mathews & MacLeod, 2005). Depressed patients demonstrate deficits in emotion perception (Kohler, Hoffman, Eastman, Healey, & Moberg, 2011) and increased attention as well as memory performance for negative affective information, likely corroborating their negative mood (Mathews & MacLeod, 2005). Besides these biases, rumination (i.e. repetitive negative self-referential thoughts) has been implicated as a second crucial cognitive symptom (Beck, 1967; Beck, 2008; Disner, Beevers, Haigh, & Beck, 2011; Nolen-Hoeksema, 2000). It is, to date, unclear in how far these distinct cognitive aspects of depression are linked and whether their co-occurrence is specific to pathology (Gotlib & Joormann, 2010). Alternatively, the fact that they are commonly affected in concert might reflect some form of natural coupling between them in the healthy brain. This point is closely related to an ongoing debate in the field of mental health concerning the conceptualization of psychiatric disorders as discrete clinical conditions versus (arbitrary) conventional divisions along dimensions of mental functioning (Widiger & Samuel, 2005) and raises the question to what extent non-pathological sad mood and clinical depression share the same neural basis.

Although a multitude of neuroimaging studies have investigated the neural mechanisms underlying depression (Rigucci, Serafini, Pompili, Kotzalidis, & Tatarelli, 2010), most have either focused on the affective aspects of the disorder or have failed to link the different aspects of the cognitive model of depression (Beck, 1967; Beck, 2008; Disner, et al., 2011) to specific and integrated neural substrates. Furthermore, to our knowledge the link between these cognitive aspects in non-pathological sad mood has not

been investigated. We have shown before that, in non-depressed healthy human participants, music can be used to induce sad mood states which manifest themselves in altered emotion perception (see chapter 4). Here, we employed functional magnetic resonance imaging (fMRI) during music listening in order to examine neural networks which underlie the different cognitive aspects that mimic those of depression symptoms (rumination and biased emotion perception). We used a novel combination of different functional connectivity measures in order to investigate the dynamic and mood-dependent interactions between these different neural networks. As a starting point, we focused on the default-mode network (DMN). The DMN is active during resting state and is deactivated in the context of a cognitive task (Raichle, et al., 2001). It has been implicated in self-referential processing and rumination and has been shown to demonstrate alterations in depressed patients in terms of deactivation and functional connectivity (Hamilton, et al., 2011; Sheline, et al., 2009). The dynamic neural interactions between the DMN (as a neural substrate of rumination) and other networks during sad, as compared to happy, musical mood induction is investigated.

Method

Participants

Fourteen healthy, right-handed participants took part in the fMRI experiment. Two of them had to be excluded due to excessive head movement, leaving twelve participants for the analysis, seven of them female (age 23.7 ± 2.9 years (mean \pm s.d.); range 19–28 years). None of the participants had a history of hearing loss or neurological abnormalities. After complete description of the study to the subjects, all gave informed consent. The ethics commission of the Faculty of Psychology and Neuroscience at Maastricht University approved the study.

Stimuli

Music. Thirty 120-second excerpts were cut out from classical music pieces, subjectively chosen for sad and happy valence. Group waveform normalization to the average level of the source files was applied. The ten happiest and ten saddest of these sequences were then selected based on emotional valence ratings by twenty participants, which were obtained using Self-Assessment Manikins (SAM; Bradley & Lang, 1994). SAM scores can be analyzed numerically on a scale from 1 to 9. The ten saddest rated sequences had a mean

valence rating of 3.29 and the ten happiest ones had a mean valence rating of 6.80. All mean ratings of individual pieces were significantly different from the overall mean valence (5.24). All sad pieces were rated significantly sadder than the overall mean ($-3.07 \geq t \geq -8.26$; all p -values $< .01$). All happy pieces were rated significantly happier than the overall mean ($4.09 < t < 12.07$; all p -values $< .01$).

Videos. Based on the NimStim set of facial expressions (Tottenham, et al., 2009), videos were synthesized for the faces of 30 actors by morphing between images classified as sad or happy, respectively, and those classified as neutral (see figure 1). This resulted in a set of 30 sad-to-neutral and 30 happy-to-neutral videos, which were each 8 seconds long and consisted of 120 frames (15 frames per second). Videos were morphed using MorphMan 4.0 (Stoik) software.



Figure 1. Still images from the video stimuli. Shown are 9 of the original 120 frames (15 frames/second). Stimuli were synthesized by morphing between images classified as sad (a) or happy (b), respectively, and those classified as neutral. This was done for the images of thirty actors from the NimStim set of facial expressions (15).

Experimental procedure

Participants listened to the happy and sad music pieces in random order. After each piece, six videos were presented showing a face gradually morphed from a happy or a sad expression, respectively, to a neutral one. Three of the six videos were sad-to-neutral and the other three were happy-to-neutral. Participants were asked to press a button with their

right index finger as soon as they could not recognize the initial emotional expression on the face anymore. Although unpublished pretests have shown that presenting the videos after, and not during, the music (as was done in the behavioral study presented in chapter 4) eliminates behavioral effects of mimicking a depressive emotion perception bias, this was done in order to avoid the introduction of task effects and the interference with resting-state brain network activity. Participants were deceived about the real purpose of the experiment and were told that its aim was the investigation of musical memory. In order to emphasize this deception, six 1.5 second excerpts of music were presented after the videos and participants were asked to identify those excerpts that belonged to the previously heard music sequence by means of a button press. Three of the six excerpts were actually taken from the respective previous pieces and three were taken from others. Auditory stimuli were presented through an MR-compatible Intercom Commander XG MRI Audio System (Resonance Technologies Inc.) using a 2-way stereo headset and visual stimuli were projected on a screen placed outside the scanner, which participants could see via a mirror mounted to the head coil. Stimuli were presented and synchronized with the MR pulses using the software package Presentation (Neurobehavioral Systems). After participation, subjects were asked to indicate their affinity toward classical music on a five point scale ('I hate it' / 'I don't like it' / 'I don't mind it' / 'I like it' / 'I love it'). Nine participants answered 'I like it' or 'I love it', 2 answered 'I don't mind it', one participant answered 'I don't like it, and no one answered 'I hate it'.

Scanning parameters

Functional MRI data were collected on a 3-Tesla fMRI scanner (head set-up, Siemens) at the Maastricht Brain Imaging Center in Maastricht, The Netherlands. For each participant, three functional runs of each 740 volumes were acquired. For functional images, a blood oxygenation level-dependent (BOLD) sensitive echo-planar imaging (EPI) sequence was used (matrix 64x64, 32 slices; slice thickness = 3 mm; field of view = 192 x 192 mm³, resulting voxel size = 3 x 3 x 3 mm³; TR = 2000 ms, TE/acquisition time slice = 30/62.5, flip angle = 90°). For later overlay, a high resolution structural scan (voxel size, 1 x 1 x 1 mm³) was collected using a T1-weighted three-dimensional (3D) ADNI sequence [TR = 2050 ms; echo time (TE) = 2.6 ms; 192 sagittal slices]. The functional runs and the structural scan were acquired in a single session for each participant.

Analysis

Data preprocessing. Functional image time series were corrected for differences in slice acquisition times with a “sinc” interpolation technique. Furthermore, 3D-motion correction was applied to correct for common small head movements by spatially aligning all volumes to the first volume by rigid body transformations. Further preprocessing steps included linear trend removal and temporal high-pass filtering to remove low-frequency nonlinear drifts of five or less cycles per time course. For group analyses, spatial smoothing with an isotropic Gaussian kernel of 5 mm full width at half-maximum was performed. Functional images were then aligned with T1 volumes and warped into Talairach space. All image data preparations (preprocessing and postprocessing) were performed in BrainVoyager QX (Brain Innovation, Maastricht, The Netherlands), version 2.2.

Single-subject and group-level ICA. One would expect the neural substrates of musically-induced mood to build up slowly over time during the presentation of music. Furthermore, within a given piece, musical valence may change dynamically. The potential contribution of both low-frequency and unpredictable phasic changes in the fMRI signal renders the use of conventional fMRI analysis techniques, such as the General Linear Model (GLM), inadequate. Hence, the data were analyzed using an approach based on independent component analysis (ICA), enabling us to identify underlying and slowly changing networks in a data-driven manner, as is commonly done in resting-state fMRI studies (van de Ven, Formisano, Prvulovic, Roeder, & Linden, 2004). Here, individual Independent Component Analysis (ICA; Hyvärinen, Karhunen, & Oja, 2001) and self-organizing group Independent Component Analysis (sogICA; Esposito, et al., 2005) were applied to the preprocessed functional (experimental) time courses using two “plug-in” extensions of BrainVoyager QX (Brain Innovation, Maastricht, The Netherlands). The single-subject ICA plug-in corresponds to an implementation of the fastICA algorithm (Hyvarinen, 1999). ICA is commonly used for the data driven extraction of functionally connected networks in the resting state (Damoiseaux, et al., 2006; Greicius, Krasnow, Reiss, & Menon, 2003; van de Ven, et al., 2004). Before the ICA decomposition, a dimensionality reduction to 50 was applied based on Principal Component Analysis (PCA). Then ICA was used to extract an equal number of spatially independent components. Independent components were estimated sequentially (deflation) using a hyperbolic tangent as nonlinearity in the fastICA contrast function. SogICA (Esposito, et al., 2005) served to cluster together similar

individual components in the three runs of an individual participant based on linear spatial correlations. A second level sogICA (hierarchical sogICA; van de Ven, Esposito, & Christoffels, 2009) then grouped together spatially similar components from these subject-wise sogICAs.

ICA-based psychophysiological interaction. From the grand sogICA, we selected the default mode network (DMN) component (Esposito & Goebel, 2011; Esposito, et al., 2005; Raichle, et al., 2001) and identified the components from the individual runs within each participant that contributed to this cluster (Esposito, et al., 2008). In order to investigate which voxels demonstrated increased functional connectivity with the default-mode network during sad mood, as compared to happy mood, induction, we applied an analysis combining ICA and psychophysiological interaction (PPI; Friston, et al., 1997). In the orthodox application of PPI, the time course of one brain region is extracted and multiplied with a task predictor. Here, instead of employing the time course of a region of interest, we extracted the component time courses from the individual independent components that contributed to the grand sogICA DMN component. These were then multiplied with task predictors that were set to +1 during the presentation of sad music and to -1 during the presentation of happy music and convolved with a double-gamma hemodynamic response function. This analysis was implemented using custom made Matlab code (MathWorks, Natick, Massachusetts, U.S.A.). Neuroelf (<http://neuroelf.net>) was employed as a BrainVoyager - Matlab interface.

The resulting PPI predictor was then fed into a General Linear Model (GLM), in which we also implemented a predictor for the component time courses. We also integrated all other task predictors, including one for the difference between sad and happy music blocks, in order to account for common effects and only identify those effects that were unique to the PPI. Group maps were created by running this analysis as a random-effects GLM and applying the contrast [PPI > baseline]. For the resulting group maps, a cluster-size threshold procedure was applied. To this end, cluster-level false-positive rates were estimated iteratively (500 iterations; Monte Carlo simulation; Forman, et al., 1995). The minimum cluster-size threshold (based on its observed relative frequency), which yielded a cluster-level false-positive rate (alpha) of 5% was then applied to the statistical maps (75 voxels).

It should be noted here that this combination of analyses does not suffer from circularity. This should become clear when considering that the ICA does not use any explicit information about the experimental protocol, i.e. the conditions. Instead, it identifies components on the basis of the whole time course in a data-driven manner. Furthermore, since the conditions were balanced in each run (sad/happy music; sad-to-neutral/happy-to-neutral videos) an implicit bias in the data can also be ruled out.

Results

Independent components

ICA is an exploratory, data-driven analysis technique, yielding an abundance of interesting data beyond the scope of the direct hypothesis testing aimed at here. More specifically, a considerable number of independent components could be identified from the sogICA that demonstrated high between-subjects consistency. For the sake of clarity, we clustered them into six groups; components likely involved in visual processing, components likely involved in auditory and motor processing, fronto-parietal components probably mostly involved in attention-related processes, components processing own and observed emotions, components involving midline structures, and miscellaneous. In the appendix to this chapter, a qualitative description of the individual components' spatial and temporal characteristics and their suggested roles in the context of our paradigm is presented.

ICA-based PPI

For the ICA-based PPI analysis, we focused on the DMN component. As mentioned before, the DMN has been suggested to be involved in self-referential processing and rumination and has been shown to demonstrate alterations in depressed patients in terms of deactivation and functional connectivity (Hamilton, et al., 2011; Sheline, et al., 2009). The DMN was identified (figure 2) from the multi-subject self-organizing group ICA (sogICA; Esposito, et al., 2008). We then tracked the individual DMN components for each run that contributed to this group component. The DMN could be robustly identified in individual runs and showed high overlap between runs and participants. The component time courses of the individual DMN components were extracted and used to calculate the psychophysiological interaction (PPI; Friston, et al., 1997). A random-effects general linear model (RFX GLM) including this PPI predictor served to identify brain regions which

showed increased functional connectivity with the DMN during sad, as compared to happy, mood induction.

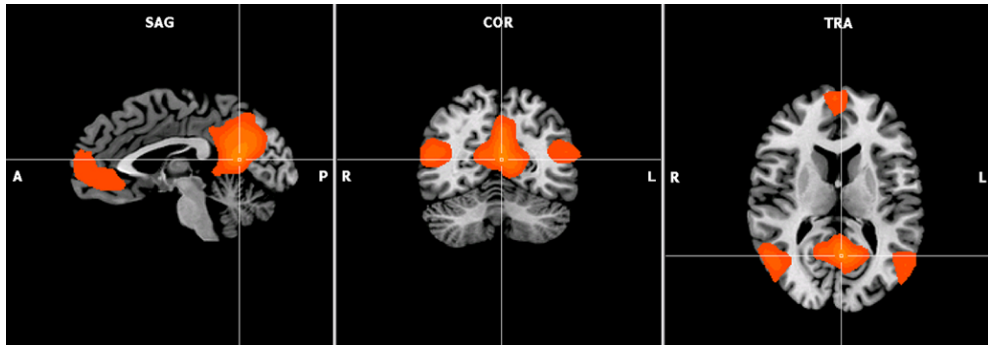


Figure 2. The default-mode network. The DMN ICA component as identified in the grand self-organizing group ICA based on linear spatial correlations between the individual ICAs. The DMN includes medial prefrontal cortex, precuneus, and inferior parietal cortices and could be robustly identified in individual runs in all subjects. Displayed is an overlay on the MNI Colin27 template.

The region of interest (ROI) that exhibited such a response pattern was bilateral posterior superior temporal sulcus (pSTS; see figure 3). Posterior STS is normally implicated in the perception of socially relevant stimuli, such as goal-directed actions (Carter, Hodgins, & Rakison, 2011) or their intention (Pelphrey, Morris, & McCarthy, 2004), as well as socially relevant eye, hand, and body movements (Ethofer, Gschwind, & Vuilleumier, 2011; Kret, Pichon, Grezes, & de Gelder, 2011) and facial expressions (Said, Moore, Engell, Todorov, & Haxby, 2010).

In order to further investigate what functional network the pSTS is involved in within the general context of our paradigm (independently of mood induction), we inspected the remaining sogICA components for overlap with this ROI. The component that showed maximal overlap was a limbic-prefrontal component, including anterior cingulate cortex (aCC), posterior cingulate cortex, bilateral dorsolateral prefrontal cortex, amygdala, and insula (see figure 4). This network has been robustly linked to emotional experience and emotion perception (Fusar-Poli, et al., 2009; Phillips, Drevets, Rauch, & Lane, 2003a). Furthermore, its structural and functional abnormalities in depression (Bennett, 2011; Carballo, et al., 2011; Mayberg, et al., 1999; Phillips, Drevets, Rauch, & Lane, 2003b) and its involvement in sad mood (Mayberg, et al., 1999) have been demonstrated. More

specifically, the limbic-prefrontal network demonstrates functional abnormality in biased depressive emotion perception (Bennett, 2011; Carballedo, et al., 2011; Mayberg, et al., 1999; Phillips, et al., 2003b), the very cognitive effect that we simulated behaviorally in our musical mood-induction paradigm (see chapter 4).

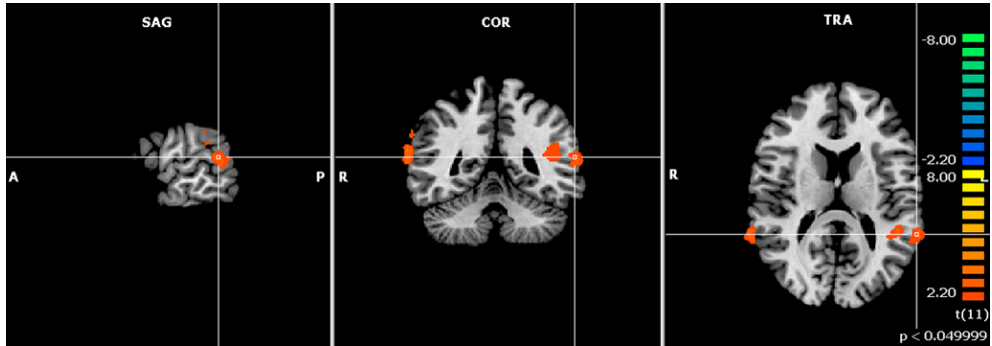


Figure 3. Results from the combined ICA-PPI analysis. Regions that showed a significant response to this contrast in an RFX group GLM demonstrated increased functional connectivity to the default-mode network during sad, as compared to happy, musical mood induction. This was true for bilateral posterior superior temporal sulcus. Maps were thresholded at $p < .05$ and then cluster-size corrected to include only clusters of 75 voxels or more, corresponding to a cluster-level false-positive rate (alpha) of 5%. Displayed is an overlay on the MNI Colin27 template.

Discussion

We have shown before that music can be employed in order to induce sad mood states which manifest themselves in biased emotion perception in healthy participants (see chapter 4). This bias in emotion perception is analogue to cognitive symptoms found in the clinically depressed population (Mathews & MacLeod, 2005). The second crucial cognitive symptom in depressed patients is rumination (Beck, 1967; Beck, 2008; Nolen-Hoeksema, 2000). It has been shown that the neural basis for these repetitive, negative self-referential thinking patterns lies in the abnormal activation and connectivity of the DMN (Disner, et al., 2011; Hamilton, et al., 2011; Sheline, et al., 2009). We demonstrate here that pSTS acts as a hub linking the neural substrates underlying rumination (the DMN) and biased emotional processing (the limbic-prefrontal network), which gets activated selectively during musical sad mood induction.

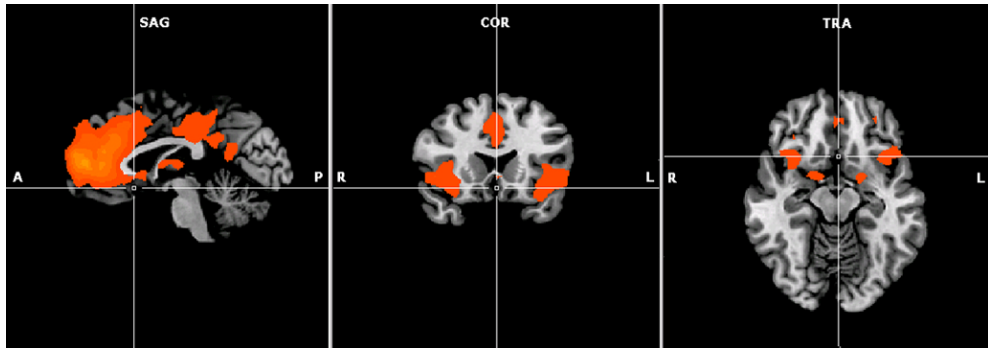


Figure 4. The limbic-prefrontal component. This network includes anterior cingulate cortex, posterior cingulate cortex, bilateral dorsolateral prefrontal cortex, amygdala, ventral striatum / nucleus accumbens and insula, and has been implicated in emotion perception in general and depressive emotion processing bias in particular. Crucially, this component also includes regions that overlap with our pSTS regions of interest. Displayed is an overlay on the MNI Colin27 template.

The importance of our findings is threefold. First, we show evidence that a neural link between these distinct cognitive aspects exists. Such a connection has been suggested on behavioral grounds (Gotlib & Joormann, 2010; Nolen-Hoeksema, 2000), but has not been demonstrated as a neural phenomenon. This coupling between cognitive symptoms has potential consequences for the conceptualization of depression and for therapeutic interventions. It has been suggested that an interplay of negatively biased processing of external stimuli and negative self-referential rumination may result in an ineffective use of emotion-regulation strategies, thereby exacerbating depressive mood (Gotlib & Joormann, 2010). We support this notion and add direct evidence that such a reciprocal aggravation derives from an actual neural link via pSTS. This may also imply that, ideally, psychotherapy should aim at tackling both cognitive aspects simultaneously. Concentrating only on either rumination or biased processing would hamper the goal of breaking through a vicious cognitive circle. The second aspect that is crucial about our findings is that we find the link between cognitive symptoms of depression in healthy participants, employing music to induce sad mood. This indicates that similar neural substrates may underlie acute and pathological sad mood. In other words, these findings suggest a dimensional concept of depression, supporting a notion that has become increasingly popular in modern psychiatry, and has been discussed animatedly in the development of the fifth edition of Diagnostic and Statistical Manual of Mental Disorders (DSM-V; Widiger & Samuel, 2005). Related to this

point, it will be a worthwhile endeavor to investigate individual differences in the susceptibility to mood induction and in the neural correlates of induced mood, rumination, and biased emotion processing. Linking these factors to increased risk levels for the onset of depression would ideally form a basis for a potential screening and prevention procedure, which may help identify non-clinical individuals at risk of developing the disorder. Lastly, our findings stress the impact that music has on our lives, not only in terms of direct affect, but actually influencing the way we perceive and process the world around us. We directly show the neural link between musically induced mood and changes in cognition and perception.

In this study, we concentrated on negative mood induction and its effect. In future research it would be desirable to further investigate ways to systematically and effectively induce positive mood. In this context, replicating the present study with clinically depressed patients will be crucial in order to investigate a potential therapeutic role of music in the treatment of this disease. Also more fine-grained distinctions of musical valence, beyond a sad-happy dichotomy and including such effects as finding joy in sad music, would be desirable.

Appendix:

Other independent components of interest

ICA is an exploratory, data-driven analysis technique. Beyond providing the basis for hypothesis testing as in the main part of this chapter, it yields a wealth of data, identifying a multitude of neural networks that are implicated in task-related and task-unrelated processes. However, in order to make specific claims about their functional roles, independent, confirmatory experimental results are needed. Here, we can only rely on the consideration of the components' spatial and temporal characteristics in order to tentatively ascribe specific functional roles to these networks.

For an illustration of the time courses of the respective components, we show the raw-data event-related averages (ERAs) of the networks, treated as one respective region of interest. Note that long ERAs were created (135 TRs = 270 seconds) in order to capture the components' behavior throughout the course of one complete mini run (one musical piece, six facial videos, six musical excerpts). See figure 5 for an illustration of the events represented in the ERAs.

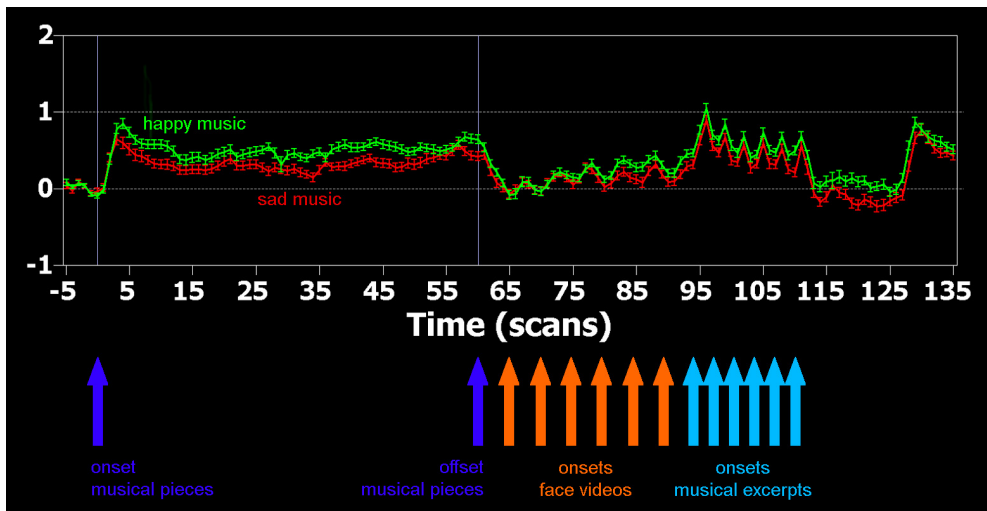


Figure 5. Illustration of the ERAs. Represented in these long ERAs (135 TRs = 270 seconds) are the on- and off-sets of the musical pieces, the onsets of the face videos, and the onsets of the musical excerpts. Happy music conditions are coded in green, sad music conditions in red. This example shows the time course of an auditory component.

Independent components involved in visual processing. Four occipital independent components showed consistent overlap between subjects. An early visual component (fig.6a), including V1, is activated exclusively during the presentation of the videos and reacts with a summed, plateau response during the whole block of videos. Here, the earliest coding of visual signals takes place, which are then further analyzed in secondary and higher-order visual cortices (fig.6b-d). Increasing modulations of the signal by the on- and off-sets of the individual videos can be observed, which suggests more specified and less generalized roles of these networks as compared to primary visual cortex. All visual components remain rather silent during the presentation of the musical pieces and excerpts.

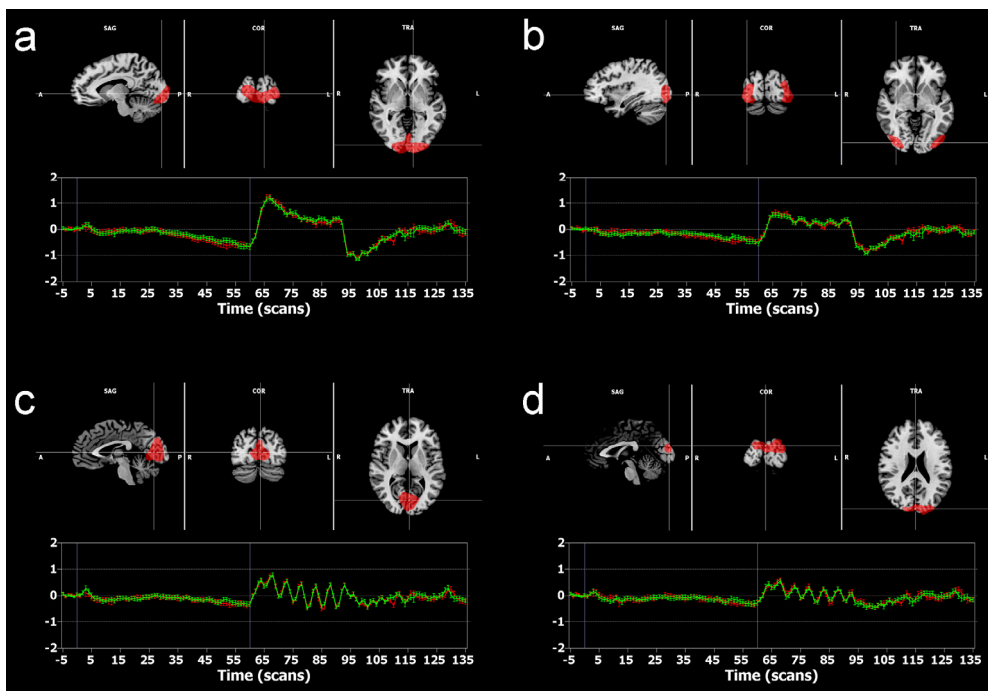


Figure 6. Independent components involved in visual processing. Displayed is an overlay on the MNI Colin27 template.

Independent components involved in auditory and motor processing. The most consistent auditory component is shown in figure 7a. It covers Heschl's gyrus (including primary auditory cortex) and higher-order auditory cortex within the superior temporal sulci and gyri. The ERA clearly shows its responsiveness to the musical pieces and the musical excerpts. During the musical pieces, it demonstrates a continuous, elevated response, which

is, however, modulated by the characteristics of the individual pieces. Related to this is the fact that happy music resulted in a slight increase in activation as compared to sad music. This is probably due to overall differences in energy. Happy music is often somewhat faster and includes a larger number of transitions and onsets. This auditory component is the only one that showed such a difference, which suggests that it did not result in overall processing differences (e.g. in emotion or resting-state networks). During the videos the component also shows a slight modulation. Auditory cortex has been suggested before to be responsive to complex visual stimuli, such as moving faces (Hoffman, Ghazanfar, Gauthier, & Logothetis, 2007), a fact that might point to more or less hardwired face-voice associations.

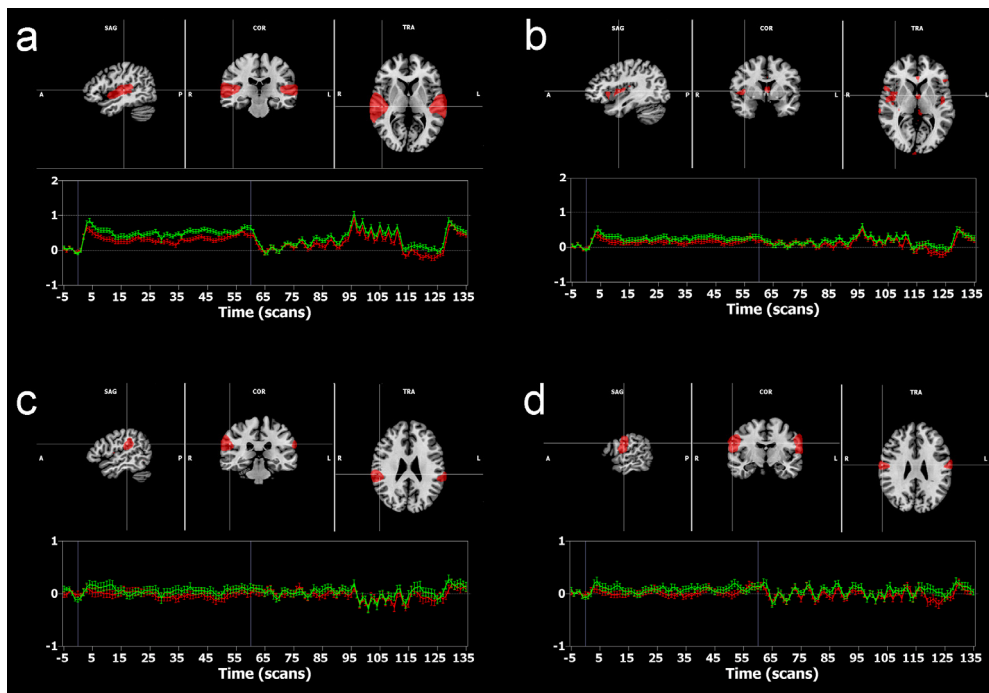


Figure 7. Independent components involved in auditory and motor processing. Displayed is an overlay on the MNI Colin27 template.

A second auditory component (fig.7b) consists mainly of regions within the Sylvian Fissure and exhibits a response pattern that is comparable to the first auditory component in shape, although the modulation is less pronounced. Also, it demonstrates less consistency between individual subjects.

A third auditory component represents regions within bilateral supramarginal gyrus (SMG; fig.7c). The SMG has been linked to short-term memory processes for musical information (Gaab, Gaser, Zaehle, Jancke, & Schlaug, 2003; Koelsch, Fritz, Schulze, Alsop, & Schlaug, 2005; Vines, Schneider, & Schlaug, 2006). This role fits well with the current paradigm and the time course represented in the ERA. More specifically, since participants are required to memorize the musical pieces and later indicate whether or not musical excerpts were part of those pieces, it is conceivable that this component displays an increased response during the presentation of the music, which stays elevated until the presentation of the musical excerpts, which then modulate the response.

The motor component in figure 7d is activated during music listening. The involvement of motor cortex in the perception of music has been shown before (Zatorre, Chen, & Penhune, 2007). Furthermore, it is activated by the videos, which may be partly due to the button press response, but may also be related to mimicry responses toward seeing another person's emotional facial expressions (Lee, Josephs, Dolan, & Critchley, 2006; Schilbach, Eickhoff, Mojzisch, & Vogeley, 2008).

Fronto-parietal independent components. The components displayed in figure 8a and 8b represent lateralized mirror-images of each other. This lateralization has been demonstrated before (van de Ven et al., 2004). Together, these networks have been linked to attention and cognitive control (Leech, Kamourieh, Beckmann, & Sharp, 2011; Vincent, Kahn, Snyder, Raichle, & Buckner, 2008). These processes are essential in the present paradigm to some extent during the recognition of the musical excerpts, but mainly during the viewing, and stopping at an appropriate time point, of the videos. This corresponds to the components' time courses, which show a small, prolonged response to the musical excerpts and a clear modulation by the individual videos.

The third fronto-parietal component is strongly left-lateralized and includes inferior frontal gyrus (IFG) and posterior superior temporal sulcus (STS) / angular gyrus. The identity of these regions and the strong lateralization point toward a language function of this network. The human language network has been shown before to be involved music processing (Koelsch, 2011; Koelsch, et al., 2002), which might be a reason for its involvement in the present paradigm. However, the fact that this component exhibits the strongest modulation during the videos suggests that it is involved in other functions as

well. One possible explanation rests on a proposed distinction between a dorsal and a ventral attention system (Fox, Corbetta, Snyder, Vincent, & Raichle, 2006). While the lateralized components in figure 8a and 8b may correspond to the dorsal attention system, which is supposedly involved in the top-down orienting of attention, this component might correspond to the ventral attention system, which has been suggested to be involved in reorienting attention toward salient sensory stimuli. It has to be noted, however, that this network is usually found to be lateralized toward the right, while here it is clearly left-lateralized.

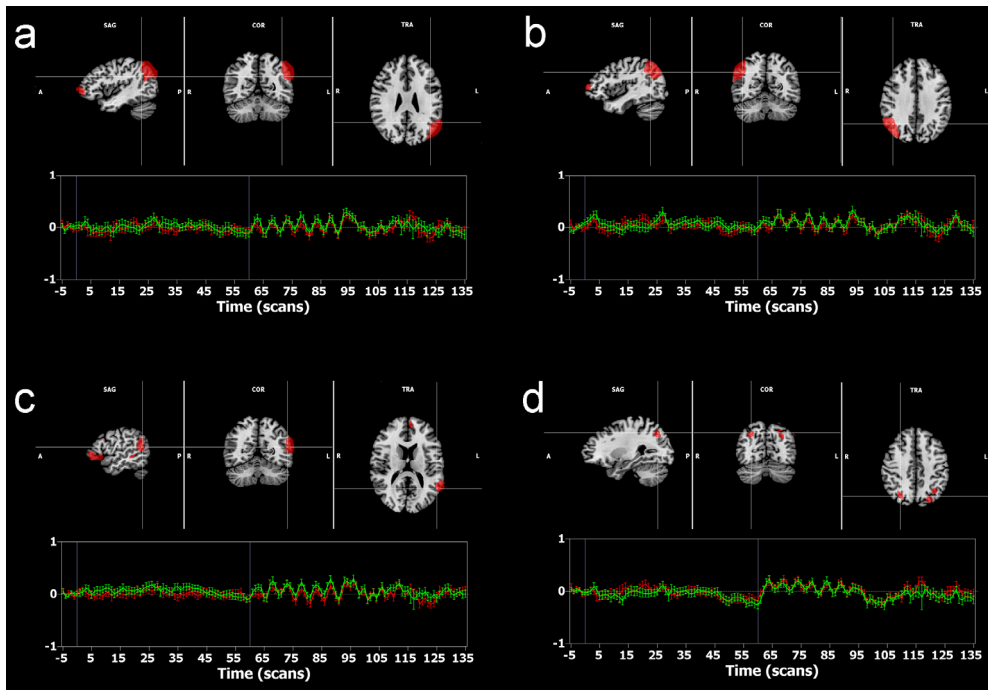


Figure 8. Fronto-parietal independent components. Displayed is an overlay on the MNI Colin27 template.

Another parietal component (fig.8d) includes bilateral intraparietal sulcus and is activated during the videos. The intraparietal sulcus has been suggested to be involved in dynamic control of action and the representation of intended action goals (Tunik, Rice, Hamilton, & Grafton, 2007). This fits with the current experiment, since participants were required to transform visual information into action (stop the videos). Furthermore,

emotional faces may convey the actors' intended goals to a certain extent. The representation of these perceived goals may also play a role here.

Independent components involved in emotion processing. The limbic-prefrontal component shown in figure 9a includes anterior cingulate cortex, posterior cingulate cortex, bilateral dorsolateral prefrontal cortex, amygdala, and insula. Its functional roles, both general and specific to the present experimental context, are examined in the main part of this chapter.

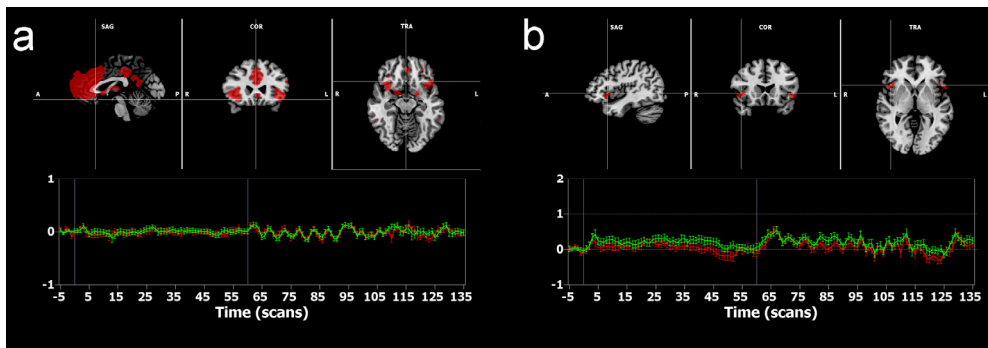


Figure 9. Emotion independent components. Displayed is an overlay on the MNI Colin27 template.

The other component that can be linked to emotion processing consists of bilateral anterior insula cortex (fig.9b). This region has been linked to a multitude of different cognitive abilities, which makes unequivocal interpretations cumbersome. However, anterior insula has consistently been implicated in the processing of emotions and, importantly, mood induction (Phan, Wager, Taylor, & Liberzon, 2002), which agrees with both the rationale of the present study and the ERA. Furthermore, it has been implicated in empathy (Singer, 2006), which may explain its modulation during the viewing of emotional facial expressions.

Midline independent components. The most consistent independent component involving midline structures is the default-mode network (fig.10a; Raichle, et al., 2001). Its functional roles, both general and specific to the present experimental context, are examined in the main part of this chapter.

The other three midline components (fig.10b-d) include regions in the precuneus and cuneus and exhibit clear modulations during the video presentation. Ascribing specific, individual roles to these components is not straightforward. However, visual and visuo-motor functions are likely.

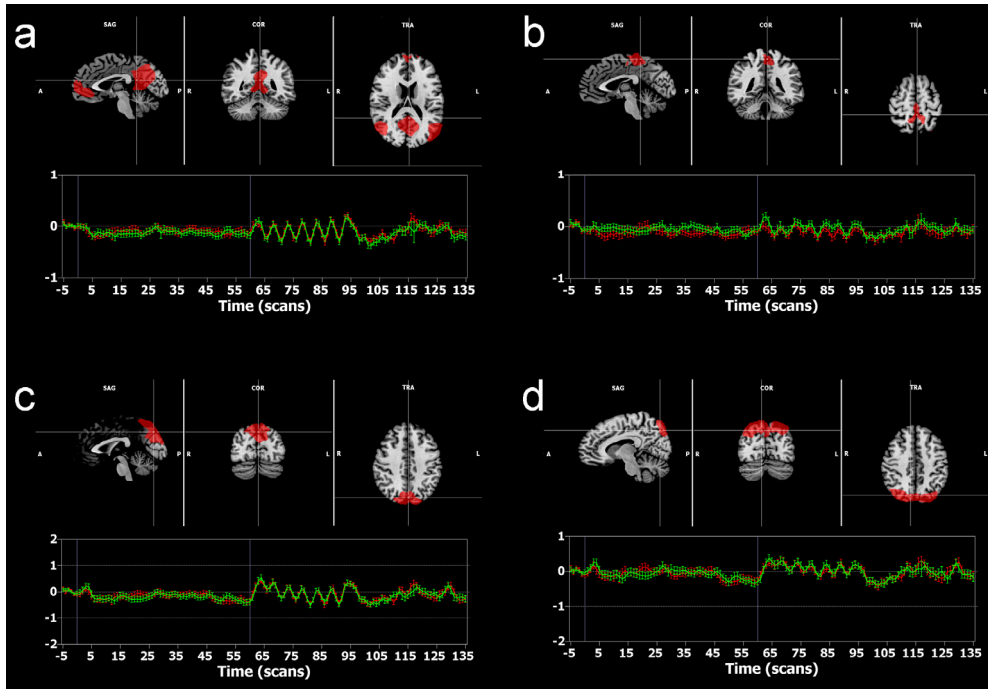


Figure 10. Midline independent components. Displayed is an overlay on the MNI Colin27 template.

Miscellaneous independent components. The component shown in figure 11a is rather noisy, but includes left superior frontal gyrus (SFG) as the most robust region. This region has been suggested to be involved in shifts of attention and cognitive sets (Nagahama, et al., 1999; Rushworth, Walton, Kennerley, & Bannerman, 2004). Careful investigation of the ERA shows the two biggest peaks of activation at the transition points between cognitive sets- moving from passive music listening to the video task, and from the video task to identifying musical excerpts. This fits nicely with the proposed role. The elevated response during the musical excerpt task may be explained by the SFG's suggested role in working memory (du Boisgueheneuc, et al., 2006).

The component displayed in figure 11b represents right-lateralized posterior temporal-occipital cortex. This region has been linked to passive viewing of biological motion, and more specifically, mouth movements (Pelphrey, Morris, Michelich, Allison, & McCarthy, 2005), explaining its involvement in the viewing of the videos.

The component in figure 11c includes bilateral posterior parietal lobe and exhibits a prolonged plateau response during the videos. It is difficult to pinpoint the exact role of this component in the task. However, one possible explanation is its suggested involvement in imitation (Molenberghs, Cunnington, & Mattingley, 2009), a subtle and involuntary process when perceiving emotional facial expressions.

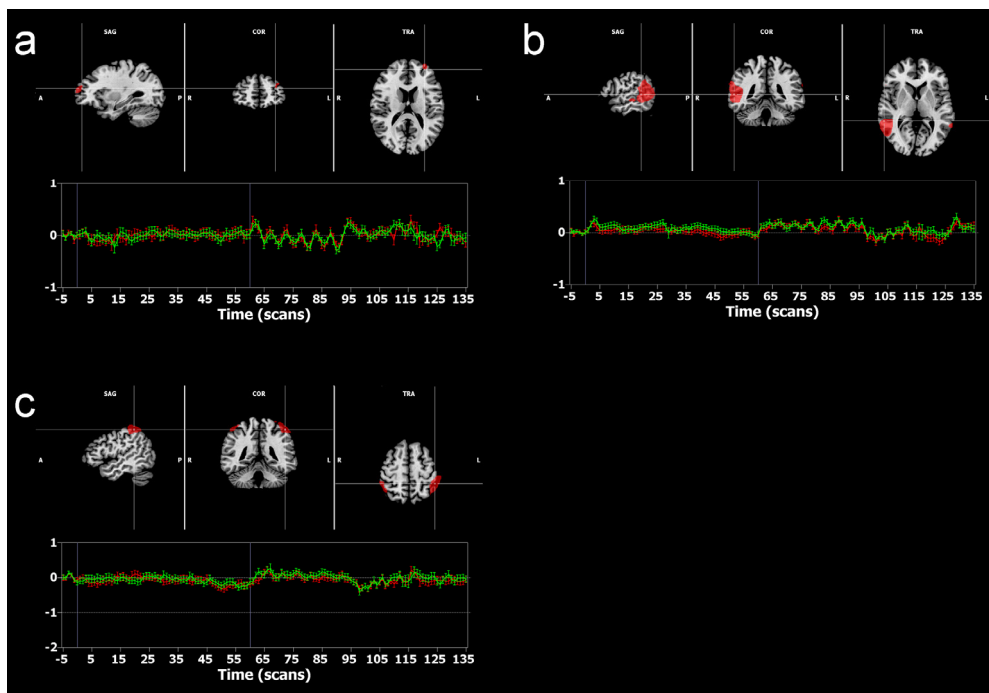


Figure 11. Miscellaneous independent components. Displayed is an overlay on the MNI Colin27 template.

References

- Beck, A. T. (1967). *Depression: Clinical, Experimental, and Theoretical Aspects*. New York: Harper & Row.
- Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates. *Am J Psychiatry*, *165*(8), 969-977.
- Bennett, M. R. (2011). The prefrontal-limbic network in depression: Modulation by hypothalamus, basal ganglia and midbrain. *Prog Neurobiol*, *93*(4), 468-487.
- Bouhuys, A. L., Bloem, G. M., & Groothuis, T. G. G. (1995). Induction of depressed and elated mood by music influences the perception of facial emotional expressions in healthy subjects. *Journal of Affective Disorders*, *33*(4), 215-226.
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *J Behav Ther Exp Psychiatry*, *25*(1), 49-59.
- Carballedo, A., Scheuerecker, J., Meisenzahl, E., Schoepf, V., Bokde, A., Moller, H. J., et al. (2011). Functional connectivity of emotional processing in depression. *J Affect Disord*, *134*(1-3):272-9.
- Carter, E. J., Hodgins, J. K., & Rakison, D. H. (2011). Exploring the neural correlates of goal-directed action and intention understanding. *Neuroimage*, *54*(2), 1634-1642.
- Damoiseaux, J. S., Rombouts, S. A., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., et al. (2006). Consistent resting-state networks across healthy subjects. *Proc Natl Acad Sci U S A*, *103*(37), 13848-13853.
- Disner, S. G., Beevers, C. G., Haigh, E. A., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nat Rev Neurosci*, *12*(8), 467-477.
- du Boisgueheneuc, F., Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., et al. (2006). Functions of the left superior frontal gyrus in humans: a lesion study. *Brain*, *129*(Pt 12), 3315-3328.
- Esposito, F., Aragri, A., Pesaresi, I., Cirillo, S., Tedeschi, G., Marciano, E., et al. (2008). Independent component model of the default-mode brain function: combining individual-level and population-level analyses in resting-state fMRI. *Magn Reson Imaging*, *26*(7), 905-913.
- Esposito, F., & Goebel, R. (2011). Extracting functional networks with spatial independent component analysis: the role of dimensionality, reliability and aggregation scheme. *Curr Opin Neurol*, *24*(4), 378-385.

- Esposito, F., Scarabino, T., Hyvarinen, A., Himberg, J., Formisano, E., Comani, S., et al. (2005). Independent component analysis of fMRI group studies by self-organizing clustering. *Neuroimage*, 25(1), 193-205.
- Ethofer, T., Gschwind, M., & Vuilleumier, P. (2011). Processing social aspects of human gaze: a combined fMRI-DTI study. *Neuroimage*, 55(1), 411-419.
- Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med*, 33(5), 636-647.
- Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc Natl Acad Sci U S A*, 103(26), 10046-10051.
- Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, 6(3), 218-229.
- Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., et al. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *J Psychiatry Neurosci*, 34(6), 418-432.
- Gaab, N., Gaser, C., Zaehle, T., Jancke, L., & Schlaug, G. (2003). Functional anatomy of pitch memory--an fMRI study with sparse temporal sampling. *Neuroimage*, 19(4), 1417-1426.
- Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: current status and future directions. *Annu Rev Clin Psychol*, 6, 285-312.
- Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc Natl Acad Sci U S A*, 100(1), 253-258.
- Hamilton, J. P., Furman, D. J., Chang, C., Thomason, M. E., Dennis, E., & Gotlib, I. H. (2011). Default-mode and task-positive network activity in major depressive disorder: implications for adaptive and maladaptive rumination. *Biol Psychiatry*, 70(4), 327-333.

- Hoffman, K. L., Ghazanfar, A. A., Gauthier, I., & Logothetis, N. K. (2007). Category-specific responses to faces and objects in primate auditory cortex. *Front Syst Neurosci*, 1, 2.
- Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw*, 10(3), 626-634.
- Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent Component Analysis*. New York: Wiley.
- Koelsch, S. (2011). Toward a neural basis of music perception - a review and updated model. *Front Psychol*, 2, 110.
- Koelsch, S., Fritz, T., Schulze, K., Alsop, D., & Schlaug, G. (2005). Adults and children processing music: an fMRI study. *Neuroimage*, 25(4), 1068-1076.
- Koelsch, S., Gunter, T. C., v Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: a cortical "language-network" serves the processing of music. *Neuroimage*, 17(2), 956-966.
- Kohler, C. G., Hoffman, L. J., Eastman, L. B., Healey, K., & Moberg, P. J. (2011). Facial emotion perception in depression and bipolar disorder: A quantitative review. *Psychiatry Res*, 188(3), 303-309.
- Kret, M. E., Pichon, S., Grezes, J., & de Gelder, B. (2011). Similarities and differences in perceiving threat from dynamic faces and bodies. An fMRI study. *Neuroimage*, 54(2), 1755-1762.
- Lee, T. W., Josephs, O., Dolan, R. J., & Critchley, H. D. (2006). Imitating expressions: emotion-specific neural substrates in facial mimicry. *Soc Cogn Affect Neurosci*, 1(2), 122-135.
- Leech, R., Kamourieh, S., Beckmann, C. F., & Sharp, D. J. (2011). Fractionating the default mode network: distinct contributions of the ventral and dorsal posterior cingulate cortex to cognitive control. *J Neurosci*, 31(9), 3217-3224.
- Mathews, A., & MacLeod, C. (2005). Cognitive vulnerability to emotional disorders. *Annu Rev Clin Psychol*, 1, 167-195.
- Mayberg, H. S., Liotti, M., Brannan, S. K., McGinnis, S., Mahurin, R. K., Jerabek, P. A., et al. (1999). Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. *Am J Psychiatry*, 156(5), 675-682.

- Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2009). Is the mirror neuron system involved in imitation? A short review and meta-analysis. *Neurosci Biobehav Rev*, 33(7), 975-980.
- Nagahama, Y., Okada, T., Katsumi, Y., Hayashi, T., Yamauchi, H., Sawamoto, N., et al. (1999). Transient neural activity in the medial superior frontal gyrus and precuneus time locked with attention shift between object features. *Neuroimage*, 10(2), 193-199.
- Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *J Abnorm Psychol*, 109(3), 504-511.
- Pacchetti, C., Mancini, F., Aglieri, R., Fundarò, C., Martignoni, E., & Nappi, G. (2000). Active music therapy in Parkinson's disease: An integrative method for motor and emotional rehabilitation. *Psychosomatic Medicine*, 62(3), 386-393.
- Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2004). Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *J Cogn Neurosci*, 16(10), 1706-1716.
- Pelphrey, K. A., Morris, J. P., Michelich, C. R., Allison, T., & McCarthy, G. (2005). Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. *Cereb Cortex*, 15(12), 1866-1876.
- Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, 16(2), 331-348.
- Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003a). Neurobiology of emotion perception I: The neural basis of normal emotion perception. *Biol Psychiatry*, 54(5), 504-514.
- Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003b). Neurobiology of emotion perception II: Implications for major psychiatric disorders. *Biol Psychiatry*, 54(5), 515-528.
- Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proc Natl Acad Sci U S A*, 98(2), 676-682.

- Rigucci, S., Serafini, G., Pompili, M., Kotzalidis, G. D., & Tatarelli, R. (2010). Anatomical and functional correlates in major depressive disorder: the contribution of neuroimaging studies. *World J Biol Psychiatry*, *11*(2 Pt 2), 165-180.
- Rushworth, M. F., Walton, M. E., Kennerley, S. W., & Bannerman, D. M. (2004). Action sets and decisions in the medial frontal cortex. *Trends Cogn Sci*, *8*(9), 410-417.
- Said, C. P., Moore, C. D., Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J Vis*, *10*(5), 11.
- Schiffenbauer, A. (1974). Effect of observer's emotional state on judgments of the emotional state of others. *Journal of Personality and Social Psychology*, *30*(1), 31-35.
- Schilbach, L., Eickhoff, S. B., Mojzisch, A., & Vogeley, K. (2008). What's in a smile? Neural correlates of facial embodiment during social interaction. *Soc Neurosci*, *3*(1), 37-50.
- Sheline, Y. I., Barch, D. M., Price, J. L., Rundle, M. M., Vaishnavi, S. N., Snyder, A. Z., et al. (2009). The default mode network and self-referential processes in depression. *Proc Natl Acad Sci U S A*, *106*(6), 1942-1947.
- Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neurosci Biobehav Rev*, *30*(6), 855-863.
- Soto, D., Funes, M. J., Guzmán-García, A., Warbrick, T., Rotshtein, P., & Humphreys, G. W. (2009). Pleasant music overcomes the loss of awareness in patients with visual neglect. *Proc Natl Acad Sci U S A*, *106*(14), 6011-6016.
- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., et al. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry Res*, *168*(3), 242-249.
- Tunik, E., Rice, N. J., Hamilton, A., & Grafton, S. T. (2007). Beyond grasping: representation of action in human anterior intraparietal sulcus. *Neuroimage*, *36* Suppl 2, T77-86.
- van de Ven, V., Esposito, F., & Christoffels, I. K. (2009). Neural network of speech monitoring overlaps with overt speech production and comprehension networks: a sequential spatial and temporal ICA study. *Neuroimage*, *47*(4), 1982-1991.

- van de Ven, V. G., Formisano, E., Prvulovic, D., Roeder, C. H., & Linden, D. E. (2004). Functional connectivity as revealed by spatial independent component analysis of fMRI measurements during rest. *Hum Brain Mapp*, 22(3), 165-178.
- Vincent, J. L., Kahn, I., Snyder, A. Z., Raichle, M. E., & Buckner, R. L. (2008). Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. *J Neurophysiol*, 100(6), 3328-3342.
- Vines, B. W., Schnider, N. M., & Schlaug, G. (2006). Testing for causality with transcranial direct current stimulation: pitch memory and the left supramarginal gyrus. *Neuroreport*, 17(10), 1047-1050.
- Widiger, T. A., & Samuel, D. B. (2005). Diagnostic categories or dimensions? A question for the Diagnostic And Statistical Manual Of Mental Disorders--fifth edition. *J Abnorm Psychol*, 114(4), 494-504.
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nat Rev Neurosci*, 8(7), 547-558.

SUMMARY

When we perceive the world around us, we usually feel that we have access to a direct sensory image of reality. This impression, however, is an illusion. Actually, our perception is heavily influenced by a plurality of contextual factors; including the interaction between different sensory modalities (e.g. what we see may alter what we hear). This thesis, titled “Cross-modal effects in the construction of perception” examines these interactions at different levels.

The first part (chapters 2 and 3) investigates how seeing lip movements can alter our perception of what we hear is being said and how the varying perceptual interpretations of physically identical auditory stimuli can be predicted on the basis of brain imaging data.

The second part (chapters 4 and 5) is concerned with audiovisual interactions at a very different level of processing and examines in how far music can influence visual emotion perception of facial expressions, mirroring cognitive symptoms of clinical depression. Furthermore, it is investigated how neural networks processing different cognitive aspects of clinical depression interact dynamically in healthy participants undergoing musical mood induction.

Chapter 1 briefly introduces the ideas of constructivism and the differentiation between sensation and perception and summarizes the state of research in multisensory neuroscience. The main methods employed in the research represented in this thesis are shortly described.

In *Chapters 2 and 3* it is demonstrated that the entanglement of sensation and perception that is usually inherent to neuroscientific investigations of perception can be overcome employing a concept called cross-modal recalibration. This effect entails the presentation of pairings of ambiguous sounds with unambiguous lip movements, effectively disambiguating the auditory percept. Furthermore, it has been shown that repeated exposure to these pairings induces a lasting perceptual bias, which affects ambiguous auditory stimuli presented later and in isolation. This provided the opportunity to inspect neural responses to stimuli that were perceived differently, while being physically identical. *Chapter 2* demonstrates, using functional magnetic resonance imaging (fMRI) and pattern recognition techniques, that it is possible to retrieve the perceptual interpretation of ambiguous phonemes – information that is fully subjective to the listener – from measurements of brain activity in auditory areas in the superior temporal cortex, most prominently on the

posterior bank of the left Heschl's gyrus and sulcus and in the adjoining left planum temporale. These findings suggest that - beyond the basic acoustic analysis of sounds - constructive perceptual processes take place in these relatively early cortical auditory networks.

Instead of analyzing the neural responses to the isolated auditory stimuli, *Chapter 3* deals with brain activation during the recalibration phase (the audiovisual pairings) in order to examine the origin of the perceptual bias that was shown in *Chapter 2* to be stored in auditory cortex. Utilizing behavioral data from the perceptual classifications of the ambiguous auditory phonemes that followed the recalibration phase later in time, we could identify a network of brain areas (bilateral inferior parietal lobe [IPL], right [and possibly left] inferior frontal sulcus [IFS], and right posterior middle temporal gyrus [MTG]), whose activation during audiovisual exposure predicted auditory perceptual tendencies later in time. It is proposed that represented here are audiovisual integrative learning processes which are responsible for the installation of the perceptual bias in auditory regions.

In *Chapter 4*, two behavioral studies are presented. Both studies employed videos which were synthesized by morphing facial expressions from sad or happy expressions, respectively, to neutral ones. The participants' task was to stop the videos when the original expression was not recognized anymore. This rationale overcomes some problems faced by studies using explicit ratings, which may be confounded by demand characteristics. In the first study, it is demonstrated that depressed patients exhibited an emotion perception bias, in that they perceived sad facial expressions to persist longer on the changing faces than controls. Study 2 utilized sad and happy pieces of classical music (pre-rated for emotional valence by independent participants) in order to differentially induce happy and sad mood states in healthy participants. It could be demonstrated that healthy participants with an affinity to classical music showed an emotion perception bias when undergoing sad mood induction, which mimicked that of depressed patients. These results suggest that acute and pathological sad mood can induce similar cognitive symptoms, a notion which is compatible with a dimensional conceptualization of depression. Furthermore, they emphasize the impact that music can have on the way we perceive the world, calling for future research into potential benefits of music in therapeutic settings.

In *Chapter 5*, a similar set-up to the one in the second study of *Chapter 4* was transferred to the fMRI environment. Again, participants were exposed to sad and happy classical music in order to investigate the neural networks underlying the mimicked cognitive aspects of depression. Using independent component analysis (ICA) and a combination of ICA and psychophysiological interaction (PPI), it could be revealed that the neural substrates underlying rumination and biased emotion perception are linked selectively during sad mood via the posterior superior temporal sulcus. These results indicate that separate cognitive symptoms of depression interact neurally, which may result in a vicious cognitive circle, and that similar neural substrates may underlie acute and pathological sad mood, suggesting a dimensional concept of depression.

SAMENVATTING

Als we de wereld om ons heen waarnemen, hebben we al snel het gevoel dat we direct toegang hebben tot de sensorische representatie van onze realiteit. Toch blijkt deze impressie een illusie te zijn. Onze waarneming wordt flink beïnvloed door meerdere contextuele factoren, waaronder de interactie tussen verschillende sensorische modaliteiten (bijvoorbeeld, wat je ziet kan veranderen wat je hoort). Dit proefschrift, getiteld ‘Cross-modal Effects in the Construction of Perception’ onderzoekt deze interactie op verschillende niveaus.

Het eerste deel (hoofdstukken 2 en 3) onderzoekt hoe het zien van lipbewegingen kan beïnvloeden wat we iemand horen zeggen. Daarnaast wordt onderzocht hoe de variërende perceptuele interpretatie van fysiek identieke auditieve stimuli kan worden voorspeld op basis van hersenactiviteit.

Het tweede deel (hoofdstukken 4 en 5) richt zich op audiovisuele interacties op een heel ander niveau van verwerking. Hier wordt gekeken in hoeverre muziek invloed heeft op het waarnemen van emotionele gezichtsuitdrukkingen, als afspiegeling van enkele cognitieve symptomen van depressie. Verder wordt bij gezonde proefpersonen onderzocht hoe neurale netwerken die betrokken zijn bij verschillende cognitieve aspecten van depressie op een dynamische wijze met elkaar interacteren tijdens deze muzikale gemoedstoestandinductie.

Hoofdstuk 1 beschrijft kort de ideeën achter constructivisme en het onderscheid tussen gewaarwording en perceptie, en vat de huidige staat van het onderzoek naar multisensorische neurowetenschappen samen. De onderzoeksmethoden die hoofdzakelijk gebruikt worden in het onderzoek dat dit proefschrift beschrijft zullen kort worden toegelicht.

In *Hoofdstukken 2 en 3* wordt gedemonstreerd dat de samenhang tussen gewaarwording (sensation) en perceptie (perception), zoals vaak gezien in neurowetenschappelijk onderzoek, overmeesterd kan worden door een concept genaamd ‘cross-modal recalibration’. Dit effect houdt in dat als er ambigue geluiden tegelijk worden gepresenteerd met onambigue lipbewegingen, het auditieve percept geïnduceerd wordt. Daarbij is gedemonstreerd dat herhaaldelijke waarneming van dergelijke stimulusparen tot een blijvende perceptuele bias kan leiden, die ambigue auditieve stimuli kan beïnvloeden die later en in isolatie gepresenteerd worden. Dit heeft de mogelijkheid gecreëerd om de

neurale activiteit te onderzoeken die hoort bij stimuli die verschillend waargenomen worden, maar feitelijk fysiek identiek zijn. *Hoofdstuk 2* demonstreert aan de hand van fMRI en patroonherkenning dat het mogelijk is om de perceptuele interpretatie van ambigue fonemen te achterhalen - deze informatie is volledig subjectief voor de luisteraar - uit metingen van hersenactiviteit in auditieve gebieden van de superior temporal cortex, voornamelijk de posterior bank van linker Heschl's gyrus en sulcus, en in de naastliggende planum temporale. Deze bevindingen laten zien dat er, naast de reguliere akoestische analyse van geluiden, constructieve perceptuele processen plaatsvinden in deze relatief vroege corticale auditieve netwerken.

In plaats van de neurale reacties op de geïsoleerde auditieve stimuli te analyseren, richt *Hoofdstuk 3* zich op de hersenactiviteit tijdens de recalibratiefase (de audiovisuele paren) om zo te onderzoeken waar de perceptuele bias die gedemonstreerd werd in *Hoofdstuk 2* werd opgeslagen in de auditieve cortex. Aan de hand van gedragsdata die verkregen was uit perceptuele classificaties van de ambigue auditieve fonemen die ná de recalibratiefase plaatsvonden, konden we een netwerk van hersengebieden (bilaterale inferior parietal lobe [IPL], rechter [en mogelijk linker] inferior frontal sulcus [IFS] en rechter posterior middle temporal gyrus [MTG]) identificeren, wiens activiteit tijdens de presentatie van audiovisuele de hierop volgende perceptie kon voorspellen. We hypothetiseren dat dit audiovisuele integrerende leerprocessen voorstellen die verantwoordelijk zijn voor het installeren van een perceptuele bias in auditieve gebieden.

In *Hoofdstuk 4* worden twee gedragsstudies gepresenteerd. Beide studies maken gebruik van video's die gesynthetiseerd waren door gezichtsuitdrukkingen te laten transformeren van verdrietig of blij, respectievelijk, naar een neutrale uitdrukking. De taak van de proefpersoon was om de video te stoppen op het moment dat de oorspronkelijke emotionele uitdrukking niet meer te herkennen was. Deze benadering omzeilt enkele problemen waar andere studies, die gebruik maakten van een expliciete scoringsmethode, tegenaan liepen. In de eerste studie wordt gedemonstreerd dat depressieve patiënten een emotie-perceptie bias hebben, waarbij ze droevige gezichtsuitdrukkingen langer waarnemen op de transformerende gezichten dan de controlegroep. Studie 2 maakte gebruik van droevige en vrolijke klassieke muziek (van tevoren beoordeeld op emotionele lading door onafhankelijke proefpersonen) om zo vrolijke en droevige gemoedstoestanden te induceren

bij gezonde proefpersonen. Er kon worden aangetoond dat gezonde proefpersonen die een affiniteit hebben met klassieke muziek een emotionele waarnemingsbias lieten zien wanneer ze een droevige gemoedstoestandinductie hadden ondergaan. Dit kwam overeen met de bias die depressieve patiënten lieten zien. Deze resultaten suggereren dat acute en pathologische droevige gemoedstoestanden gelijke cognitieve symptomen kunnen induceren. Deze notie is compatibel met een dimensionale conceptualisatie van depressie. Daarnaast laten ze de impact zien die muziek heeft op de manier waarop we de wereld zien en geeft zo aan dat toekomstig onderzoek naar de potentiële voordelen van muziek in een therapeutische setting wenselijk is.

In *Hoofdstuk 5* werd een analoge opzet als de tweede studie van *Hoofdstuk 4* omgezet naar een fMRI studie. Ook hier werden proefpersonen blootgesteld aan droevige en vrolijke klassieke muziek om zo de neurale netwerken die de cognitieve aspecten van depressie nabootsten te kunnen onderzoeken. Met gebruik van independent component analysis (ICA) en een combinatie van ICA en psychofysiologische interactie (PPI), kon aangetoond worden dat het neurale substraten die aan getob (rumination) en de emotieperceptie bias ten grondslag liggen, selectief bij een droevige gemoedstoestand via de posterior superior temporal sulcus verbonden zijn. Deze resultaten geven een indicatie dat aparte cognitieve symptomen van depressie neuraal interacteren, wat zou kunnen resulteren in een vicieuze cognitieve cirkel. Daarnaast zou dit kunnen betekenen dat er gelijke neurale substraten ten grondslag kunnen liggen aan acute en pathologische droevige stemmingen, wat wijst op een dimensionaal concept van depressie.

Vertaling / translation: Job van den Hurk

ACKNOWLEDGEMENTS

My predecessors have warned me that- while more than 4 years of scientific work may be tough- the toughest part is writing the acknowledgements. I will proceed, in a constant fear of forgetting important people, not finding the right words and alienating readers because of the order and length of these words of gratitude, in thanking my supervisors, colleagues, family, and friends.

Dear Elia, I owe so much to your mentorship, your supervision, and your friendship that I can truthfully say, there is no way I could have done this without you. Thank you very much for all your help, your openness even for ideas and plans that were not quite up your alley, your support, and for being a true role model. In times that were not always easy for you and your family, you never seemed to be there for us and to spread optimism and enthusiasm. Much of my decision to pursue this PhD in the first place was due to having met you during the third-year course Research Methods and much of the reason for never having regretted it is due to your continued support. Although you got busier through the years, I always knew I had two aces up my sleeve to get your attention- talking about Luca and talking about S.S.C. Napoli. I am very happy to see that both are doing so well!

Dear Rainer, although we did not work together on an everyday basis, you were always a role model for me due to your enthusiasm and incredible knowledge in the field. The meetings that we had were invaluable and stimulating and had a big impact on how I evaluated and adjusted my research plans. Thank you for always having an open ear in spite of your incredibly busy schedule and for co-creating such a positive and stimulating work atmosphere.

Dear Jean, thank you for your support and our collaboration. The ingenuity of your behavioral work was essential for a big chunk of this thesis. You were always very trusting and supportive and I appreciate that a lot.

Thank you, Giancarlo, first and foremost, for being a true friend in the office next door. I have always thoroughly enjoyed our sarcasm and humor, your love for the wrong kinds of music, your cooking, and your calm friendliness even in the face of devastating losses at FIFA. However, you were not only a friend, but also my unofficial third supervisor. Your humble way of explaining and re-explaining concepts over and over without a hint of disdain was essential in enabling me to run my research the way I wanted to. I am proud of our collaborative publication!

Martin van Frost, my brother from another mother- we were in this together, from the beginning of the Research Master program right up until now, and I am very grateful for that. It makes me happy that we went through all these different phases of our professional and personal development together and that the foundation of our friendship was never once shaken. You were, and still are, my go-to guy for humor, diversion, talking s**t, talking business, self-doubts, self-aggrandizement, urbanity, sophistication, savoir vivre, and fundamental debates about science, religion, moral, politics and ideals. If I were religious, you would be my leader. Even though I can never be your real friend ;-), you are one of my best friends fo' shnizzle. I will hurtfully miss our meticulous, point-by-point reports of our respective weekends or holidays and I will never be able to hear the words "did you know" the same way again. Thank you, homeboy.

Prosciutto the gentle, you also belong into this category- true friends. You are the only one who I did all the following with: had suits fitted by a 50-year old gay gentleman over glühwein before drunkenly confessing terrible deeds and heartache, played table tennis at work until soaked, played table tennis in a bar in Manhattan at two in the morning until soaked, swam with a freaking whale (!), expressive danced and sang My Heart Will Go On, ate too many Magnus Mints, sang Girl from Ipanema in the streets of Barcelona, got drunk at work, got drunk at concerts, got drunk on a boat, planned the communist revolution. How many people can one say one did this with? Thank you, my friend.

Tom, thank you for our FIFA nights and for enduring my childish rants, for our road trip, and for not dying on the most under-prepared hike in Yosemite history. Thank you, Alard for being an amazing roommate and for enduring German phone calls and weekend planning, loud music, and people invading our office in search of Nespresso. Thank you, Fabrizio, for all your help. Your enthusiasm and ideas have helped me a lot and I am glad to have you as my third Italian mentor. Thank you, Federico. You also helped me quite a number of times, which makes you Italian mentor #4. Also: poker, pizza, champions league. Thank you, Job, for translations, poker nights and for doing a great job in general (sorry ☺). Thank you, Annemie, Christl, and Riny, for your invaluable support. You have always helped me and just how indispensable you are became clear once more in the organization of all THIS! Lars, Sven, and Tim thank you for letting me kick your a**es on the squash court and for feeling like a banker from the 80s doing it and thanks for kicking my a** on the soccer court (same to the rest of my soccer friends). Thank you, Aline. Although we did not share enough parties and nights out, you made a lot of my work days a

little friendlier. Thank you, Roberta, Marin, Anke, Michelle, Britta, Kamil, Matteo, Pegah, Milene, Marieke, Vincent, Joël, Judith, Amanda, Lars, Bettina, Valerie, Leo and all the other colleagues for the great atmosphere in our department! Thank you for all your advice (I have learned from all of you) and thank you for all the fun! From international dinners, parties, and poker sessions to drunken nights in Melbourne, San Francisco, Barcelona, and Québec City- it's been a blast and I will truly miss it! Nienke, thank you so much for letting me walk in your foot steps! You taught me about multimodal processing way back in our little reading club, you introduced me to everyone in New York, and now you were in the reading commission of my thesis. Thank you, Bernie, for your kindness and your constant support. You were a great teacher, head of the department, dean, mentor, and head of my corona. You always showed me that you liked me and were truly happy for my successes. Thank you, Peter, for being a great teacher and colleague and for being in my reading committee. I am sure Belgian soccer will pick up again ☺. Thank you, Fren, for the discussions and insights. Thank you to Michael Beauchamp for admitting me to the defense of my thesis!

Dear Alex and dear Nassim, Mariam, Sarah, Yvonne, Henry, Roz, Cyrus, Justine, Atusa, Carlos and all my other Californian friends, thank you for an incredible time in LA in 2004/2005 and thank you for still being my friends now (as impressively demonstrated last year at Alex's and Nassim's wedding). I hope to see you in NYC frequently!

Dear Aude, if I had mentioned you as a side note to my thanks to Martin, it would not have done our friendship justice. Thank you for being my friend in your own right, for running together, eating together, laughing together, for your kindness and interest, for enduring our FIFA nights and for being the perfect fiancé for my friend.

Time to switch to German...

Ich war ziemlich lange in Maastricht und zwischendurch war es ganz schön hart, da zu bleiben, als meine Herzensfreunde aus den ersten Jahren die Stadt verließen. Zusammen kochen, feiern, Wein trinken, Allgemeinbildungsabende und Brunches zelebrieren, lernen und vieles mehr wird nie wieder denselben magischen Touch haben wie in dieser Zeit. Liebe Hanna, du hast für mich eine Ära geprägt und warst immer der wichtigste Mensch für mich in Maastricht. Auch wenn es nicht immer leicht war, bin ich sehr froh, dass wir uns immer noch so nahe stehen und, allen Befürchtungen zum Trotz,

diese Freundschaft zwischen uns weiter haben gedeihen lassen. Danke, dass Du immer da warst und immer da bist. Du weißt, dass dasselbe auch umgekehrt gilt und immer gelten wird. Rebeccs, Du bist zum Glück zurückgekommen und das war irrsinnig wichtig für mich. Was hätte ich mit all der Zeit gemacht, die wir durch die Wälder Aachens und die Wiesen Maastrichts gelaufen sind? Mit wem hätte ich wann diese Gespräche geführt, den Alltag und Tiefergehendes besprochen? Mit wem wäre ich von der Arbeit nach Hause (!), durch den Central Park und vor allem den Kölnmarathon gelaufen, hätte ich in der Weinbar an der Upper West Side Stunden verbracht und New York erkundet? Danke für alles, Beccs! Ich werde das sehr vermissen und hoffe, wir bleiben nicht ewig in Kiel und New York, sondern wohnen irgendwann wieder in derselben Stadt. Toni (Lubert), Du bist ein wahrer Freund und ich habe gerade bei meinem Besuch in Heidelberg noch mal gedacht, wie sehr ich Dich vermisse. Mit niemandem kann ich mich so intellektuell reiben, Gitarre spielen und singen, und über Leben und Liebe philosophieren. Ich habe großen Respekt vor Dir und freue mich, dass wir immer noch eine so gute Art miteinander haben. Anne Schü., danke Dir für tiefe Gespräche, für Parties in der Frankenstraat, und für Deine Freundschaft. Ich bin froh, dass ich bald eine Zahnärztin habe, auf die ich mich so verlassen kann. Jules, ganz am Anfang warst Du noch gar nicht so in unserem kleinen Kreis, dafür aber dann mit Kawumm! Ich weiß noch genau, wie ich Dich gefragt habe, ob Du mit auf den Weihnachtsmarkt kommst und bin immer noch froh, dass Du mitgekommen bist. Du bist eine tolle Frau und eine tolle Freundin und ich vermisse Dich! Dinchen, Du warst eine der ersten Freundinnen, die ich in Maastricht hatte und ich war tierisch stolz auf Dich, als Du mir jetzt hiermit ein paar Monate zuvorgekommen bist. Schön, dass es immer noch so zwischen uns ist. Claudia, Frauke, Kim, Anne Scho., Nina und Nina- ich danke Euch für alles! Ich werde unsere Zeit nie vergessen und hoffe, Euch alle bald wiederzusehen!

Meine andere Freundesheimat habe ich seitdem ich 16 war und ist für mich wie eine Familie. Wir sind zusammen groß geworden und jetzt werden wir sogar in Anzahl noch größer! Tamara, wo und wie wäre ich ohne Dich? Ich will es mir nicht vorstellen. Du wirst immer besonders für mich sein und Du wirst immer auf mich zählen können. Ich bin stolz darauf, wie wir es geschafft haben, so eng befreundet zu bleiben und ich freue mich wie verrückt über Mascha. Philipp, aus diesen Vorzeichen nicht nur Akzeptanz, sondern eine eigenständige Freundschaft wachsen zu lassen, finde ich großartig! Ich danke Dir dafür. Dass Du Dich allerdings so lange vor dem Comeback im Soccer-Cage gedrückt hast,

bis ich wegziehe, geht natürlich eigentlich nicht klar. Benjamin, Du bist der Beste (Kunde im Lokal) und seit der ersten Zigarette auf dem Dach des Düsseldorfer Flughafens nicht mehr aus meinem Leben wegzudenken. Eine Doktorarbeit reicht nicht aus, um zu besprechen, was wir miteinander erlebt haben- eigentlich so ziemlich alles, oder? Danke, mein Freund, und Hick-Fick-Cha-Cha-Cha! Jens, ich bin mit niemandem so auf einer Wellenlänge, wenn es um Humor und politisch-philosophische Debatten geht, wie mit Dir. Ich habe großen Respekt vor Deiner scharfen, ungeschönten Logik und Deinem Verstand. Seit den Mittwochabenden in der Müllhalde (oder war rechts Dein Zimmer?) ist so viel Zeit vergangen und soviel passiert und trotzdem können wir immer noch genau solche Abende verbringen. Herrlich! Thomas, auch uns vereinen eine solch lange Geschichte und so viele Erlebnisse! Die Unterschiedlichkeit dieser ist verblüffend- denke nur zurück an Basketball mit Herrn Hupp, unsere Autofahrten durch Blens an diesem einen Wochenende, unser gemeinsames Wohnen in Maastricht oder alles dazwischen, danach und überhaupt! Ich habe ja immer gesagt, dass Du der erste von uns sein würdest und so ist es auch gekommen. Es freut mich von Herzen, dass Ihr so glücklich seid und Joan so toll ist! Naasewicz, mit Stolz kann ich sagen, dass ich schon mit Dir auf der Bühne gestanden habe! Du warst mir immer ein verlässlicher und treuer Freund und ich danke Dir dafür! Nie habe ich mit so viel Spaß einen Kühlschranks enteist (benebelt und mit Hammer, Messer und Fön), oder Gloria in excelsis Deo gesungen, wie mit Dir! Es ist schön, dass Du Deinen Platz gefunden hast! Pats, wir sind zusammen nach Maastricht gegangen und haben die Opiumhöhle in der Wyckerbrugstraat zusammen bezogen. Man kann uns nicht vorwerfen, wir hätten dieser Wohnung nicht alles abverlangt! Wyck in flames!! Es freut mich, dass wir uns immer noch nahe stehen nach all den Veränderungen und dass Du immer noch kein Wortspiel auslässt, und sei es noch so schlecht. Maria, Petra, Sabine und Barbara- ich hätte mir Euch nicht besser aussuchen können. Schön, dass Ihr Euch der unehrvollen Aufgabe gewachsen seht, meine komischen Freunde glücklich zu machen.

Carsten, es freut mich, dass wir uns so angefreundet haben und unsere gemeinsamen Zeiten beim Pokern, Fußballspielen, Bierchen trinken und quatschen werden es mir schwer machen, Aachen zu verlassen (auch wenn wir das alles viel zu selten machen)! Danke auch dafür, dass Du mir offen gesagt hast, Du seiest ja dagegen ☺. Gut, dass bei Yassi und Dir noch eine NY-Reise aussteht!

Michael, schön, dass wir immer wieder an früher anknüpfen können, egal wie lange wir uns nicht sehen und was zwischendurch passiert. Du bist ein wahrer Freund.

Liebe Susanne, Katja, Gabriele, Maren, Tom und Rest der Gruppe 13- vielen Dank für dieses schöne Jahr, die Wertschätzung, die Einblicke, das Lachen und den Austausch. Es ist ein Jammer, dass ich nicht weiter bei Euch bleiben kann!

Chris und Georg, Ihr seid die besten Eltern, die ich mir vorstellen kann. Im Selbsterfahrungsteil meiner Therapieausbildung kamen den Kollegen die Tränen vor (natürlich wohlgesonnenem) Neid bei der Beschreibung meines Elternhauses und ich finde zu Recht! Es ist ein unvorstellbarer Luxus durchs Leben zu gehen mit dem Wissen, rückhaltlos, vollkommen und uneigennützig unterstützt und geliebt zu werden. Ich habe großen Respekt vor Eurer fachlichen Kompetenz, vor Eurem Welt- und Menschenbild, vor Eurer sozialen Intelligenz, vor Eurer Offenheit und Eurer unstillbaren Reiselust. Der Systemiker sagt „Respekt vor Menschen, kein Respekt vor Ideen“. Das habe ich durch Euch lange vor meiner systemischen Ausbildung verinnerlicht und das hat mir im Studium und bei meiner Arbeit sehr geholfen. Es war immer anders bei uns als bei anderen. Wie wertvoll und schön anders es war und ist, wird mit jedem Jahr älter werden klarer für mich. Ihr seid mein wichtigster Bezugspunkt, mein Rückhalt und meine Vorbilder in vielen Aspekten. Diesen fundamentalen Dank auszusprechen ist ihm schon fast nicht würdig. Wie alles, ist ja auch das erfolgreiche Erlangen eines Dokortitels das Resultat einer gelungenen Gene-Environment-Interaction. Da sowohl Gene, als auch Environment zu einem großen Teil, von Euch stammen, gehört damit dieser Dokortitel, zumindest hiermit in der Widmung, auch Euch. Nina, Du machst das gerade Geschilderte natürlich erst komplett. Du bist die beste Schwester der Welt und ich genieße jede Phase unserer geschwisterlichen Beziehung. Gut- die Kesselflickerphase der Kindheit können wir da mal lassen, aber danach warst Du mein wichtigster Orientierungspunkt im Erwachsenwerden. Ich wollte immer so leben wie Du, war stolz und ein bisschen neidisch auf Dich. Dann hat sich mit der Zeit eine Vertrautheit und Verbundenheit auf Augenhöhe entwickelt, die ich unbeschreiblich schön finde. Ich konnte und kann mit Dir immer über alles sprechen und weiß, dass Du mir immer mit Offenheit und Support begegnest. Das jüngste Beispiel ist Deine rückhaltlose Unterstützung der New-York-Pläne nachdem Du vor der Bekanntgabe der Stipendiumsvergabe unmissverständlich klar gemacht hattest, dass Du eigentlich nicht willst, dass ich so weit weg gehe. Mit Dir private und fachliche Gespräche zu führen, zu

lachen, zu weinen, zu reflektieren, zu reisen, zu tauchen, zu kochen, zu trinken, zu laufen und vieles mehr gehören zu den schönsten Dingen in meinem Leben. Jetzt freue ich mich auf die nächste Phase unserer Geschwisterbeziehung- Du als Mutter und ich als Onkel! Lieber Jens, danke, dass Du genau der richtige Mann für meine Schwester und genau der richtige Schwager für mich bist und genau der richtige Vater für meinen Neffen sein wirst. Die nächste Hoffnung für den FC ist im Anmarsch! Liebe Sine! Haaaaai!! Ich kann Dir gar nicht sagen, wie ich mich freue, dass wir diese Kiste hier in so kurzem Abstand voneinander geschaukelt haben! Jetzt habe ich Dir schon den Marathon und den Doktor nachgemacht, so wie es sich für einen kleinen Cousin gehört. What's next? Ich, für meinen Teil, hätte das hier jedenfalls ohne unsere regelmäßigen Ausgehmarathons in Köln, unsere Gespräche und unsere Ski- und anderen Urlaube nicht geschafft. Mit einer Frau zusammen aufgewachsen zu sein, die so witzig, intelligent und easy ist, ist eine große Ehre. Außerdem bist Du mein ungebiaster Blick auf den weiblichen Geist. Danke für Alles! Liebes Goldvittchen, Billhelm, Herr Scholl-Latour- Maastricht und vor Allem Aachen waren durch Dich unvergleichlich witziger und schöner. Ich schätze Deinen Humor, Deine politische Bildung und Meinung, Deine Gedanken, Deine blöden Fragen, Deine Antworten auf blöde Fragen, unser Singen, Deine Pokerkunst und Deine Übersicht auf dem echten und auf dem virtuellen Platz sehr. Vielen Dank für den Ausgleich und all den Spaß! Familie ist halt einfach Familie, was? Ali und Margot, danke für die unschlagbar witzige und laute Familienatmosphäre, für einen Schlafplatz nach durchzechten Nächten, für Familienfeste, Skiurlaube und eine gemeinsame Geschichte. Es ist schön, dass es Euch gibt! Liebe Uli und lieber Jonah, danke für Eure Zuneigung und Eure Unterstützung. Danke, dass ich mich als Patenonkel (oder gerne auch Godfather) versuchen darf. Es hat mich gerührt, dass Ihr jetzt dieselben Plätzchen backt, wie wir beide schon, Uli, als ich so alt war wie Jonah jetzt. Das ruft Erinnerungen hervor an Gondeln über dem Rhein, an Plätzchenteig roh essen, an Woolworthbesuche und Rubbellose auf der Domplatte, an Videogucken im Bett und an Schneekugeln. Ich hoffe, ich kann auch in Zukunft ein bisschen von dieser Liebe an Jonah weitergeben.

We can't all be heroes, because somebody has to sit on the curb and clap as they go by.

Will Rogers (1879-1935)

You are all heroes. Consider me clapping.

CURRICULUM VITAE

Niclas J. Kilian-Hütten was born in Düren (Germany) on October 3rd, 1981. In 2001 he completed his secondary school education at the “Gymnasium am Wirteltor” in Düren. He studied Psychology at Maastricht University from 2002 and obtained a Bachelor of Science, cum laude, in Psychology in 2005. In 2004/2005 he spent one semester at the University of California, Los Angeles (UCLA), where he conducted empirical work for his Bachelor Thesis titled “Hemispheric differences in the implicit concept of self”, under supervision of Professor Eran Zaidel. Starting in 2005, he followed the Research Master program in Cognitive Neuroscience at Maastricht University. In his Master Thesis, under supervision of Professor Elia Formisano, he used functional magnetic resonance imaging to investigate the phenomenon of cross-modal recalibration of auditory speech perception. He graduated, cum laude, in 2007. Funded by an NWO TopTalent grant, he started his Ph.D. program under supervision of Professor Elia Formisano and co-supervision of Professor Rainer Goebel at the Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University. Funded by an NWO Rubicon grant, he will start a postdoctoral fellowship at Columbia University, New York, under supervision of Professor Charles Schroeder in 2012.

