

**Theory of Mind and Emotion Recognition**  
**The Influence of Authenticity on the Perception of Emotional Prosody**

**Dissertation**

for the award of the degree

“Doctor of Philosophy”

Division of Mathematics and Natural Sciences  
of the Georg-August-Universität Göttingen

within the doctoral program Systems Neuroscience (GGNB)  
of the Georg-August University School of Science (GAUSS)

Submitted by

**Matthis Drolet**

from Uster, Schweiz

Göttingen 2013

Doctoral Thesis Committee:

**Prof. Dr. Julia Fischer (Advisor)**  
Cognitive Ethology Lab  
German Primate Center (DPZ)  
Kellnerweg 4, 37077 Göttingen

**Prof. Dr. Christiane Thiel**  
Biological Psychology  
Carl von Ossietzky University Oldenburg  
Ammerlaender Heer Str. 114-118, 26111 Oldenburg

**Prof. Dr. Jens Frahm**  
Max Planck Institute for Biophysical Chemistry  
Biomedical NMR Research  
Am Fassberg 11, 37077 Göttingen

Defense Board Members:

**Prof. Dr. Julia Fischer (First Referee)**  
Cognitive Ethology Lab  
German Primate Center (DPZ)  
Kellnerweg 4, 37077 Göttingen

**Prof. Dr. Christiane Thiel (Second Referee)**  
Biological Psychology  
Carl von Ossietzky University Oldenburg  
Ammerlaender Heer Str. 114-118, 26111 Oldenburg

**Prof. Dr. Jens Frahm**  
Max Planck Institute for Biophysical Chemistry  
Biomedical NMR Research  
Am Fassberg 11, 37077 Göttingen

**Dr. Peter Dechent**  
Göttingen University Medical School  
Dept. of Cognitive Neurology  
Robert-Koch-Str. 40, 37075 Göttingen

**Prof. Dr. Annekathrin Schacht**  
Experimental Psycholinguistics  
University of Göttingen  
Nikolausberger Weg 23, 37073 Göttingen

**Prof. Dr. med. Oliver Gruber**  
Department of Psychiatry and Psychotherapy  
University of Göttingen  
von-Siebold-Str. 5, 37075 Göttingen

Date of the Oral Defense: May 8, 2013

Herewith I declare that I have written this thesis independently and with no other aids or sources than quoted.

---

Matthis Drolet

Göttingen, 30.3.2013

# Table of Contents

<b>Summary</b> .....	<b>ii</b>
<b>Zusammenfassung</b> .....	<b>vi</b>
<b>1. General Introduction</b> .....	<b>1</b>
1.1 Social Cognition .....	1
1.2 Emotions .....	5
1.3 Emotional Expressions .....	7
1.4 Authenticity .....	10
1.5 Objectives .....	11
<b>2. Study I: Authenticity Affects the Recognition of Emotions in Speech:</b> .....	<b>13</b>
2.1 Abstract .....	14
2.2 Introduction .....	15
2.3 Methods .....	16
2.4 Results .....	21
2.5 Discussion .....	24
2.6 Concluding Remarks .....	28
2.7 Acknowledgments .....	28
2.8 Supplements .....	29
<b>3. Study II: Encoding conditions affect recognition of vocally expressed emotions across cultures</b> .....	<b>33</b>
3.1 Abstract .....	34
3.2 Introduction .....	35
3.3 Material and Methods .....	38
3.4 Results .....	41
3.5 Discussion .....	46
3.6 Conclusion .....	48
3.7 Acknowledgements .....	49

<b>4. Study III: Explicit authenticity and stimulus features interact to modulate BOLD response induced by emotional speech.....</b>	<b>51</b>
4.1 Abstract .....	52
4.2 Introduction.....	53
4.3 Methods .....	55
4.4 Results .....	59
4.5 Discussion .....	63
4.6 Concluding Remarks .....	67
<b>5. Study IV: Recognizing the authenticity of emotional expressions: F0 contour matters when you need to know .....</b>	<b>69</b>
5.1 Abstract .....	70
5.2 Introduction.....	71
5.3 Methods .....	73
5.4 Results .....	78
5.5 Discussion .....	81
<b>6. General Discussion.....</b>	<b>85</b>
6.1 Brain Activation .....	85
6.2 Behavior.....	89
6.3 Conclusions.....	92
6.3 Outlook .....	94
<b>7. References .....</b>	<b>97</b>
<b>8. List of Figures and Tables .....</b>	<b>118</b>
8.1 Figures .....	118
8.2 Tables .....	119
<b>9. Acknowledgments.....</b>	<b>120</b>
<b>10. Curriculum Vitae .....</b>	<b>122</b>



## Summary

Perceiving and understanding social stimuli, including facial expressions, body language and vocalizations, is essential to human interactions. The ability to correctly identify internal states of others through such stimuli, also called social cognition, is necessary for understanding and predicting the behaviors of others. The identification of internal states through stimulus information and previous experiences is known as a “Theory of Mind” (ToM). ToM, or mentalising, is the ability to attribute specific mental states (e.g. desires, beliefs, and intentions) to others. Emotional expressions are a particularly important category of stimuli in social interactions. In this project, I focus specifically on vocal expressions of emotion. The recognition of emotion in the voice relies on the nonverbal component of vocal expressions, also called intonation or prosody. However, the recognition of emotional prosody has been found to depend on a variety of other factors, such as context or the cultural background of the speaker or perceiver.

Here I investigate the recognition of emotional expression and authenticity to examine interactions between vocal expression and context. The difference between authentic and play-acted emotional expressions provides an ideal platform to examine these interactions. These categories differ in context due to the different intentional states of the individuals expressing the emotion. The authentic recordings used in this project came from radio interviews and were internally initiated, while the play-acted ones were produced by actors who were told the text, context, and emotion to be expressed and were, therefore, externally initiated. Because the recording categories were characterized by differences in the mental state of the speakers, I expected that the detection of authenticity and its effect on emotion recognition would depend on ToM processing.

One goal of this dissertation was to determine whether stimulus authenticity influences emotion recognition and modulates activity in the neuronal substrates involved in that recognition. Experimentally, participants judged either the authenticity or emotional content (anger, fear, joy, or sadness) of the recordings while being scanned by way of functional magnetic resonance imaging (fMRI). Authenticity affects emotion recognition, such that anger was recognized best when play-acted and sadness when authentic, and the ToM network in the brain was recruited for explicit judgments of authenticity. In addition, authentic recordings were found to enhance activity in the medial prefrontal cortex. This demonstrates the necessity of ToM processes in the recognition of authenticity, the concurrent influence on the perception of emotional expressions, and the recruitment of additional cognitive resources for perception of authentic expressions.

To clarify the nature of the effect of authenticity, I additionally compared the findings across three countries to examine the influence of universals and cultural-specificities. Participants from Germany, Romania, and Indonesia were tested on a similar behavior-only recognition experiment. Participants in all three countries were equally poor at distinguishing the authenticity of expressions. Nevertheless, the effect of authenticity on emotion recognition was nearly identical across all three cultural groups. Although German subjects revealed a slight advantage in recognizing emotions overall, differential cultural effects became particularly apparent through differential biases in emotion attribution. German participants exhibited a bias towards choosing anger for all stimuli, while Romanian and Indonesian participants were biased differently for authentic expressions, preferentially categorizing these as sadness. This suggests that complex interactions between human universals and cultural specificities influence emotion and authenticity recognition.

Further examination of the effect of authenticity was focused on understanding the proximate reasons for both the behavioral and brain effects. To do this, I tested the direct stimulus-induced influence of authenticity on brain activation and whether top-down knowledge about authenticity would interfere with the stimulus effect. While participants were tasked only with emotion recognition, two-thirds of stimulus playbacks were cued as either authentic or play-acted, and either congruent or incongruent to the stimulus. During uncued trials, play-acted stimuli spontaneously up-regulate activity in the auditory cortex and areas associated with processing of emotional prosody, but the stimulus-driven effect of authenticity in the medial prefrontal cortex was not replicated. While cueing did not affect emotion recognition, cue and stimulus interaction did up-regulate activation in the posterior superior temporal sulcus and the anterior cingulate cortex. When a play-acted stimulus followed a cue indicating an authentic stimulus, activation in the temporoparietal junction also increased, indicating additional perspective-taking in ToM processes.

What remained unclear, however, is how authenticity is perceived acoustically. Examining the effect of fundamental frequency contour, which varies more in play-acted expressions than authentic ones, I wished to determine whether contour affects not only authenticity but also emotion recognition and correlated brain activation. Stimuli with lower contour were preferentially categorized as either sad or authentic, while activation in the primary auditory cortex was up-regulated task-independently by increased contour. Contour and task additionally interacted in a network including medial prefrontal cortex, with increased activity related to low-contour stimuli during explicit perception of authenticity versus an increase for high-contour stimuli during explicit perception of emotion. Contour-induced effects appear to be purely stimulus-driven in early auditory and intonation perception, while being strongly task-dependent in regions involved in higher cognition.



The behavioral and functional results show that authenticity of emotional prosody is an important property that influences human responses to such stimuli, with implications for studies using play-acted emotions. The effect of stimulus authenticity on brain activity suggests that perceiving intention influences the recognition of vocally expressed emotions. This effect can be simultaneously task-dependent, as seen in parts of the ToM network, and stimulus-driven, as seen in early auditory and intonation processing. In addition, contextual information can further modulate the brain response to differences in the authenticity of emotionally expressive speech. In combination with the cross-cultural data this supports the view that the influence of speaker intention on emotion recognition relies on complex interactions of universals and cultural-specificities in emotion expression.



## Zusammenfassung

In zwischenmenschlichen Interaktionen sind die Wahrnehmung und das Verstehen von sozialen Äußerungen, zu denen Gesichtsausdrücke, Körpersprache und Lautgebung gehören, unentbehrlich. Die Fähigkeit den mentalen Zustand anderer Personen identifizieren zu können, auch als soziale Kognition bekannt, ist notwendig um das Verhalten anderer verstehen und interpretieren zu können. Diese Identifikation des mentalen Zustandes durch die Kombination von Reizinformation und eigener Erfahrung ist als "Theory of Mind" (ToM) bekannt. ToM ist die Fähigkeit anderen Personen bestimmte mentale Zustände (Wünsche, Glauben, Absichten und Emotionen) zuzuschreiben. Emotionale Ausdrücke stellen dabei besonders wichtige Reize in sozialen Interaktionen dar, wobei sich diese Arbeit spezifisch auf vokale emotionale Ausdrücke beschränkt. Das Erkennen von Emotionen in der Stimme geschieht durch nicht-verbale Komponenten, die bei der Lautproduktion entstehen und u.a. als Intonation oder Prosodie bekannt sind. Allerdings wird die Erkennung von emotionaler Prosodie stark von anderen Faktoren beeinflusst, so wie Kontext oder Kultur des Sprechers und Zuhörers.

Um die Interaktion zwischen emotionalem Ausdruck und Kontext weiter aufzuklären, habe ich die Erkennung von Ausdrücken in der Stimme in Zusammenhang mit der Authentizität untersucht. Der Unterschied zwischen authentischen und gespielten emotionalen Ausdrücken bietet eine ideale Grundlage für die Erforschung dieser Interaktion. Beide Stimulus-Arten unterscheiden sich im Kontext der Produktion durch die unterschiedliche Intention des jeweiligen Sprechers. Die in diesem Projekt verwendeten authentischen Tonaufnahmen stammen aus Radiointerviews und waren intern initiiert, während die gespielten Tonaufnahmen von Schauspielern produziert wurden, die den Text, den Kontext der Aufnahme und die jeweilige Emotion zugewiesen bekamen, und daher extern initiiert waren. Meine Erwartung war, dass die Erkennung der Authentizität durch eine ToM Verarbeitung geschehen würde, da die unterschiedlichen Tonaufnahmen durch einen Unterschied im mentalen Zustand des Sprechers charakterisiert sind.

Ich wollte feststellen, ob die Erkennung der emotionalen Ausdrücke und die damit einhergehende Gehirnaktivierung von der Authentizität des Stimulus Materials beeinflusst werden. Experimentell sollten Individuen, die in einem Magnetresonanztomographen lagen, entweder die Authentizität oder die Emotion (Wut, Angst, Freude, Trauer) der Tonaufnahmen identifizieren. Authentizität hatte einen deutlichen Einfluss auf die Erkennung von Emotionen, wobei Wut bei gespielten Aufnahmen besser erkannt wurde, Trauer hingegen bei authentischen. Im Gehirn wurde dabei das ToM Netzwerk für die explizite Bewertung der Authentizität rekrutiert. Zusätzlich wiesen die Versuchspersonen bei authentischen Aufnahmen, anders als bei gespielten, eine erhöhte Aktivierung im medialen

prefrontalen Cortex auf. Dies weist auf die Notwendigkeit von ToM bei der Erkennung von Authentizität, den gleichzeitigen Einfluss auf die Wahrnehmung von emotionalen Ausdrücken, und die erhöhte Rekrutierung kognitiver Ressourcen für authentische Ausdrücke hin.

Um die Eigenschaften des Effekts von Authentizität zu verdeutlichen, habe ich diesen Befund auf universelle und kultur-spezifische Einflüsse in drei unterschiedlichen Ländern geprüft. Individuen aus Deutschland, Rumänien und Indonesien wurden in einem ähnlichen Experiment, aber nur auf Verhalten, getestet. Obwohl Teilnehmer aus den unterschiedlichen Ländern die Authentizität der Aufnahmen nur schwer erkennen konnten, war der Authentizitätseffekt auf die Emotionserkennung für alle Kulturen sehr ähnlich. Die wichtigsten Unterschiede kamen durch Neigungen für bestimmte Emotionen zustande. Die deutschen Versuchspersonen waren bei den gespielten wie auch bei den authentischen Aufnahmen eher dazu geneigt Wut zu wählen, während Versuchspersonen aus Rumänien und Indonesien sich bei den authentischen Emotionen, im Gegenteil zu den Deutschen, vermehrt für Trauer entschieden. Dies weist auf eine komplexe Interaktion universeller und kultur-spezifischer Effekte bei der Authentizitäts- und Emotionserkennung hin.

Für weitere Aufklärung des Authentizitätseffektes konzentrierte ich mich schließlich auf die proximalen Ursachen des Verhaltens und der Verarbeitung im Gehirn. Zunächst habe ich untersucht, ob der Authentizitätseffekt durch vorheriges Wissen beeinflusst werden kann, indem den Teilnehmern in Zwei-Dritteln der Versuche durch Hinweisung mitgeteilt wurde, ob die vorgespielten Aufnahmen authentisch oder gespielt waren. Diese Aussage war entweder richtig (kongruent) oder falsch (nicht-kongruent). Ganz ohne Hinweise verursachten gespielte Reize eine erhöhte Aktivierung im primären auditorischen Cortex und in Arealen, die wichtig für die Verarbeitung von Prosodie sind. Der Effekt im medialen prefrontalen Cortex war hingegen nicht mehr zu erkennen. Obwohl die Erkennung der emotionalen Ausdrücke durch die gegebenen Hinweise („gespielt“ oder „echt“) nicht beeinflusst wurden, hatten die Hinweise einen klaren Einfluss auf die Gehirnaktivierung. Die allgemeine Interaktion von Hinweis und Authentizität des Reizes verursachten eine erhöhte Aktivierung im superioren temporalen Sulcus und im anterioren Cingulum. Wenn auf den Hinweis eines authentischen Ausdruckes eine gespielte Aufnahme folgte, erhöhte sich die Aktivierung im temporoparietalen Cortex, was auf eine Komponente der ToM Fähigkeit, der gesteigerten Perspektivenübernahme, hindeutet.

Unklar blieb, wie Authentizität akustisch wahrgenommen wird. Die Kontur der Grundfrequenz weist eine größere Variabilität bei gespielten als bei authentischen Aufnahmen auf. Durch eine Analyse des Effektes der Kontur-Variabilität wollte ich feststellen, ob dieser Faktor einen Einfluss auf die Erkennung der emotionalen Ausdrücke und auf die Gehirnaktivierung hat. Aufnahmen mit niedriger Kontur-Variabilität wurden bevorzugt als Trauer oder authentisch kategorisiert, während die

Aktivierung im Gehirn durch eine höhere Kontur-Variabilität im primären auditorischen Cortex aufgabenunabhängig erhöht wurde. Hinzu kam, dass die Kontur und die Aufgabe der Versuchsperson (Emotionserkennung versus Authentizitätserkennung) im medialen prefrontalen Cortex interagierten. Eine erhöhte Aktivierung fand bei der Emotionserkennung statt, wenn die Kontur-Variabilität niedrig war, während bei der Authentizitätserkennung eine erhöhte Aktivierung festgestellt wurde, wenn die Variabilität der Kontur hoch war. Der Authentizitätseffekt im Verhalten und in der Gehirnaktivierung scheint also durch Kontur-Variabilität beeinflusst zu sein, ist aber stark aufgabenbedingt in Arealen, die für soziale Kognition wichtig sind.

Die Verhaltens- und Aktivierungsdaten zeigen, dass Authentizität ein wichtiger Faktor bei der Emotionserkennung ist, das auch Auswirkungen auf Studien hat, die gespielte emotionale Ausdrücke benutzen. Die Effekte zur Gehirnaktivierung des ToM Netzwerkes suggerieren, dass die Erkennung von Intention einen Einfluss auf die Wertung von emotionalen Ausdrücken in der Stimme hat. Der Einfluss der Authentizität des Stimulus kann gleichzeitig aufgabenabhängig, beispielsweise in Teilen des ToM Netzwerkes, und Stimulus-angetrieben, im primären auditorischen Cortex, sein. Weiterhin können Kontextinformationen die Gehirnaktivierung, die durch Authentizität moduliert wird, zusätzlich beeinflussen. Wenn diese Ergebnisse gemeinsam mit den Unterschieden zwischen den getesteten Kulturen betrachtet werden, wird deutlich, dass der Einfluss der Intention des Sprechers auf die Erkennung von Emotionen auf einer komplexen Interaktion von universellen und kultur-spezifischen Effekten beruht.



## 1. General Introduction

The accurate and timely recognition of social stimuli is of major importance to daily human interactions. The ability to correctly identify internal states in others can be particularly useful for predicting and understanding behavior. One such internal state that greatly affects behavior is emotion, and its recognition relies strongly on the perception of emotional expressions. Importantly, emotion expression and recognition can vary depending on cultural background, context, and intention. In the following, I first introduce the concepts involved in social cognition, including Theory of Mind. I then delve more deeply into theories of emotion and neuroscience of emotion expression, along with the impact of culture. Finally, I introduce the overlap of Theory of Mind and emotion that is central to this project, the difference between authentic and play-acted vocal expressions of emotion.

### 1.1 Social Cognition

#### 1.1.1 Social-Cognitive Theories

Arguably, secondary only to its somatosensory and motoric roles, the most important function of the mammalian brain is the regulation of social interactions (Dunbar, 1998). Cognitive processing of social interactions is a major topic in psychology, behavioral biology, philosophy, and cognitive neuroscience, focusing on the understanding and prediction of the actions of others (Haggard, 2008; Tomasello *et al.*, 2005), and examining the brain structures responsible for these abilities (Frith and Frith, 2007). Central to theories about the function of the brain in social cognition is how others' mental states are represented. Historically, two models have been proposed and tested to explain these processes: theory-theory and simulation-theory (Emery, 2004). The goal of both theories is to explain how humans perform the higher-level social cognition of understanding and predicting the behaviors of individuals around them, also commonly called mentalising or Theory of Mind. Theory-theory posits independent theories or models of others' mental states, such that a mental-state model can be triggered by a particular stimulus and is directly associated with particular behavioral patterns. This theory assumes social cognition to function as the interaction of mental state concepts (belief, desire, etc.) and, in cognitive neuroscience, that particular brain regions exclusively represent each of these concepts (Apperly, 2008). Simulation-theory, on the other hand, assumes a more domain-general organization for self and other perception, relying on the integration of individual low-level processes that, combined, can either affect our own or represent others' mental states (Apperly, 2008). Under simulation-theory the brain produces our own mental state and simulates

those of others in overlapping networks of regions specialized in more low-level processes relevant to overall cognition and recruited for social cognition as the need arises (Mitchell, 2005).

### 1.1.2 Theory of Mind

Theory of Mind (ToM; Premack and Woodruff, 1978) initially appears to be closely associated with theory-theory. In fact, it is defined simply as “the ability to explain and predict the behavior of ourselves and others by attributing to them independent mental states, such as beliefs, desires, emotions or intentions” (Frith and Frith, 2003). Thereby, it places more importance on the ultimate use of ToM rather than its proximate composition<sup>1</sup>. As with Premack and Woodruff (1978), much of the early research on ToM, as defined here, occurred in the field of animal behavior and presented a framework for the comparative study of various mental-state attributions. Various concepts related to ToM, such as metacognition (Smith *et al.*, 2004) and mirror self-recognition (Reiss and Marino, 2001), are present in other species as well. The concept of ToM has produced a field of research of its own and, while it can be incorporated in either of the two major theories above, it provides a framework for an objective examination of social cognition open to either interpretation.

In humans, ToM has been examined behaviorally quite extensively through child development research. With the now famous Sally-Anne task (Baron-Cohen *et al.*, 1985; Wimmer and Perner, 1983) research first showed that it develops by the age of 4 (Call and Tomasello, 1998). More recent studies found that lower-level behavioral and perceptual abilities related to ToM are present even earlier (Rakoczy, 2009; Rakoczy *et al.*, 2004). This early development shows how important this ability is to our species. Some have even hypothesized that it may be at the root of other abilities such as language (Ferstl and Von Cramon, 2002) or culture (Hollan, 2000). Further research focusing on the neuroscience of ToM in adults showed that particular brain regions were preferentially recruited during tasks requiring ToM abilities. Currently, the core regions agreed-upon as being central to ToM are medial prefrontal cortex (mPFC), temporoparietal junction (TPJ) / posterior superior temporal sulcus (pSTS), and the posterior cingulate cortex (PCC) / precuneus (Frith and Frith, 2003; Gallagher and Frith, 2003; Saxe, 2006). While activation in other regions, such as the temporal poles and the orbitofrontal cortex, is known to correlate with specific tasks requiring mentalising, it remains unclear whether inconsistent results are due to functional differentiation (Olson *et al.*, 2007; Sabbagh, 2004) or due to artifacts related to methodological constraints (Weiskopf *et al.*, 2006, 2007).

---

<sup>1</sup> **Proximate** explanations address the immediate cause of an event or behavior, seeking to identify “how” something occurs. **Ultimate** explanations address the larger picture that leads to proximate causes, seeking to identify “why” something occurs.



The regions comprising the ToM network have specific functions unique to each. The mPFC is involved in various socially relevant situations (Krueger *et al.*, 2009), but also in more domain-specific tasks such as action-outcome prediction (Alexander and Brown, 2011) in coordination with the cerebellum (Stoodley and Schmahmann, 2009), and the regulation of emotion (Kober *et al.*, 2008). However, it has been suggested that this may be due to some degree of mentalising in each of these tasks (Van Overwalle, 2011). TPJ/pSTS is involved in directing attention to salient and task-relevant stimuli (Corbetta and Shulman, 2002), as well as the perception of biological motion (Puce and Perrett, 2003; Wheaton *et al.*, 2004). The posterior cingulate cortex is a hub for access to autobiographical memory. In fact, tasks requiring access to autobiographical memory require a network composed of PCC/precuneus and mPFC, indicating that such information is necessary in ToM tasks as well (Spreng *et al.*, 2009). Finally, this wide network is also active in what is called the default mode. This activation is seen in imaging studies examining activity in the brain when not focusing on a particular task. It is considered by some to be the pattern of activity during conscious but undirected thought (Buckner *et al.*, 2008), indicating that access to autobiographical memory and/or thinking about social situations may be one of the most prevalent processes in brain (Schilbach *et al.*, 2008) or, at the very least, that undirected thought and mentalising require similar processes, including long-term memory retrieval (Spreng *et al.*, 2009).

### *1.1.3 Top-down versus Bottom-up Processing*

Research has provided a strong basis for understanding the various facets of social cognition. Until the end of the twentieth century much of social cognition research was focused on the influence of individual stimuli or individual stimulus components. However, social cognition can require the processing of various stimuli in parallel, as well as the integration of previous knowledge (Mesulam, 1998). Such integration involves both bottom-up and top-down processing. Bottom-up processing is influenced only by the stimulus, such that only sensory information about the stimulus is required. In the visual domain, for example, expressions of fear can carry a clear message, the response to which is an increase in attention towards the source, in part due to biologically coded pathways (Adolphs *et al.*, 1995). Top-down processing, on the other hand, incorporates other factors not directly originating from the stimulus source. This is possible either through knowledge available to the perceiver from previous experiences (Rogan *et al.*, 1997) or information about the context in which the stimulus was produced (Polyn and Kahana, 2008; Watanabe and Sakagami, 2007).

Due to increased awareness of top-down influences and the availability of more sensitive technologies, there is now an increase in research focused on these and, in particular, their effects on bottom-up stimulus perception. For example, the influence of context on emotion perception has been found to be significant (Albert *et al.*, 2010; Barrett and Kensinger, 2010; Brück *et al.*, 2011;

Chung and Barch, 2011; Gruber and McDonald, 2012; Jaeger and Rugg, 2012; Kappas and Poliakova, 2007). These data required a new framework to develop hypotheses relevant to the study of such interactions. Teufel and colleagues (2010) developed one such framework as of a model involving both bottom-up stimulus processing and top-down cognitive control (Fig. 1.1). This model integrates the bottom-up perception of stimuli with the top-down influence of various information sources such as context, prior knowledge, and ToM. Examining bottom-up and top-down cognition, and its interaction with emotion in particular, will enable us to better understand how these interact to produce the very complex behaviors seen in humans. The top-down influence of emotion on cognitive processes has been examined extensively (Kompus *et al.*, 2009; Shafer *et al.*, 2012). However, the top-down effects of cognitive processes, such as ToM, on emotion recognition have not been examined in detail. A unique study that implicated complex interactions between emotive and cognitive processes by Blair (2007) did find that goal-directed tasks can decrease emotional stimulus-induced activation. They also found that connectivity between mPFC and amygdala was decreased during task performance in the presence of additional emotional distractors, which provided some of the first evidence for complex interactions between cognitive task performance and emotion perception. Here, regarding this dichotomy between cognition and emotion, I specifically consider cognition to include processes other than emotive ones, with a particular focus on ToM. Accordingly, the focus of this project was to examine the top-down influence of mentalising on emotion perception.

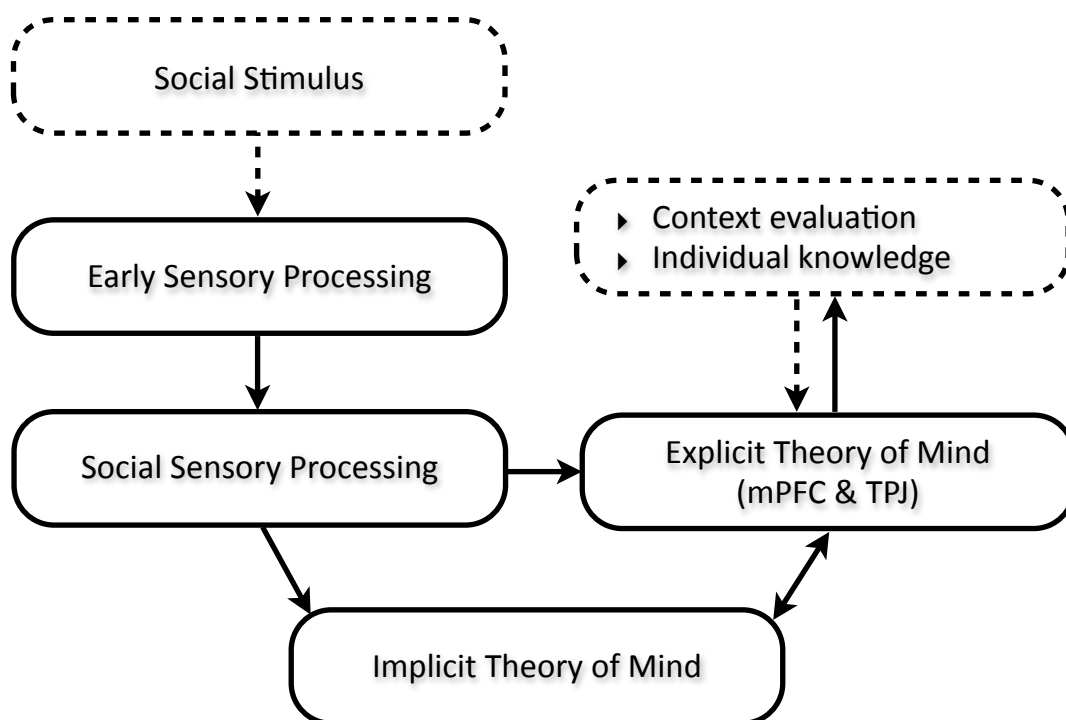


Fig. 1.1 Model of bottom-up and top-down processes in social cognition. Adapted from Teufel *et al.* (2010)

## 1.2 Emotions

### 1.2.1 Theories of Emotion

In “The Expression of the Emotions in Man and Animals” Darwin (1872) commented on the similarities in behavior between species and the apparent gradual adaptation in basic behavioral and emotional patterns. While this laid the groundwork for modern evolutionary theories of emotion, it did not seek to examine proximate causes of emotion. Instead, Darwin focused on their ultimate relevance regarding environment and conspecifics. However, he did not specifically address the proximate source of emotion.

One of the first theories to address this was the Somatic Theory of emotion, developed in parallel by William James (1884) and Carl Lange (Cannon, 1927). They suggested that the physiological responses we associate with emotions are the emotions themselves, while the conscious experience and outward expressions of emotion are secondary and dependent on the physiological:

*“...we feel sorry because we cry, angry because we strike, afraid because we tremble, and [it is] not that we cry, strike, or tremble, because we are sorry, angry, or fearful, as the case may be.” (James, 1890, p. 450)*

These theories were novel and largely counterintuitive to commonly held beliefs and led to techniques to examine emotions physiologically. However, the James-Lange theory did not make clear distinctions between internal and external influences or the conscious versus unconscious experience. The lack of a clear separation and definition (James, 1890) led to many of the topics of contention present today (Scherer, 2005). While later developments, such as the Cannon-Bard Theory (Cannon, 1931), clarified this distinction between physiology and conscious experience, physiology was still considered central to all emotion. Early neuropsychological research appeared to confirm these findings, with regions apparently central to emotions, including the parahippocampal gyrus, the cingulate gyrus, and the amygdala, commonly called the limbic system (LeDoux, 2000). The amygdala, in particular, is involved in both facial emotion expression and perception (Adolphs *et al.*, 1995; Morris *et al.*, 1996; Phan *et al.*, 2002) and acts as a major hub for behavior related to fight-or-flight responses (Seymour and Dolan, 2008), while the entire cingulate cortex is consistently active for the internal perception of pain and emotion in general (Vogt, 2005).

While these regions are involved in emotion and emotion perception, they are also active for a variety of cognitive processes independent of emotion, which indicated that the earlier findings were not as clear-cut as assumed. The somatic view was also challenged in part due to findings that physiological properties could not explain the full diversity of emotions (Scherer, 2005). In addition,

the hippocampus and parahippocampal gyrus, for example, are the hub for autobiographical memory (Bliss and Collingridge, 1993), which interacts with emotional processes (Gruber and McDonald, 2012), but does not necessarily produce or control them. Also, the cingulate gyrus becomes active in situations where emotion and cognition interact, in particular when task performance requires some form of attention control related to emotion perception (Bush *et al.*, 2000), but also tasks independent of emotion (Braver *et al.*, 2001). Proponents of cognitive theories of emotion emphasize this apparent overlap of cognitive processes in many areas of the brain. They argue that the cognitive perception and classification of both internal and external stimuli is needed before either the physiological response or the feeling experience, so that all facets of emotion are dependent primarily on unconscious brain processes rather than bodily responses (Frijda, 1986). Richard Lazarus (1991) and Ira Roseman (1984) argued that cognitive appraisal leads to internal physiological changes, conscious experience (feelings) and external actions. Cognitive theories now focus largely on the various types of such appraisals that occur in the brain and how they lead to specific emotions.

Overall, research has not pointed in either the cognitive or physiological direction exclusively. Cognitive appraisal can have a clear and measurable impact on the experience of emotion (Ross, 2011). However, the lack of a physiological component can greatly affect the feeling experience as well (Damasio, 2008). Further developments in emotion theories, such as the Two-factor model (Schachter and Singer, 1962) or the Component Process Model (CPM), try to integrate the various findings, incorporating cognitive and physiological elements to varying degrees. Of particular interest along this line is the CPM, which is based in part on the large body of research originating from the group of Klaus Scherer, an important proponent of cognitive-appraisal theories of emotion. This model incorporates the various elements discussed above, from physiology to cognitive appraisal, such that the lack of a single component can affect the whole (Grandjean *et al.*, 2008).

### *1.2.2 Defining Emotions*

In part due to the discussion about the nature of emotions, the question arose of how emotions should be defined in the first place. While various models have been suggested, they can be largely subsumed under two categories: basic models and dimensional models. The basic model of emotion assumes strict categories of emotion that are coded to specific stimulus-response patterns (Ekman, 1992). As such, emotions are represented as discrete, non-overlapping categories that elicit, but are separate from, phenomena such as physiological responses and cognitive appraisal. Panksepp (1998) argues that the discovery of specific brain circuits associated with emotional responses supports this view, while others have proposed that discrete emotions may be “constructed” from more

elementary biological and psychological components, as in the constructionist model (Gendron and Barrett, 2009; Schachter and Singer, 1962).

The idea of components has been applied in a different way in dimensional models. These place emotions along gradients of other psychological and physiological processes. While early theorists tried to place all emotions along a single gradient, contemporary proponents make use of multiple dimensions (Scherer, 2000). Two of the most well-known dimensions are valence and arousal, while many include a third, such as dominance or control (Davitz, 1964; Goudbeek and Scherer, 2010; Pereira, 2000). Recent neuropsychological research has also provided evidence for dimensional models, with particular brain regions found to be differentially activated by these dimensions. Dolcos and colleagues (2004), for example, presented evidence for differential mPFC activation based on both arousal and valence, with dorso- and ventromedial activity sensitive to arousal and valence, respectively. James Russell, a well-known proponent of dimensional models (Russell and Barrett, 1999; Russell, 1980), has been involved in research that found a much more widespread network of activation, including the thalamus, caudate, parahippocampal gyrus and amygdala in processing of valence and arousal (Colibazzi *et al.*, 2010).

As with the somatic theories, many contemporary researchers have attempted to combine elements of these models to explain the similarities and differences in the results to date and form more integrative ones. Within this project I focused on specific categories of emotions because the stimuli were acquired based on the use of such categorical labels by the individuals expressing emotions. The following discussion, however, is based on the notion that cognition and physiology represent components of a whole, with the assumption that discrete categories are more language-based than biological (Davidoff, 2001; Roberson *et al.*, 2010).

## **1.3 Emotional Expressions**

### *1.3.1 Universals versus Culture*

The basic model of emotion initially gained support from research into emotion expressions and their ultimate use in different cultures. Not only do individuals use the same expressions, such as smiling, from a very early age (Freedman, 1964; Söderling, 1959) but these expressions extend well beyond cultural boundaries. Eibl-Eibesfeldt (1968) was one of the first to systematically compare behavior, including facial emotional expressions, between various cultures around the world. The expressions were largely similar and occurred in similar contexts, which appeared to refute the notion that expression categorization could be solely culture-specific or random. Ekman (1988) further examined the physiology of those facial expressions to determine how they are formed. He assembled an

extensive database of the muscles involved in facial expressions and compared the perception of those expressions across numerous cultural groups. He found that there was high agreement between individuals of even very disparate groups (Ekman *et al.*, 1969). Later research examining emotional expressions of various types found similar patterns (Elfenbein and Ambady, 2002). Two of the strongest proponents, Ekman (1992) and Izard (2009), argue that the automatic and distinctive nature of emotional expressions, biological foundations for these, and invariance between cultures, all point to universal emotions that are biologically coded.

While there are some strong similarities in expressions, there are also differences, and they can elicit very different responses and reactions depending on the cultural group. Thus, emotion categories as such may be socially constructed (Averill, 1980). Since the evaluation or appraisal of emotion can be included as part of the emotional process (Scherer, 1984; Smith and Ellsworth, 1985) these effects needed to be addressed. The socio-cultural model of emotion expression, which posits culturally constructed and learned emotions, was bolstered by findings on the differences in emotion recognition between different groups (Bryant and Barrett, 2008; Jack *et al.*, 2012), which led to the question of whether a hard-coded emotion model (physiological or cognitive) was appropriate (Markus and Kitayama, 1991). Even the stringent basic emotion theories were updated to incorporate the influence of cognition, as with Izard's emotion schemas (Izard, 2009) that he used to explain this interplay. Lisa Feldman Barrett (2011) recently published a review of research over the past century in which she strongly accentuates the symbolic value of emotion expressions as opposed to the more direct value as a signal of internal state. Thus, it seems that emotional expressions, while being grounded in evolutionary behavioral and brain structures, are strongly modified by learning. Thus, it may be more fruitful for further research to examine how and to what extent biological universals interact with learning and culture.

### *1.3.2 Vocal emotion expression*

While the human voice is an important modality in the transmission of emotional information (Banse and Scherer, 1996; Hammerschmidt and Jürgens, 2007; Juslin and Laukka, 2003), the majority of research regarding emotion expressions has focused on facial expressions. Over the past half century, however, findings on vocal emotion expression have influenced thinking on both emotion expression and emotion overall. Vocal emotion expressions can be categorized based on their quality, intensity, and valence, using such measures as mean, range and variability of pitch, speech rate, and intensity (Banse and Scherer, 1996). Initial research indicated that these prosodic features and their recognition were largely universal (Frick, 1985). As with facial expressions, however, group-differences were found as research continued (Scherer *et al.*, 2001). Nevertheless, as with discussions on emotions themselves, the question of whether vocal expressions of emotion are more

physiological or cognitive, universal or learned, remains quite active. Bryant and Barrett (Bryant and Barrett, 2008) found very similar responses to vocal emotion expressions between American English and hunter-gatherer cultures in the Amazon. However, an extensive meta-analysis looking at facial and vocal expressions by Elfenbein and Ambady (2002) found that individuals are better able to distinguish emotional expressions produced by someone of their own group, now referred to as an in-group advantage. This accumulated data indicates a certain level of innate expression and recognition of emotion that can be modified by learning and culture.

### *1.3.3 Emotion Recognition Neuroscience*

As with emotion research, the development of advanced neuroscientific methods has had a dramatic impact on vocal emotion research. The main source of knowledge on the production of emotional prosody comes from neurology and lesion studies. This data led to the first discussion of a right-lateralized processing of emotional prosody in the human brain (Tucker *et al.*, 1977) as compared to the widely discussed left-hemisphere specialization for verbal language (Binder *et al.*, 1997). Major proponents of this view include Ross and colleagues who performed one of the first direct manipulations of right-hemisphere processing by injecting sodium amytal to temporarily inhibit the right hemisphere in epileptic patients (Ross *et al.*, 1988), which caused speech to be more flat in intonation, findings that have been replicated since (George *et al.*, 1996; Shapiro and Danly, 1985). However, a large body of research has found that right-lateralized brain damage, while appearing to restrict production of fundamental frequency (pitch) range through inferior frontal lesions and prosody perception through temporal lesions, does not necessarily lead to major prosodic deficits or aprosodia (Baum and Pell, 1999).

While the discussion of lateralization is ongoing, the use of functional magnetic resonance imaging (fMRI) has focused research on the particular regions of the brain involved in perceiving emotional prosody. While there was no specific evidence for lateralization of prosody perception, Adolphs and colleagues found that, in contrast to facial expressions, neither the amygdala (1999) nor the temporal lobes (2001) are necessary for recognition of emotional prosody. The involvement of the superior and middle temporal regions, however, is quite extensive (Wiethoff *et al.*, 2008). This activation extends further into the temporal cortex than just the primary auditory cortex, also known as the transverse temporal gyrus (TTG), which is the region where cortical auditory processing begins (Binder *et al.*, 1994; Celesia, 1976). Further, the posterior superior temporal sulcus differentiates emotional from verbal content (Wildgruber *et al.*, 2006). The involvement of the superior temporal gyrus and sulcus in the processing of speech sounds overall has been solidified by a meta-analysis by Vigneau and colleagues (Vigneau, 2006), who found that processing phonemes recruits a widespread region extending well past the primary and secondary auditory cortices.

Buchanan (2000) initially observed that the inferior frontal cortex was activated during explicit recognition of emotional prosody (as opposed to simply listening to emotional prosody) and Adolphs (2002) later confirmed this finding with an extensive lesion-patient study. This led to a model in which the superior temporal and inferior frontal cortices are involved in basic acoustic and task-dependent cognitive processing, respectively (Schirmer and Kotz, 2006). While basic acoustic cues are processed in superior temporal gyrus, explicit evaluation occurs in the inferior frontal gyrus and orbitofrontal cortex. The orbitofrontal cortex is known to be important for various tasks involving emotion evaluation. It plays a role in reward and punishment-related behavior (Rolls, 1999) thereby influencing decision-making processes. More generally, it may be central to making decisions based on the emotions associated with available choices (Bechara *et al.*, 2000). Finally, the anterior cingulate cortex is activated for explicit perception of emotional prosody as well (Bach *et al.*, 2008), likely due to its involvement in tasks requiring the integration of emotion and cognition (Allman *et al.*, 2001).

### **1.4 Authenticity**

Most research into emotional expressions makes use of acted expressions (actors instructed to portray an emotion by a third party). In doing so the focus of research has been to either examine prototypical behavior or assume that expressions are largely identical irrespective of the source, both for production by the sender and perception by the receiver (Belin *et al.*, 2008; Ethofer, 2006; Ethofer *et al.*, 2006; Hietanen *et al.*, 1998; Laukka, 2005). Research using acted emotions is important to determine the nature of prototypical categories of emotion expression, perception, and correlated brain activation (Bänziger and Scherer, 2007). It is precisely the control afforded by acting that has made such research possible (Scherer, 2013), while the perceived difficulty of acquiring and applying emotional expressions from real-life non-instructed social situations has slowed expansion in this direction (but see Laukka *et al.*, 2007; Scherer, 2013). However, findings on feigning emotion expression (Wilting *et al.*, 2006) indicate that such acted behavior may differ significantly from more naturally occurring ones, and the question of the relevance of findings from acted behavior to actual social contexts is just now beginning to be addressed (Audibert *et al.*, 2008a, 2008b).

Two areas that come closest to examining effects of feigning emotion are deception and spontaneous emotion expressions. A significant amount of ongoing research examining the various processes surrounding deception, such as lying or pretense, has found explicit behavioral correlates of both deceiving and detecting deception (Adolphs, 2009; Ekman *et al.*, 1990; Frick, 1985; Grezes *et al.*, 2006; Scherer, 1986). The human ability to feign an emotional state through expression has been examined for decades (Anolli, 1997; Ekman *et al.*, 1976) and, though detecting pretense can be



difficult, it is not impossible (Burgoon *et al.*, 2008; Streeter and *et al.*, 1977; Warren *et al.*, 2008). In fact, research on deception has implicated ToM understanding (Shany-Ur *et al.*, 2012) and related brain networks (Herve *et al.*, 2013) in detecting deception. However, deception cannot be equated to other kinds of acting or posing of expressions (though empirical studies of these differences are also lacking). For example, acted or posed emotional expressions can be more intense than spontaneous expressions, thereby making the emotion clearer (Laukka *et al.*, 2012). Since deception relies on similarity to a “true” state, such exaggeration of intensity would not be useful. The use of experimentally induced spontaneous expressions has tried to circumvent this issue. Some significant similarities in acoustic correlates and recognition have been found in the perception of posed and spontaneous expressions, albeit with some differences in effect sizes (Laukka *et al.*, 2011). However, other research has found that posed expressions are not identical to spontaneous expressions (Houston and Holmes, 1975; Laukka *et al.*, 2012) and that this can influence recognition of those expressions (Audibert *et al.*, 2008a, 2008b). Interestingly, one study examining cultural effects, for example, did not find an in-group advantage for spontaneously expressed emotions (Matsumoto *et al.*, 2009). Such differences vary greatly, however, based on individual differences and what modality is used (Audibert *et al.*, 2008b). Thus, further research is required to truly understand the differences and similarities involved.

### **1.5 Objectives**

I hypothesize that, as with the detection of deception, ToM abilities are crucial in detecting the differences between spontaneous and acted expressions due to the fact that the intentional state of the person expressing the emotion is different in these two conditions. I predict that, either implicitly or explicitly, the detection of these different categories of emotional expression can affect emotion recognition. In the following I use the terms authentic and play-acted to distinguish spontaneous and acted expressions. Operationally, I define authentic expressions as those produced spontaneously in a non-experimental setting and describing a personally experienced event. Play-acted expressions are those created in a laboratory setting with actors instructed as to the context, emotion, and text to be spoken. Apart from significance to methodology and emotion expression research, these ideas combine two perceptual categories that provide a platform for examining both emotion perception and ToM and how these interact within the context of top-down and bottom-up processes in social cognition.

These predictions were examined using eighty non-instructed authentic speech recordings of emotionally expressive speech intonation (neutral verbal content) acquired from German radio archives. Each was reenacted by a professional actor or actress provided with context, text and

emotion. **Study I** first examined the basic influence authenticity has on emotion recognition and the process of explicit perception of authenticity itself. Specifically, I was interested in whether authenticity could be perceived by the listener, whether the ToM network is active during such explicit categorization of the context in which the stimuli were produced, and whether this has an impact on the recognition of the emotional stimuli. **Study II** was developed to determine whether the effects seen in **study I** were specific to the cultural group categorizing the emotion, with the goal of advancing the universality versus culture debate on the nature of emotional expressions. The behavioral experiment was repeated with three different groups of participants from Germany, Romania and Indonesia. **Study III** examined whether top-down cognitive processing is involved in the effect authenticity has on emotion perception by explicitly manipulating the information available about a stimulus by way of cueing. Finally, to determine exactly how the perception of authenticity and its effect on emotion perception are triggered bottom-up by stimulus properties, **study IV** examined the effect that an acoustic correlate of authenticity has on brain activation and behavior.

## Chapter 2

### Study I: Authenticity Affects the Recognition of Emotions in Speech: Behavioral and fMRI evidence

Matthis Drolet<sup>1,2</sup>, Ricarda I. Schubotz<sup>2</sup> & Julia Fischer<sup>1</sup>

<sup>1</sup> Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany

<sup>2</sup> Minerva Group Motor Cognition, Max Planck Institute for Neurological Research, Köln, Germany\*

Cognitive, Affective, & Behavioral Neuroscience 12(1), 2012

10.3758/s13415-011-0069-3

#### Author Contributions

MD and RS designed the study; MD collected and analysed the data; MD, RS and JF wrote the manuscript; all authors discussed the results and commented on the manuscript.

### 2.1 Abstract

The aim of this study was to determine how authenticity of emotion expression in speech modulates activity in the neuronal substrates involved in emotion recognition. Within an fMRI paradigm participants judged either the authenticity (authentic or play-acted) or emotional content (anger, fear, joy, or sadness) of recordings of spontaneous emotions and reenactments by professional actors. Contrasting between task types, active judgment of authenticity, more than active judgment of emotion, indicated potential involvement of the Theory of Mind (ToM) network (medial prefrontal cortex, temporoparietal cortex, retrosplenium) as well as areas involved in working memory and decision-making (BA 47). Subsequently, trials with authentic recordings were contrasted with those of reenactments to determine the modulatory effects of authenticity. Authentic recordings were found to enhance activity in part of the ToM network (medial prefrontal cortex). This effect of authenticity suggests that individuals integrate recollections of their own experiences more for judgments involving authentic stimuli than for those involving play-acted stimuli. The behavioral and functional results show that authenticity of emotional prosody is an important property influencing human responses to such stimuli, with implications for studies using play-acted emotions.

## 2.2 Introduction

The ability to feign emotions is presumed to be uniquely human, and is relevant not only to the ability to deceive others, but also forms the foundation for the dramatic arts. In turn, pretense (Cowie *et al.*, 2001; Ekman *et al.*, 1988), pretend play (Rakoczy and Tomasello, 2006) and social deception are all key topics of interest in Theory of Mind research (Bänziger and Scherer, 2007; Frith and Frith, 2007; Hoekert *et al.*, 2010; Premack and Woodruff, 1978). Importantly, while little is known about the potential effects of acting on emotion expression and perception (Goldstein, 2009), a large portion of research into emotion perception makes use of stimulus material obtained from actors.

Studies on human emotion expression that employ acted behavior either focus on prototypical behavior or assume acting to be analogous to authentic emotional expression, both in terms of its production by the sender and its effect on the receiver (Belin *et al.*, 2008; Ethofer *et al.*, 2006; Hietanen *et al.*, 1998; Laukka, 2005). However, findings on social deception (Adolphs, 2009; Ekman *et al.*, 1990; Frick, 1985; Grezes *et al.*, 2006; Scherer, 1986) and, recently, feigning emotions (Scheiner and Fischer, 2011; Wilting *et al.*, 2006){Formatting Citation} indicate that acted behaviors may differ from more naturally occurring ones. Scheiner & Fischer (2011) collected 80 non-instructed “authentic” speech recordings of four emotional expression categories (20 each of anger, sadness, joy & fear) from German radio archives and had each reenacted by a professional actor. Recordings were presented to subjects from three separate cultures (Germany, Romania, Indonesia) in forced-choice rating tasks of emotional content and recording condition (“authentic” versus “play-acted”). While subjects were poor at distinguishing between conditions (only slightly above chance), these nevertheless influenced recognition of emotional content: across all three cultures, anger was recognized more accurately when play-acted and sadness when authentic.

The principal goal of the current study was to determine what specific neuronal substrates are active during judgments on emotion and authenticity and, in a second step, whether stimulus authenticity additionally modulates the activation during these tasks. Participants were presented with emotion- and authenticity-judgment tasks in a functional Magnetic Resonance Imaging (fMRI) setup to determine the activation loci involved during these explicit tasks. Contrasts based on stimulus authenticity could then be applied to determine its effect on activation during either task. As in the study by Scheiner & Fischer (2011), those expressions that were obtained during radio interviews are defined as non-instructed and are labelled as “authentic”, while acted emotions were produced by professional actors following scripted instructions. While radio interviews may not exactly mirror

more private expressions of emotion, these nevertheless represent unscripted emotional expressions in a non-laboratory setting.

Explicit judgment of authenticity was predicted to activate the so-called Theory of Mind (ToM) network. The ToM network is involved in the perception of oneself and other individuals as cognitive entities and, more specifically, in the determination of mental states (Bolton, 1902; Povinelli and Preuss, 1995; Premack and Woodruff, 1978). U. Frith and Frith (2003) reviewed a series of neuroimaging studies indicating that the medial prefrontal cortex (PFC), the temporal poles and the posterior superior temporal sulcus (STS) (extending into the temporoparietal junction (TPJ)) are components of the ToM network. Van Overwalle and Baetens (2009) additionally incorporated the posterior cingulate (PCC) extending into the precuneus (PC) as part of a review of cognitive neuroscience research on mentalising.

Along with the ToM network, the stimuli were likely to activate regions of the brain involved in the processing of emotion and prosody. In this study, explicit judgments of emotional content versus a word detection control condition were expected to activate loci in the orbital and inferior frontal, superior temporal, and inferior parietal cortices (as reviewed by Schirmer and Kotz, 2006). In addition, differential activation to individual emotions has been described in the cingulate cortex (Bach *et al.*, 2008; Buchanan *et al.*, 2000; Wildgruber *et al.*, 2004, 2005).

Considering the previous results by Scheiner & Fischer, we predicted that, on a behavioral level, anger would be detected correctly most often when play-acted, and sadness when authentic. Authentic and play-acted recordings were predicted to each have detection rates slightly above chance. Accordingly, stimulus authenticity (as defined above) was expected to modulate activity in both prosody-specific regions and the ToM network such that play-acted emotions lead to higher activations than authentic emotions (bottom-up/stimulus-driven effects). This prediction was made on the assumption that actors are aware of the dichotomy between the emotion to be communicated and their current environment and state. This, if encoded in the stimuli, potentially represents additional information encoded in play-acted as opposed to authentic recordings.

## 2.3 Methods

### 2.3.1 Participants

Neuroanatomical and performance effects on behavioral tasks can differ as a function of gender (Everhart *et al.*, 2006; Gur *et al.*, 2002). Specifically, emotion recognition accuracy and reaction time show a distinct advantage for women (Belin *et al.*, 2008; Hall and Schmid Mast, 2007; McClure, 2000; Schirmer *et al.*, 2005). Moreover, differences in the size and shape of the cingulate cortex have been

found between men and women (Kochunov *et al.*, 2005). Therefore, 24 female participants 20-30 years of age (mean = 24 years of age; right-handed; German as mother-tongue) were selected and contacted using the Cologne MPI database for fMRI experiments. Only individuals without a history of neurological or psychological complications (including the use of psychiatric medication) were included. Participants were informed about the potential risks of magnetic resonance imaging and screened by a physician. They gave informed consent before participating and were paid afterwards. The experimental standards were approved by the local ethics committee and data were handled pseudonymously.

### 2.3.2 Stimulus Selection

Authentic recordings were selected from German radio archives. The recordings (mono wave format; sample rate of 44.1kHz) were selected from interviews with individuals talking in an emotional fashion (fear, anger, joy, or sadness) about a highly charged ongoing or recollected event (e.g. parents speaking about the death of their children, people winning in a lottery, in rage about injustice, or threatened by a current danger). Emotion was ascertained through the verbal content of the spoken text, as well as the context either described in the recording details or in the recordings themselves. While the potential influence of social acting can never be completely excluded, this effect was minimized by excluding clearly staged settings (e.g. talk-shows) and relying on non-scripted interviews. Of the 80 speech tokens 35 were made outdoors and varied in their noise surroundings, but all selected recordings were of good quality with minimal background noise and did not contain any keywords that could allow inference of the expressed emotion. To ensure inference-free verbal content, text-only transcripts were rated by naïve subjects. Recordings for which the respective emotion was recognized better than chance were replaced. The final original stimulus set consisted of 20 samples per emotion (half from male and half from female speakers) resulting in a total of 80 recordings made by 78 different speakers.

Play-acted stimuli were performed by actors from Berlin, Hanover, and Göttingen, in Germany (each replicated a maximum of 3 recordings of equivalent emotional content; total of 42 actors). The actors were told to express the respective text and emotion in their own way, using only the text, identified context, and emotion (the segment to be used as stimulus was not indicated and the actors never heard the original recording). To remove category effects between authentic and play-acted stimuli the environment for the play-acted recordings was varied by recording in different locations (30 recorded outdoors and 50 recorded indoors) while avoiding excessive background noise. Segments selected as stimuli (mean length 1.76 sec  $\pm$  1.01SD; range 0.357 – 4.843 sec) did not contain any keywords that could allow inference of the expressed emotion. The average amplitude of all stimuli was equalized with Avisoft SASLab Pro Recorder Version 4.40 (Avisoft Bioacoustics, Berlin, Germany).

### 2.3.3 Trial and Stimulus Presentation

The program NBS Presentation (Neurobehavioral Systems, Inc., Albany, California) controlled the details of each experimental run with trial structure, timing and order preprogrammed by way of scripts. Each experimental run (one per participant) included 178 trials (only limited by the number of stimuli available), of which 72 were used for an emotion judgment task (EJ) and 72 for an authenticity judgment task (AJ). In addition, 2 control tasks were included: 16 word detection trials (using the respective 8 authentic and 8 play-acted stimuli not used in the experimental task trials), in which participants had to count occurrences of the word “und” (“and”), and 18 empty trials with pink-noise playback. The word detection task was included as additional control due to interest in the ToM network in this experiment. As has been shown previously (Buckner *et al.*, 2008) the so-called default-mode network that is activated during rest trials partly mirrors the ToM network. To detect potential ToM activation we therefore applied this additional control task as opposed to the empty trials, which could mask actual ToM activity if contrasted with the experimental trials (Spreng *et al.*, 2009). For emotion judgments four responses were possible: anger, sadness, happiness, fear (presented in German as: “Wut”, “Trauer”, “Freude”, “Angst”), while for authenticity judgment responses were presented as authentic (“echt”) and play-acted (“theater”; described to participants beforehand as “gespielt”, i.e. play-acted). To minimize eye movement the maximal line-of-sight angle for visual information was kept under 5 degrees. Trial type and stimulus type pseudo-randomizations were performed using conan (UNIX shell script: Max-Planck Institute (MPI) for Neurology in Leipzig, Germany) to reduce any systematic effects that could have otherwise occurred with simple randomization. Each participant was shown a button sequence on-screen (800x600 pixel video goggles: NordicNeuroLab, Bergen, Norway) complementary to the response box layout (10x15x5 cm gray plastic box with a row of 4 black plastic buttons). For both emotion judgment and word detection all buttons were assigned a possible response. For authenticity judgment only the 2 leftmost buttons were used. (Fig. 2.1)

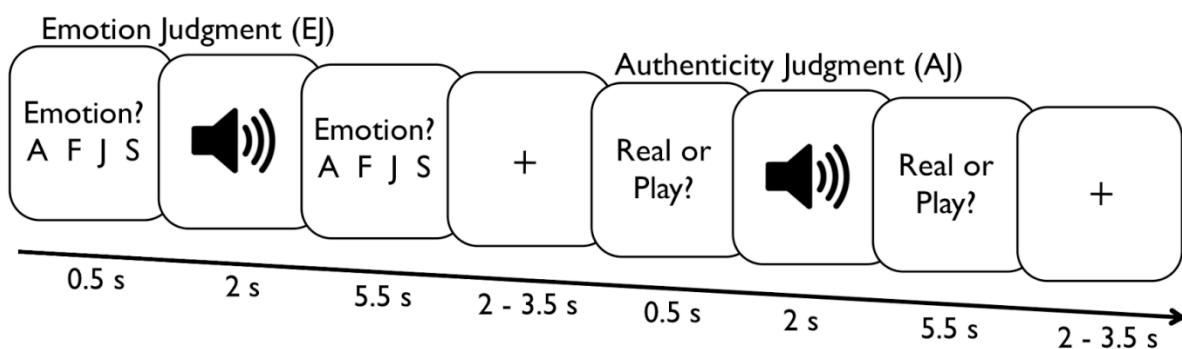


Fig. 2.1 Experimental trial sequences for emotion and authenticity judgment trials with durations in seconds (s). A F J S represent the 4 possible emotion responses (forced-choice design: anger, fear, joy, sadness, respectively).



#### 2.3.4 Experimental Procedure

Participants were fitted with headphones for audio playbacks (NNL: NordicNeuroLab, Bergen, Norway) after they were placed in a supine position on the fMRI table. Imaging was performed with a 3T Siemens MAGNETOM TrioTim (Cologne, Germany) system equipped with a standard birdcage head coil. Participants were placed with their four fingers (excluding thumb) positioned on the response buttons of the response box. Formfitting cushions were used to prevent head, arm, and hand movements. Twenty-two axial slices (210 mm field of view; 64 x 64 pixel matrix; 4 mm thickness; 1 mm spacing; in-plane resolution of 3 x 3 mm) parallel to the bicommissural line (AC–PC) and covering the whole brain were acquired using a single-shot gradient EPI sequence (2000 msec repetition time; 30 msec echo time; 90° flip angle; 1.8 kHz acquisition bandwidth) sensitive to BOLD contrast. In addition to functional imaging, 22 anatomical T1-weighted MDEFT images (Norris, 2000; Ugurbil *et al.*, 1993) were acquired. In a separate session, high-resolution whole-brain images were acquired from each participant to improve the localization of activation foci using a T1-weighted 3D-segmented MDEFT sequence covering the whole brain. Functional data were mapped onto this 3D average using the 2D anatomical images made immediately following the experiment. Including a visual and auditory test before the experiment, one experimental run lasted approximately 45 minutes.

#### 2.3.5 Behavioral Statistical Analysis

Recognition accuracy was analyzed using the R Statistical Package v2.8 (R Development Core Team, 2008). The Generalized Linear Mixed Model (binomial family function) was implemented to determine the best model fit for recognition rates using the *glmer* function from the *lme4* package. The Akaike Information Criterion (AIC) was used to select the model that best fit the data. Choice theory (Agresti, 2007) was implemented as a baseline-category logit model using the *multinom* function from the *VR* package (Smith, 1982). Model predictions are presented in figures using the logit values calculated by R. Respective probability of correct responding is indicated. Reaction times  $\pm$  standard deviation in seconds (reaction times are compared using a standard ANOVA) are also stated. The statistical analyses performed on the data showed main effects of and interactions between emotion and authenticity while accounting for individual participant effects. For the authenticity task the  $d'$  value from Signal Detection Theory was calculated from the recognition probability estimates as a measure of signal discrimination ( $d' = z(\text{correct assessments}) - z(\text{false alarms})$ ).

#### 2.3.6 Functional MRI Statistical Analysis

The functional data were processed using the software package LIPSIA v1.5.0 (Lohmann *et al.*, 2001). This software package is available under the GNU General Public License ([www.cbs.mpg.de/](http://www.cbs.mpg.de/)

institute). Functional data were motion-corrected offline with the Siemens motion correction protocol (Siemens). To correct for temporal offset between the slices acquired in one image a cubic-spline interpolation was employed. Low-frequency signal changes and baseline drifts were removed using a temporal high-pass filter with a cutoff frequency of 1/95 Hz. Spatial smoothing was performed with a Gaussian filter of 5.65 mm FWHM. To align the functional data slices with a 3D stereotactic coordinate reference system, a rigid linear registration with six degrees of freedom (three rotational, three translational) was performed. The rotational and translational parameters were acquired on the basis of the MDEFT slices to achieve an optimal match between these slices and the individual 3D reference dataset. The MDEFT volume dataset with 160 slices and 1-mm slice thickness was standardized to the Talairach stereotactic space (Talairach and Tournoux, 1988). The rotational and translational parameters were subsequently transformed by linear scaling to a standard size. The resulting parameters were then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system, thus generating output data with a spatial resolution of 3 x 3 x 3 mm (27 mm<sup>3</sup>).

Statistical evaluation was based on a least-squares estimation using the general linear model for serially auto-correlated observations (Friston *et al.*, 1998; Worsley and Friston, 1995). The design matrix was generated with a gamma function, convolved with the hemodynamic response function. Brain activations were analyzed time-locked to recording onset and the analyzed epoch was individually set for each trial to the duration of the respective stimulus (mean 1.75 sec., range 0.35-4.84 sec.). The model equation, including the observation data, design matrix, and error term, was convolved with a Gaussian kernel of dispersion of 5.65 sec FWHM to account for temporal autocorrelation (Worsley and Friston, 1995). In the following, contrast images (i.e. beta value estimates of the raw-score differences between specified conditions) were generated for each participant. As all individual functional datasets were aligned to the same stereotactic reference space, the single-subject contrast images were entered into a second-level random effects analysis for each of the contrasts. One-sample t-tests were employed for the group analyses across the contrast images of all participants that indicated whether observed differences between conditions were significantly distinct from zero. The t-values were subsequently transformed into z-scores. To protect against false-positive activation in task-based contrasts, the results were corrected for multiple comparisons by the use of cluster-size and cluster-value thresholds obtained by Monte Carlo simulations with a significance level of  $P < 0.05$  (Lohmann *et al.*, 2008). To protect against false-positive activation in stimulus-based contrasts, a voxel-wise z-threshold was set to  $Z = 2.33$  ( $p = 0.01$ ) with a minimum activation area of 270 mm<sup>3</sup>. Small volume correction (SVC) was performed, correcting the results for a restricted search volume using spheres with a radius of 20 voxels ( $p < 0.05$ ). The volumes of the individual spheres, each centered at a previously published coordinate

(mPFC: -1 -56 33; TPJ: -51 -60 26 & 54 -49 22 ; retrosplenium: -3 51 20) based on the review by van Overwalle & Baetens (2009; Wurm *et al.*, 2011), were summed to calculate the alpha level of each individual activation cluster within the *a priori* hypothesized regions of the ToM network.), based on the published coordinates reviewed by van Overwalle & Baetens (2009).

## 2.4 Results

### 2.4.1 Behavioral Results

Recognition rates and significant differences as calculated by way of the generalized linear mixed model are presented in Table 2.1. Recognition rates for the emotion task produced significant interaction effects with authenticity of stimulus (Fig. 2.2; Supplement 2.1). Anger was detected correctly more often when play-acted than when authentic, while sadness was detected correctly more often when authentic than when play-acted. Fear was only detected near chance levels for both authentic and play-acted recordings. Recognition of joy was at chance levels for both authentic and play-acted recordings. Participants appear to have detected joy less well during the fMRI experiment than reported by Scheiner & Fischer (2011). As the model for authenticity recognition with emotion category as a factor did not produce any significant effects (Supplement 2.2), results are presented without emotion as a factor. Recognition was greater for authentic than for play-acted recordings, but the Reaction Times (RT) did not differ significantly (RT authentic stimuli:  $2.34 \pm 0.39$ s versus Play-acted stimuli:  $2.41 \pm 0.46$ s) indicating that this was not due to general task difficulty. These recognition rates are above chance levels and indicate that authentic recordings were detected correctly more often than play-acted recordings (Fig. 2.3).

Table 2.1 Recognition Accuracy for Emotion and Authenticity Trials

Recording Type		Emotion Trials		Authenticity Trials		
		PE	SE	PE	SE	
Authentic	Anger	0.39	0.04	**		
	Fear	0.23	0.04	***		
	Joy	0.27	0.04	**	0.70	0.05 ***
	Sadness	0.60	0.05	***		
Play-acted	Anger	0.60	0.05	***		
	Fear	0.24	0.06	**		
	Joy	0.28	0.07	**	0.64	0.03 *
	Sadness	0.45	0.07	***		

*Probability estimates (PE) calculated from the Generalized Linear Mixed Model with standard errors (SE). Emotion trials required a determination of the emotional content and authenticity trials a determination of the authenticity content of the recordings. Values for authenticity trials were calculated without effect of recording emotion as this produced a better fit for the model. (\* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$ )*

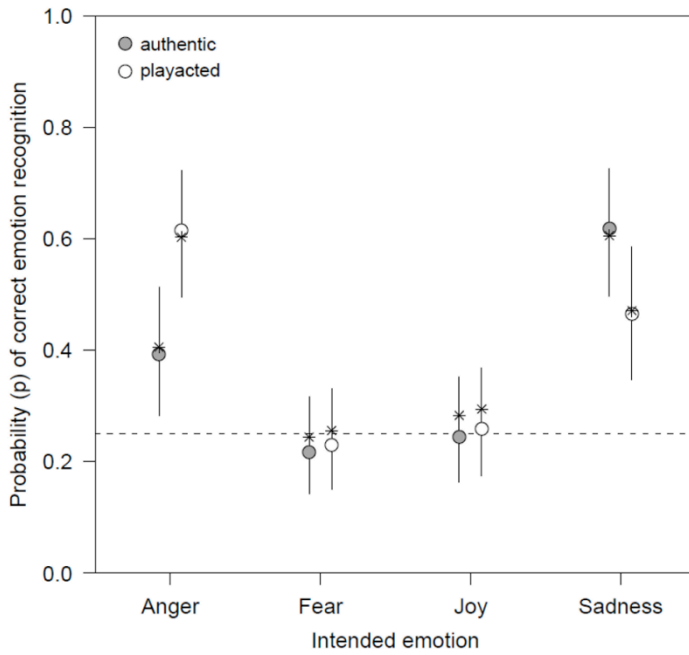


Fig. 2.2 Behavioral results from fMRI for emotion judgments. Actual emotion of recordings on x-axis, probability of recognition of emotion on left y-axis, logit function from statistical model on right y-axis. Circles indicate values calculated by model with 95% confidence intervals. Asterisks indicate actual mean recognition values.

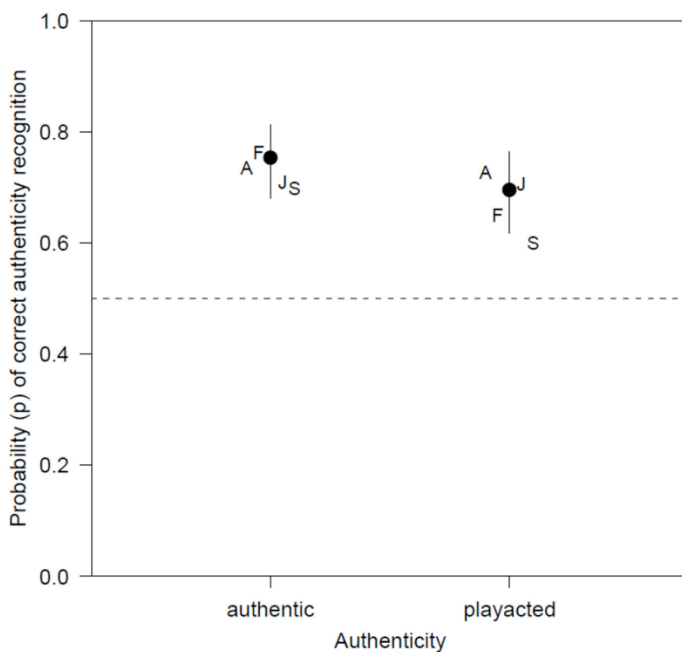


Fig. 2.3 Behavioral results from fMRI for authenticity judgments. Actual authenticity of recordings on x-axis, probability of recognition of authenticity on left y-axis, logit function from statistical model on right y-axis. Circles indicate values calculated by model with 95% confidence intervals. Letters indicate actual mean recognition values by emotion (A = anger, F = fear, H = happy, S = sad).

An ANOVA of reaction times (Table 2.2) did not indicate any significant main effects of stimulus emotion ( $F=2.27$ ,  $p>0.05$ ) or stimulus authenticity ( $F=3.33$ ,  $p>0.05$ ), or interactions ( $F=1.38$ ,  $p>0.05$ ) between emotion and authenticity. It did indicate a main effect for task type reaction times ( $F=9.58$ ,  $p<0.001$ ) with authenticity task (RT  $2.38 \pm 0.41$ s) faster than emotion task (RT  $3.24 \pm 0.50$ s). Due to the significant difference between reaction times in the two experimental tasks, all contrast designs have each trial weighted by reaction time to ensure that any significant effects are not due to differences in task difficulty.

Using the probability estimate for correct authenticity detection as 'hits' and incorrect responses of authentic for play-acted trials as 'false alarms' (Signal Detection Theory analysis) resulted in a  $d'$  value of 0.62 (indicates difficult discrimination between stimuli) as well as a decision criterion of -0.08 (very slight bias for play-acted responses). The Choice Theory analysis indicated that participants had a significant bias in assigning play-acted stimuli to anger and assigning authentic stimuli to sadness, and a significant bias against selecting joy (Fig. 2.4).

### 2.4.2 fMRI Results

As an additional control for the effect of task difficulty in this contrast, the contrast of authenticity versus emotion judgment included reaction time as a random factor (regressor of no interest). Each trial was coded with the actual reaction time recorded for that trial. Activation loci with significantly greater activation for the authenticity judgment [conjunction (authenticity > emotion)  $\cap$  (authenticity > word-detection)] included right inferior frontal gyrus (BA 47), right frontoparietal operculum, bilateral hippocampus, bilateral retrosplenial cortex, bilateral dorsal superior frontal gyrus, medial PFC, bilateral ventromedial PFC, bilateral temporoparietal junction, left superior temporal gyrus, left anterior cingulate cortex, and left anterior middle temporal gyrus (Fig. 2.5, Table 2.3). However, task weightings cannot completely eliminate a potential confounding influence of task difficulty.

Instances of subjects listening to authentic and play-acted recordings were contrasted to detect stimulus effects due to stimulus authenticity. To ensure that significant modulation of resonance is not influenced

Table 2.2 Reaction times by stimulus type

	Authentic		Play-acted	
	RT(s)	SE	RT(s)	SE
Anger	3.25	0.60	2.65	0.53
Fear	3.22	0.63	3.19	0.50
Joy	3.43	0.70	3.23	0.59
Sadness	3.31	0.48	3.58	0.54

RT = reaction time. SE = standard error.

Table 2.3 Network for judging authenticity ([Authenticity judgment > Emotion judgment]  $\cap$  [Authenticity judgment > Word detection])

Area	x	y	z	Z
medial PFC	1	57	9	3.26
	-2	57	9	3.20
ventromedial PFC	-8	15	-15	3.08
	4	33	-9	3.12
hippocampus	-29	-36	-9	4.03
	-29	-18	-12	3.24
	19	-33	-12	3.35
	16	-17	-15	3.02
ventral IFG	31	30	-9	4.97
Retrosplenium	-8	-60	21	5.76
	2	-57	24	5.21
TPJ	-44	-72	36	5.07

Anatomical specification, Talairach coordinates and maximum Z value of local maxima. Abbr.: PFC = prefrontal cortex; TPJ = temporoparietal junction; STG = superior temporal gyrus; SFG = superior frontal gyrus; IFG = inferior frontal gyrus; MTG = middle temporal gyrus.

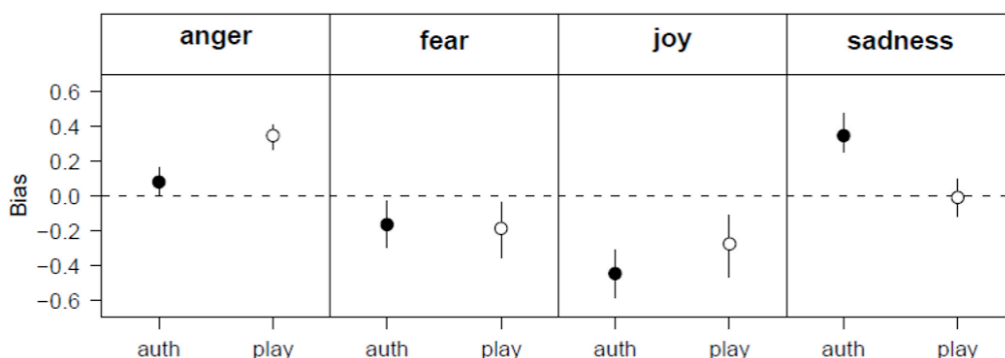


Fig. 2.4 Analysis of emotion recognition data in terms of the choice theory. Given is the response bias for each of the four possible choices (anger, fear, joy, sadness) dependent on authenticity (auth - authentic, play - play-acted). Data are given as means  $\pm$  95% confidence interval. In the absence of any bias, all four log-transformed bias values would be zero. Positive values indicate a bias towards choosing the response named in the headline, a value below zero indicates a bias against choosing the respective response.

by variable numbers of speakers between authentic and play-acted stimuli, the contrast of authentic versus play-acted included speaker repetition as a random factor (regressor of no interest). Repetition was defined for each participant separately, such that the first playback by a specific speaker was labeled as 1, the second as 2 and the third as 3 (as appropriate for specific stimuli). This contrast was first performed for authenticity-judgment tasks to determine whether the network found above is influenced by stimulus authenticity. This revealed significant activation loci in right medial PFC, left retrosplenium, left angular gyrus and left dorsal superior frontal gyrus (Fig. 2.6; Table 2.4).

The determination of whether stimulus authenticity affects emotion recognition was performed with the same authentic versus play-acted contrast but for emotion judgments, which revealed

similar effects of authenticity on the activation seen in right medial PFC and left retrosplenium. Considering the average reaction times that were measured from participant task completion and the similar recognition rates for authentic and play-acted stimuli, it is unlikely that any differences in activation were due to variability in participant concentration level or stimulus difficulty. Additional modulation detected by this contrast included left hemisphere activation of the middle frontal gyrus, precentral gyrus, superior parietal lobule, and middle occipital gyrus. (Fig. 2.6, Table 2.4).

## 2.5 Discussion

The present results demonstrate that the authenticity of emotional prosody affects emotion recognition and modulates cerebral activation as reflected by BOLD fMRI. Explicit judgments about the authenticity of a stimulus, as compared to judgments about the emotional valence, may involve parts of the so-called "Theory of Mind" network. Activation was further boosted for authentic as compared to play-acted stimuli, as was shown for both judgments on emotion and on authenticity, which points to a significant influence of stimulus authenticity during the processing of emotional prosody, even when authenticity is not attended to explicitly. This task-independent modulation

Table 2.4 Direct effects of stimulus authenticity (authentic > play-acted)

Area	x	y	z	Z
<i>During AJ</i>				
dorsal SFG	-23	24	45	3.43
medial PFC	4	54	18	3.03
ventromedial PFC	-8	36	-9	3.26
	1	36	-9	3.46
ventral IFG	-41	27	-6	3.35
anterior MTG	-62	-12	-12	2.81
Retrosplenium	-11	-54	12	3.17
TPJ	-47	-69	36	3.19
<i>During EJ</i>				
medial PFC	1	51	15	2.88
MFG	-23	-15	57	3.51
PrG	-32	-27	51	3.72
SPL	-26	-54	57	2.87
Retrosplenium	-17	-48	9	2.85
	-5	-57	27	2.65
MOCg	-38	-90	0	3.84

Anatomical specification, Talairach coordinates and maximum Z value of local maxima ( $p < 0.05$ , corrected).

Abbreviations: SFG = superior frontal gyrus; PFC = prefrontal cortex; IFG = inferior frontal cortex; MTG = middle temporal gyrus; TPJ = temporoparietal junction; MFG = middle frontal gyrus; PrG = precentral gyrus; SPL = superior parietal lobule; MOCg = middle occipital gyrus.

included the medial prefrontal cortex, a region of the ToM network known to be activated by stimuli with social context and in tasks requiring multi-modal integration (Krueger *et al.*, 2009; Van Overwalle, 2009), as well as the retrosplenial cortex.

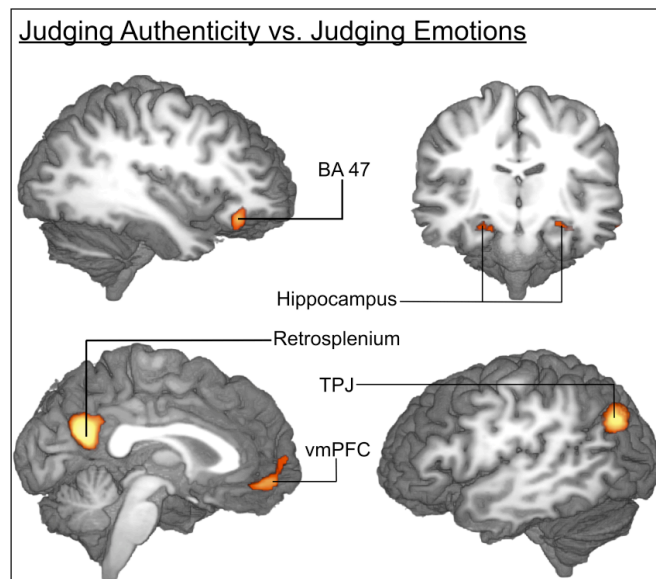


Fig. 2.5 Brain activation correlates of experimental tasks. Group-averaged ( $n = 24$ ) statistical maps of significantly activated areas for authenticity versus emotion judgment trials. To correct for false-results a voxel-wise  $z$ -threshold was set to  $Z = 2.33$  ( $p = 0.01$ ) with a minimum activation area of  $270 \text{ mm}^3$  and mapped onto the best average subject 3D anatomical map. Top left: right view sagittal section through temporal lobe. Top right: anterior view coronal section through thalamus. Bottom left: right view through corpus callosum. Bottom right: left view sagittal section through temporal cortex. TPJ = temporoparietal junction; vmPFC = ventromedial prefrontal cortex.

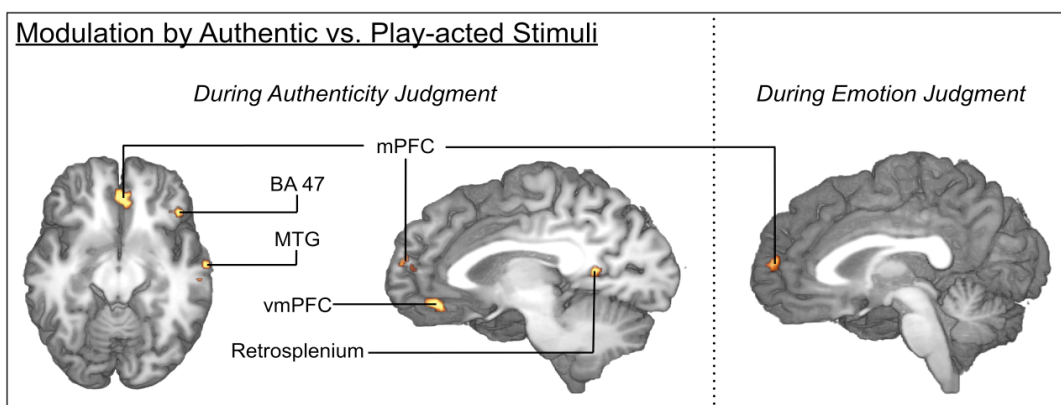


Fig. 2.6 Brain activation correlates of experimental tasks. Group-averaged ( $n = 24$ ) statistical maps of significantly activated areas for authentic versus playacted stimuli (authenticity judgments: left and middle; emotion judgments: right). To correct for false-positive results a voxel-wise  $z$ -threshold was set to  $Z = 2.33$  ( $p = 0.01$ ) with a minimum activation area of  $270 \text{ mm}^3$  and mapped onto the best average subject 3D anatomical map. Left: ventral view axial section through temporal lobes. Middle: left view sagittal section through corpus callosum. Right: left view sagittal section through corpus callosum. MTG = middle temporal gyrus; vmPFC = ventromedial prefrontal cortex; mPFC = medial prefrontal cortex.

Authenticity judgments were contrasted with emotion judgments and a word detection control condition to ensure that any significant activation was indeed specific to the processing of authenticity. This contrast indicated an increase in activity for authenticity judgments in the medial PFC, the retrosplenium and the right and left TPJ. Medial PFC and TPJ subserve a variety of functions linked to ToM and social cognition (Abraham *et al.*, 2008; Frith and Frith, 2003). (Gallagher and Frith, 2003), communicative intent (Frith and Frith, 2007) and self-referential thought (Ochsner and Gross, 2005). TPJ activation points more specifically towards the inclusion of perspective-taking, seen in both spatial perspective-taking (Gallagher and Frith, 2003) and representation of mental states (Saxe, 2006).

Particularly intriguing in the above contrast was the activation in the hippocampi, which points to increased access to previous episodic experiences (Awad *et al.*, 2007; Gilboa *et al.*, 2004; Lackner, 1974). Although the retrieval of past experiences is of key importance in mentalising, hippocampal activation is not always seen to increase specifically for ToM tasks. Combined with the medial PFC, and connected by way of the retrosplenium (Vann *et al.*, 2009), this points to an increase in the active comparison of stimuli with memories in the determination of authenticity. As this is contrasted with the emotion task, it implies that this active comparison with self and social experiences is of greater importance in the determination of social context than in the judgment of the emotion itself. However, while the contrast of task type included reaction time as a per-trial weighting and as a random effect, this cannot completely eliminate a potential confounding influence of task difficulty. Potential differences in task difficulty may induce a similar pattern in the so-called default network, which overlaps with part of the activation seen in the task-based contrast.

While the task-based effects were of additional interest, the focus of this study was the influence of stimulus authenticity on emotion recognition and correlated brain activation. As modulation by stimulus authenticity was examined for each task-type separately, this analysis contrasts trials of identical tasks and similar task difficulty and eliminates the difficulty confounds mentioned above. The ToM network was shown (based on *a priori* ROIs) to be modulated by differences in authenticity, both during emotion and authenticity judgments. Contrasting authentic versus play-acted stimuli showed up-regulation by authentic recordings in the medial PFC and retrosplenium, while discrimination of authenticity was shown to be quite poor through the signal detection analysis ( $d' = 0.62$ : difficulty recognizing authenticity) indicating that authenticity may have a greater influence than immediately apparent simply from behavioral data. While play-acted stimuli were hypothesized to cause greater activation due to the actor's intention to express the instructed emotion, the contrast of authentic versus play-acted emotions showed the opposite effect. Modulation of mPFC activity may be related to its role in integrating various processing streams through its involvement in the representation of mental states (Gallagher and Frith, 2003), communicative intent (Frith and



Frith, 2007), and self-referential thought (Ochsner and Gross, 2005). Participants' lack of knowledge of the exact context of the authentically spoken text may induce greater activation, as opposed to play-acted recordings for which, if recognized, the intent is known. In this case, it is not a difference in intentionality of stimulus production that requires greater activation of the mPFC, but a lack of knowledge about the context in which that stimulus was originally produced.

While differences in production context could have a significant effect during the authenticity task, this would not necessarily be the case during the emotion task, as recognition of the emotion expressed may not necessarily require knowledge of the intention behind that expression. Modulation of the retrosplenium points to a second, non-mutually exclusive explanation for the present activation patterns. Authenticity may lead to significantly different integration of (mPFC), and greater access to (retrosplenial cortex: Vann *et al.*, 2009) previous experiences. This would suggest that individuals integrate recollections of their own experiences more for judgments involving authentic stimuli than for play-acted stimuli. The higher variability of previous authentic experiences thereby lead to greater activation when accessed while play-acted emotions have been experienced less often and in less variable contexts as they are restricted to settings involving acting. Therefore, activation is up-regulated when the stimulus allows for more open-ended assumptions about social context, as with authentic emotions, while for play-acted recordings acting represents the social context and would thereby not induce an increase in ToM activation.

This effect could also explain how a bias towards identifying play-acted stimuli as angry and authentic stimuli as sad, as uncovered through the Choice Theory analysis, could occur with relatively low recognition rates for both emotion (in comparison with other studies) and authenticity (compared with chance rates). While explicit discrimination of authenticity is quite difficult, though largely unbiased in this study (small criterion value from SDT), it nevertheless differentially modulates activity in the ToM network to bias the recognition of emotion. Therefore, authentic and play-acted stimuli have the potential to differentially influence experiments on both behavior and brain activation, even in tasks not explicitly related to stimulus authenticity. While an acoustic analysis of all recordings (Jürgens *et al.*, 2011) comparing the authentic and play-acted sets showed that the two groups of recordings (original; reenacted) did not differ significantly in pitch (Hz:  $188.77 \pm 66.45$ ;  $188.7 \pm 64.61$ ), speech rate (syllables/sec:  $5.62 \pm 1.62$ ;  $5.66 \pm 1.66$ ), or harmonic to noise ratio (mean:  $0.525 \pm 0.081$ ;  $0.517 \pm 0.084$ ), pitch variability and specific vowel voice quality parameters were found to correlate significantly with recording authenticity. Further research will be required to determine which of these stimulus properties are relevant to the authenticity effects discussed here.

Contrasting authentic versus play-acted stimuli also elucidated further authenticity-modulated activity besides the network described above. Increased activation in anterior middle temporal cortex

during the playback of authentic recordings similarly points to the wide range of possible social contexts available for authentically expressed emotions, accentuated by the higher level of integration implied by more anterior activation. While the involvement of the anterior temporal cortex has been related specifically to prosody recognition (Adolphs *et al.*, 2002; Ethofer *et al.*, 2006; Wiethoff *et al.*, 2008), this specific link may be more generally due to its role in social cue perception in general, as was first demonstrated by lesions in the macaque monkey (Bachevalier and Meunier, 2005) and more specifically for social emotions in humans (Burnett and Blakemore, 2009).

### **2.6 Concluding Remarks**

While play-acted stimuli could be expected to cause greater activation in the ToM network than authentic stimuli due to the speaker's additional intention to act, it instead appears that some element of the authentic recordings, as compared with play-acted, increased activation in parts of the ToM network and in social-context processing. This effect may be driven by the greater social significance that authentic emotions have in real life (as compared to play-acted emotions) or, alternatively, authentic emotions may just be more potent than play-acted emotions in triggering spontaneous mentalising. Combining these behavioral and functional results clarifies that the emotional authenticity of prosody is an important property influencing human responses to such stimuli. This has important implications for future research on human emotions and should be taken into account for studies using play-acted prosody.

### **2.7 Acknowledgments**

MD would like to thank Franziska Korb for her exceptional Presentation and Perl scripting introductions and her continued help in ironing out programming kinks. We also want to thank Ralph Pirow for his course in "R" and guidance in the complex world of statistics.

## 2.8 Supplements

### Supplement 2.1a

#### Estimates for Emotion Trials with Authenticity

Fixed Effect		logit	Std. Error	z value
Authentic	Anger	-0.44	0.17	-2.59 **
	Fear	-1.20	0.21	-3.55 ***
	Joy	-1.00	0.21	-2.68 **
	Sadness	0.41	0.20	4.23 ***
Play-Acted	Anger	0.41	0.20	4.23 ***
	Fear	-1.15	0.30	-2.63 **
	Joy	-0.93	0.30	-2.65 **
	Sadness	-0.19	0.28	-5.12 ***

Values from the Generalized Linear Mixed Model fit by the Laplace approximation.

(\* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$ )

Formula: Correctness  $\sim$  Emotion \* Authenticity + (1 | Subject)

### Supplement 2.1b

#### Estimates for Emotion Trials without Authenticity

Fixed Effect	logit	Std. Error	z value
Anger	-0.02	0.13	-0.12
Fear	-1.17	0.15	-7.72 ***
Joy	-0.97	0.15	-6.51 ***
Sadness	0.11	0.14	0.88

Values from the Generalized Linear Mixed Model fit by the Laplace approximation.

(\* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$ )

Formula: Correctness  $\sim$  Emotion + (1 | Subject)

### Supplement 2.1c

#### ANOVA of Model Fit for Emotion Trials

Model	Df	AIC	BIC	logLik	Chisq	Df
f2	5	2168.20	2195.50	-1079.10		
f1	9	2148.90	2198.00	-1065.40	27.31	4 ***

Values for goodness of fit of models compared. (\* =  $p < 0.05$ ; \*\* =  $p < 0.01$ ; \*\*\* =  $p < 0.001$ )

f1: Correctness  $\sim$  Emotion \* Authenticity + (1 | Subject)

f2: Correctness  $\sim$  Emotion + (1 | Subject)

Supplement 2.2a

Estimates for Authenticity Trials with Emotion

Fixed Effect		logit	Std. Error	z value
Authentic	Anger	0.88	0.27	3.29 **
	Fear	1.07	0.24	0.83
	Joy	0.80	0.23	-0.33
	Sadness	0.67	0.23	-0.94
Play-Acted	Anger	0.91	0.23	0.13
	Fear	0.51	0.32	-1.84
	Joy	0.72	0.32	-0.34
	Sadness	0.26	0.32	-1.39

Values from the Generalized Linear Mixed Model fit by the Laplace approximation.

(\* = p<0.05; \*\* = p<0.01; \*\*\* = p<0.001)

Formula: Correctness ~ Emotion \* Authenticity + (1 | Subject)

Supplement 2.2b

Estimates for Authenticity Trials without Emotion

Fixed Effect	logit	Std. Error	z value
Authentic	0.85	0.23	3.78 ***
Play-Acted	0.59	0.11	-2.34 *

Values from the Generalized Linear Mixed Model fit by the Laplace approximation.

(\* = p<0.05; \*\* = p<0.01; \*\*\* = p<0.001)

Formula: Correctness ~ Authenticity + (1 | Subject)

Supplement 2.2c

ANOVA of Model Fit for Authenticity Trials

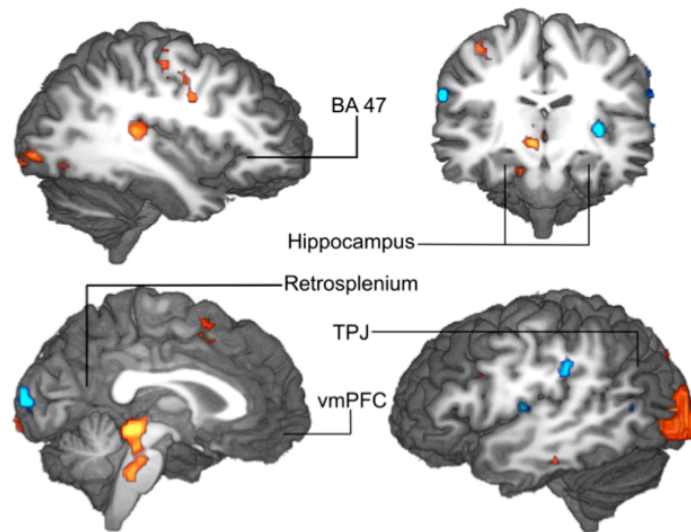
Model	Df	AIC	BIC	logLik	Chisq	Df
f2	3	1969.93	1986.30	-981.97		
f1	9	1968.90	2017.99	-975.45	13.03	6 *

Values for goodness of fit of models compared. (\* = p<0.05; \*\* = p<0.01; \*\*\* = p<0.001)

f2: Correctness ~ Authenticity + (1 | Subject)

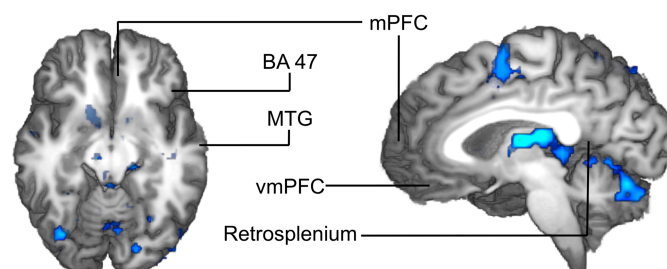
f1: Correctness ~ Emotion \* Authenticity + (1 | Subject)

### Parametric Contrast of Reaction Time



Supplement 2.3 Parametric brain activation correlates of reaction time regressor (of no interest) with network from task-based contrast (labelled). Group-averaged ( $n = 24$ ) statistical maps of significantly activated areas. False-positive correction by voxel-wise  $z$ -threshold set to  $Z = 2.33$  ( $p = 0.01$ ) with minimum activation area of  $270 \text{ mm}^3$  and mapped onto the best subject average 3D anatomical map. Top left: right view sagittal section; top right: anterior view coronal section; bottom left: right view sagittal section. Bottom right: left view sagittal section. TPJ = temporoparietal junction; vmPFC = ventromedial prefrontal cortex.

### Parametric Contrast of Speaker Repetition



Supplement 2.4 Parametric brain activation correlates of speaker repetition regressor (of no interest) with network from stimulus-based contrast (labelled). Group-averaged ( $n = 24$ ) statistical maps of significantly activated areas. False-positive correction by voxel-wise  $z$ -threshold set to  $Z = 2.33$  ( $p = 0.01$ ) with minimum activation area of  $270 \text{ mm}^3$  and mapped onto best subject average 3D anatomical map. Left: ventral view axial section through temporal lobes. Middle: left view sagittal section through corpus callosum. Right: left view sagittal section through corpus callosum. MTG = middle temporal gyrus; vmPFC = ventromedial prefrontal cortex; mPFC = medial prefrontal cortex.

## Authenticity versus Emotion Recognition

### Supplement 2.5 Sample transcripts of original radio recordings

Emotion	Context	Original Transcript	Transcript Translation	Gender
Anger	Two fighting dogs bit to death 6-year-old V. in the schoolyard. Fighting dogs are a big problem in the area and people do not feel protected by the police. The people are shouting at a police officer, blaming him for being too late.	Der Kiosk ruft vor Viertelstund an, "nach Viertelstund" kommt ihr erst, oder was?"	The kiosk called 15 minutes ago, you only come "after 15 minutes" or what?	Male
Fear	The hundred year flood at the Oder threatens whole villages. The water is rising and an inhabitant of an especially low-lying house reports her fears.	Grade unser Haus liegt ziemlich tief. Also 1947 stand das Wasser da schon "bis zum Fensterkreuz". Und wenn das noch schlimmer werden sollte, schätz ich, dass das Haus bald gar nicht mehr zu sehen ist im Wasser. Ja, ich hab ganz dollen Angst	Especially our house lies pretty low. Well, 1947 the water was already up to the window crossbar. And if it should get worse, I guess, that the house won't be visible anymore in the water. Yes, I am very much afraid.	Female
Joy	A married couple has won a new car in the lottery. They report what trick they used to get their ticket on top for the drawing.	Ja. "Mein Mann wollte schon immer im Anfang" der ganzen Sache die 5 Mark einzahlen. Ich schob das ja immer noch ein bisschen hinaus. Eben ein kleiner Schnack, ne.	Yes. "My husband already at the beginning" of the whole thing wanted to pay the 5 Mark. Well, I always delayed that a bit. Just a little joke, huh.	Female
Sadness	In a school massacre, R. killed 16 people before killing himself. One of the victims was the spouse of E. One year after the gun rampage he reports on how he lives with the loss of his spouse. It is still difficult for him to accept her death. Besides his grief, he also felt a heavy pain about the helplessness of his friends.	Und sie sitzen da plötzlich in Tränen "und alles lacht", oder ist dann plötzlich sprachlos, weil das dann passiert und sie das einfach nicht steuern können.	And you are sitting there in tears "and everybody laughs" or are speechless, because it just happened and you are not able to control it.	Male

The segments of transcripts used as stimuli are indicated by quotation marks.

## Chapter 3

### Study II: Encoding conditions affect recognition of vocally expressed emotions across cultures

Rebecca Jürgens, Matthis Drolet, Ralph Pirow<sup>1</sup>, Elisabeth Scheiner<sup>2</sup>, Julia Fischer

Cognitive Ethology Laboratory, German Primate Center

<sup>1</sup> now at the Ecole d'Humanité, Goldern, Switzerland

<sup>2</sup> now at the Institut für Zoophysiologie, Universität Münster

Frontiers in Psychology, 4 (2013)

10.3389/fpsyg.2013.00111

#### Author Contributions

MD and RJ contributed equally to the work. ES and JF designed the study; ES collected the data; RP, MD, and RJ analysed the data; MD, RJ and JF wrote the manuscript; all authors discussed the results and commented on the manuscript.

### 3.1 Abstract

Although the expression of emotions in humans is considered to be largely universal, cultural effects contribute to both emotion expression and recognition. To disentangle the interplay between these factors, play-acted and authentic (non-instructed) vocal expressions of emotions were used, on the assumption that cultural effects may contribute differentially to the recognition of staged and spontaneous emotions. Speech tokens depicting four emotions (anger, sadness, joy, fear) were obtained from German radio archives and reenacted by professional actors, and presented to 120 participants from Germany, Romania, and Indonesia. Participants in all three countries were poor at distinguishing between play-acted and spontaneous emotional utterances (58.73% correct on average with only marginal cultural differences). Nevertheless, authenticity influenced emotion recognition: across cultures, anger was recognized more accurately when play-acted ( $z = 15.06$ ,  $p < .001$ ) and sadness when authentic ( $z = 6.63$ ,  $p < .001$ ), replicating previous findings from German populations. German subjects revealed a slight advantage in recognizing emotions, indicating a moderate in-group advantage. There was no difference between Romanian and Indonesian subjects in the overall emotion recognition. Differential cultural effects became particularly apparent through differential biases in emotion attribution. While all participants labeled play-acted expressions as anger more frequently than expected, German participants exhibited a further bias towards choosing anger for spontaneous stimuli. In contrast to the German sample, Romanian and Indonesian participants were biased towards choosing sadness. These results support the view that emotion recognition rests on a complex interaction of human universals and cultural specificities. Whether and in which way the observed biases are linked to cultural differences in self-construal remains an issue for further investigation.



### 3.2 Introduction

Emotions are an important part of human social life. They mediate between the internal state and external world and they prepare the organism for subsequent actions and interactions. Although there is an ongoing debate about the definition of emotions (see special section of the journal *Emotion Review* (On defining emotions, 2012)), there is a growing consensus among theorists that emotion needs to be viewed as a multi-component phenomenon (Frijda, 1986; Lazarus, 1991; Scherer, 1984). The three major components of emotions are neurophysiological response patterns in the central and autonomic nervous systems; motor expression in face, voice and gesture; and subjective feelings. Many theorists also include the evaluation or appraisal of the antecedent event and the action tendencies generated by the emotion as additional components of the emotional process (Frijda, 1986; Lazarus, 1991; Scherer, 1984; Smith and Ellsworth, 1985).

Different theoretical frameworks have been put forward as to whether emotions are universal and evolved adaptations (Darwin, 1872) or whether they are socially constructed and vary across cultures (Averill, 1980). Both approaches are, however, not mutually exclusive, and it has recently been argued that the dichotomy between nature and nurture should be abandoned (Juslin, 2012; Mason and Capitano, 2012; Prinz, 2004). Matsumoto (1989), for example, argued that although emotions are biologically programmed, cultural factors have a strong influence on the control of emotional expression and perception.

Scherer and Wallbott (1994) conducted a series of cross-cultural questionnaire studies in 37 countries to investigate the influence of culture on the experience of emotions and found strong evidence for both universality and cultural specificity in emotional experience, including both psychological and physiological responses to emotions. Ekman and colleagues (Ekman and Friesen, 1971; Ekman and Oster, 1979; Ekman *et al.*, 1969) tested the universality of facial expressions and demonstrated that a standardized set of photographs depicting different emotion expressions was correctly judged by members of different, partly preliterate, cultures. At the same time, recognition accuracy was higher for members of the cultural background from which the facial expressions were obtained. Thus, facial expressions are considered to be largely universal (but see Jack *et al.*, 2012), while cultural differences are observed in the types of situations that elicit emotions (Matsumoto and Hwang, 2011), in small dialectic-like differences (Elfenbein *et al.*, 2007) and in the culture-specific display rules that alter facial expressions (Ekman *et al.*, 1969; Matsumoto *et al.*, 2008).

The human voice is also an important modality in the transmission of emotional information, both through verbal and nonverbal utterances (Banse and Scherer, 1996; Hammerschmidt and Jürgens, 2007; Juslin and Laukka, 2003; Sauter *et al.*, 2010). Expression of emotion in the voice occurs through

modifications of voice quality (Gobl and Ni Chasaide, 2003) and prosody in general (Scherer, 1986). Initial research on vocal emotion recognition indicated that the patterns in prosodic recognition were largely universal (Frick, 1985), which paralleled the results from facial expressions (Elfenbein and Ambady, 2002). Ratings of vocalizations by listeners showed that they were able to infer vocally expressed emotions at rates higher than chance (Banse and Scherer, 1996; Juslin and Laukka, 2003). In a classic study, Scherer and colleagues (2001) compared judgments by Germans and members of eight other cultures on expressions of emotions by German actors'. They found that with increasing geographical distance from the speakers the recognition accuracy for emotional expressions decreased. Additionally, recognition accuracy was greater for foreign judges whose own language was closer to the Germanic language family. A meta-analysis on emotion recognition within and across cultures revealed that the in-group advantage found by Scherer and colleagues (2001) for German judges is a typical finding in cross-cultural emotion recognition studies (Elfenbein and Ambady, 2002). This meta-analysis included studies that used different types of stimuli, from facial and whole-body photographs to voice samples and video clips. Emotions were universally recognized at better-than-chance levels. However, there was also a consistent in-group advantage: accuracy was higher when emotions were both expressed and recognized by members of the same national, ethnic, or regional group. This advantage was smaller for cultural groups with greater exposure to one another, measured in terms of living in the same nation, physical proximity, and telephone communication (Elfenbein and Ambady, 2002).

Cultural variations in emotion recognition can not only be explained by differences in the emotion encoding, but also by response biases on part of the recipient due to culture-dependent decoding rules (Elfenbein *et al.*, 2002; Matsumoto, 1989). For example, revealing that Japanese participants were less accurate in recognizing anger, fear, disgust and sadness, Matsumoto (1992) suggested a bias against negative emotions in collectivistic societies as an important factor to maintain group stability (but see Elfenbein *et al.*, 2002 for divergent results).

Much of the research cited above has been performed on stereotypical and controlled expressions of emotions often produced by actors. Though actors spend many years perfecting the authenticity and clarity of their portrayals of human behavior and emotions (Goldstein and Bloom, 2011), acted emotional expressions may still be more stereotyped and more intense than spontaneous expressions (Laukka *et al.*, 2011; Wilting *et al.*, 2006; but see Jürgens *et al.*, 2011; Scherer, 2013) and are thought to be more strongly bound by social codes (Hunt, 1941; Matsumoto *et al.*, 2009). In addition, preselected, stereotypical expressions might conceal possible effects of response biases in cross-culture studies due to their clear and unmistakable expression patterns (Elfenbein *et al.*, 2002; Wagner, 1993).

In a series of previous studies we presented listeners with emotional speech tokens produced without external instruction (“authentic”) obtained from radio archives, as well as corresponding tokens reenacted by professional actors (“play-acted”). We found that (German) listeners were poor at distinguishing between authentic and play-acted emotions. Intriguingly, the recording conditions nevertheless significantly affected emotion recognition. Anger was recognized best when play-acted, while sadness was recognized best when authentic (Drolet *et al.*, 2012). Moreover, using an fMRI approach, we found that both explicit recognition of the source of the recording, i.e. whether it was authentic or play-acted (compared to the recognition of emotion) and authentic stimuli (versus play-acted) lead to an up-regulation in the ToM network (medial prefrontal, retrosplenial and temporoparietal cortices). Moreover, acoustic analyses revealed significant differences in the F0 contour, with a higher variability in F0 modulation in play-acted than authentic stimuli (Jürgens *et al.*, 2011).

Considering these findings, we here aim to expand our understanding of the recognition of play-acted and authentic stimuli and biases in emotion recognition. By testing participants from different cultures we intended to gain insights into the influence culture has on our findings. We selected Romanian and Indonesian participants because they differ regarding the geographical distance from the German sample, with a higher degree of overlap between the Romanian and German cultures than between Indonesian and German. Moreover, Romania and Indonesia have been described as collectivistic societies in contrast to the individualistic German society (Hofstede, 1980, 1996; Trimbitas *et al.*, 2007), which allows a comparison of listeners’ culture-dependent response biases on non-instructed, more ambivalent speech tokens (Elfenbein *et al.*, 2002; Matsumoto, 1992). If the observed interaction between emotion recognition and recording condition is based on universal processes in emotion recognition, we would predict a similar pattern across the three cultures. Specifically, more stereotyped displays should be recognized more easily across cultures (Elfenbein *et al.*, 2007). If, in contrast, acting reflects a socially learned code, then the higher recognition of play-acted anger should disappear in the other two cultures (Hunt, 1941; Matsumoto *et al.*, 2009), with a stronger effect in Indonesian than Romanian participants, due to cultural distance. If collectivistic societies foster a response bias against negative emotions, Romanian and Indonesian participants should reveal a bias against judging an emotion as anger, fear or sadness in contrast to the German participants (Elfenbein *et al.*, 2002; Matsumoto, 1992). This effect should be increased in cases in which the stimulus material is less clear and less stereotypical (Elfenbein *et al.*, 2002; Wagner, 1993).

### 3.3 Material and Methods

#### 3.3.1 Recordings

We focused on four emotions that differ in valence, dominance and intensity: anger, fear, joy and sadness (de Vignemont and Singer, 2006; Bryant and Barrett, 2008; Ethofer *et al.*, 2009). These are the most commonly used emotions in this field of research (Sobin and Alpert, 1999; Scherer *et al.*, 2001; Juslin and Laukka, 2003) and were accessible in the radio interviews used for stimulus material. Neutral prosody, while interesting for comparative reasons, is rare and hard to control in real-life settings. One possibility, news anchors, whose voices are characterized by neutral prosody, unfortunately represent a way of speaking more related to acting than to natural speech. We compared emotional expressions that were obtained during radio interviews to reenacted versions of the same stimuli. The authentic speech recordings were selected from the database of a radio station and consisted of German expressions of fear, anger, joy or sadness. The recordings were made during interviews with individuals talking in an emotional fashion about a highly charged ongoing or recollected event (e.g. parents speaking about the death of their children, people winning in a lottery, being in rage about current or past injustice, or threatened by a current danger). Emotions were ascertained through the content of the text spoken by the individuals, as well as the context. While the possibility of social acting can never be completely excluded we aimed to minimize this effect by excluding clearly staged settings (e.g. talk-shows). Stimuli were saved in wave format with 44.1 kHz sample rate and 16 bit sampling depth. Only recordings of good quality and low background noise were selected. Before the experiment, we asked 64 naïve participants to rate the transcripts for emotional content to ensure that the stimulus material was free of verbal content that could reveal the emotion. Text segments that were assigned to a particular emotion above chance level were shortened or deleted from the stimulus set. Thus, the stimuli that were used in the experiment did not contain any keywords that could allow inference of the expressed emotion, as for example "I have known him for 43 years" (translation; original German: "Ich kenn ihn 43 Jahr") was used as a sad stimulus, and "up to the window crossbar" (German: "bis zum Fensterkreuz") as a fear stimulus. Of the chosen 80 speech tokens, 35 were made outdoors and varied in their noise surroundings. The final stimulus set consisted of 20 samples of joy and sadness, 22 samples of anger and 18 samples of fear, half of which were recorded from female speakers, resulting in a total of 80 recordings made by 78 different speakers. Segments had a mean length of 1.9 seconds (*SD*: 1.2 seconds). These wave files represent the so-called authentic stimuli. An information sheet was prepared for each authentic stimulus, which indicated the gender of the speaker, the context of the situation described, and a transliteration of the spoken text surrounding and including the respective selection of text.

The play-acted stimuli were produced by 21 male and 21 female actors (including 31 professional actors, 10 drama students, and 1 professional singer) recruited in Berlin, Hanover, and Göttingen, Germany. Actors were asked to reproduce 2-3 of the authentic recordings. Using the recording information sheet, the actors were told to express the respective text and emotion in their own way, using only the text, identified context, and emotion (the segment to be used as stimulus was not indicated and the actors never heard the original recording). Each actor could practice as long as needed, could repeat the acted reproduction as often as they required, and the recording selected for experimental use was the repetition each actor denoted as their first choice. To reduce any category effects between authentic and play-acted stimuli, the environment for the play-acted recordings was varied and 30 out of 80 randomly selected reenactments were recorded outside. Nevertheless, care was taken to avoid excessive background noise. The relevant play-acted recordings (wave format, 44.1 kHz, 16 bit sampling depth) were then edited so they contained the same segment of spoken text as the authentic recordings. The average amplitude of all stimuli was equalized with AvisoftSASLab Pro Recorder v4.40 (Avisoft Bioacoustics, Berlin, Germany).

### *3.3.2 Ethics*

It was not possible to obtain informed consent from the people whose radio statements were used, as these were not individually identified. The brevity of the speech samples also precluded individual identification; we thus deemed the use of these samples as ethically acceptable. Actors gave verbal informed consent and were paid €20; experimental participants gave written informed consent and were paid €5 for their participation. Both actors and participants were informed afterwards about the purpose of the study.

### *3.3.3 Procedure*

Due to the unequal numbers of speakers in the two conditions, we split the dataset in two and presented the two sets (playback A and playback B) to different groups of listeners. This also served to avoid participant exhaustion. Each set contained 5 authentic and 5 corresponding play-acted duplicates per speaker gender and intended emotion, resulting in a total of 80 stimuli (40 authentic, 40 play-acted) per set. Apart from 3 exceptions the playbacks were prepared in such a way that each actor was present in one set only once and related recordings (authentic versus play-acted) were presented in a pseudo-randomized fashion with the stipulation that speech token pairs were not played immediately after each another to make direct comparisons between recording pairs unlikely.

Each of the two sets of stimuli was presented to 20 listeners (10 female and 10 male) per country, resulting in 40 participants per country. In Germany, all participants were native German speakers recruited at the Georg-August University, Göttingen. Thirty-six were students, three were Ph.D. students and one was an assistant lecturer. The age of German listeners varied between 20 and 33

years, the average age was  $M = 24.4$ ,  $SD = 2.8$  years for the listeners of playback A and  $M = 25.1$ ,  $SD = 3.0$  years for the listeners of playback B. The 40 Romanian listeners were recruited at the Lucian-Blaga-University of Sibiu, Romania. All of them were students. The age of Romanian listeners varied between 18 and 22 years, the mean age was  $M = 20.0$ ,  $SD = 1.2$  years for the listeners of playback A and  $M = 19.5$ ,  $SD = .7$  years for the listeners of playback B. The 40 Indonesian listeners were recruited at the Jakarta University, Indonesia. All Indonesian participants were students aged 18 to 31 years. The mean age was  $M = 20.7$ ,  $SD = 2.8$  years for the listeners of playback A and  $M = 20.5$ ,  $SD = 1.9$  years for the listeners of playback B. Neither the Romanian nor the Indonesian participants spoke any German. Romanian participants were, however, more familiar with German due to a large German community in the town of Sibiu. We did not collect any information about the emotional state of the participants before or during the experiments.

The stimuli were played back using a laptop (Toshiba Satellite with a Realtek AC97 Soundcard) with a program called Emosurvey (developed by Martin Schmeisser). Participants heard the stimuli through earphones (Sennheiser HD 497). They could start the playback of the stimuli themselves and each stimulus could be started a maximum of three times. The ratings were made by way of mouse clicks on the screen. When all questions were answered, the next stimulus could be started. The listeners' ratings were automatically saved in a log file, which could afterwards be transferred to other software packages for analysis. In a forced-choice design participants were asked to determine, for each stimulus, the emotion expressed (emotion-rating: joy, fear, anger, sadness), and whether the emotion was authentic or play-acted (dichotomous authenticity-rating: authentic, play-acted).

### 3.3.4 Statistical Analysis

All models were implemented in the R statistical computing environment (R Development Core Team, 2008). We analyzed the authenticity ratings as well as the emotion ratings with generalized linear mixed models (GLMM) using the `glmer` function from the `lme4` package for binomial data (Bates, 2005). The responses for correct authenticity rating and for correct emotion rating were tested with the predictor variables Country, Intended emotion, Stimulus authenticity, as well as their interactions and the random factors Participant and Text stimulus (model formulation: correct recognition  $\sim$  Country \* Emotion \* Authenticity + Random factor Text stimulus + Random factor Participant). Both models (Authenticity rating and Emotion rating) were compared to their respective null models (including only the intercept and the random factors, model formulation: correct recognition  $\sim$  1 + Random factor Text stimulus + Random factor Participant) using a likelihood ratio test (function `anova` with the test argument "Chisq"). This comparison revealed differences, such that each of the full models accounted for more variance than the null models. Based on the chosen model we specified a set of experimental hypotheses that we tested post-hoc using the `glht` function from the `multcomp`

package (Hothorn *et al.*, 2008), adjusting the p-values for multiple testing using the single-step method.

Assessing recognition accuracy by simply counting hit rates, without addressing potential false alarms or biases (a strong preference towards one response), can be misleading (Wagner, 1993). For instance, if participants strongly prefer rating stimuli as “authentic”, then one would obtain high hit rates for “authentic” speech tokens, but also many wrongly classified play-acted ones (called false alarms). Although the mean recognition rate in this case is quite high, the true ability to recognize authenticity is low. This example shows the importance of calculating biases for understanding rating behavior. A standardized method for analyzing the true discrimination ability for two response options was first introduced as Signal Detection Theory (SDT; Tanner and Swets, 1954). This technique offers both a measure of discriminatory ability  $d'$  (also called sensitivity), which is the true ability to discern one stimulus from another, and a measure of the response bias towards one category, which is independent of sensitivity (criterion  $c$ ). As the emotion recognition task in our study included four response options (four emotions), we analyzed the ratings using Choice Theory (Luce, 1959, 1963; Smith, 1982). Choice theory is a logit-model analogue to SDT, which allows the analysis of more than two discrete response categories. A Choice Theory analysis provides (1) the participants' relative bias ( $b$ ), which is the equivalent criterion  $c$  and (2) dissimilarity values ( $\alpha$ ), which are equivalent to the discriminatory ability  $d'$ .

We implemented the choice theory analysis as a baseline-category logit model (Agresti, 2007). We used the fitted intercept and slope coefficients to derive the bias and similarity parameters of choice theory. The binomial ‘mixed’ model for authenticity recognition (binomial due to the two response options “authentic” and “play-acted”) was calculated in R using the `glmer` function of the `lme4` package (Bates, 2005). The multinomial ‘mixed’ model for emotion recognition was programmed under WinBUGS (Lunn *et al.*, 2000) using the R2WinBUGS interface package (Sturtz *et al.*, 2005) to account for the four response options (“anger”, “fear”, “sadness” and “joy”).

## 3.4 Results

### 3.4.1 Authenticity Recognition

Across cultures, recognition accuracy for authenticity was only slightly above chance ( $M = 58.73\%$ ,  $SD = 8.84\%$ ), with a higher recognition rate for authentic ( $M = 67.81\%$ ,  $SD = 12.37$ ) than for play-acted speech tokens ( $M = 49.58\%$ ,  $SD = 16.78$ ). Post-hoc tests confirmed this difference in recognition rates ( $z = 18.39$ ,  $p < .001$ ; Fig. 3.1). German raters, correct in 62.43 % of cases, were, on average, more accurate in their authenticity ratings than either Romanian (57.20%) or Indonesian raters (56.67%; German - Romanian  $z = 2.99$ ,  $p = 0.028$ ; German– Indonesian  $z = 2.95$ ,  $p = 0.031$ ).

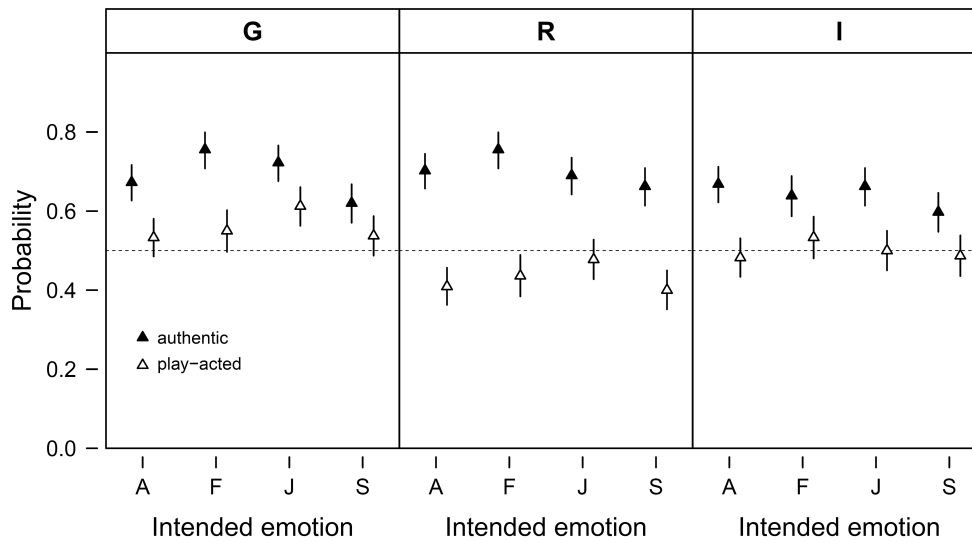


Fig. 3.1 Probability of correct authenticity recognition by intended emotion (A - anger, F – fear, J – joy, S – sadness) and stimulus authenticity (authentic or play-acted). The data are split by cultural affiliation (G - Germany, R - Romania, I - Indonesia). Given are means and 95% confidence intervals. The probability of correct authenticity recognition by chance is .5 as indicated by the dashed horizontal lines.

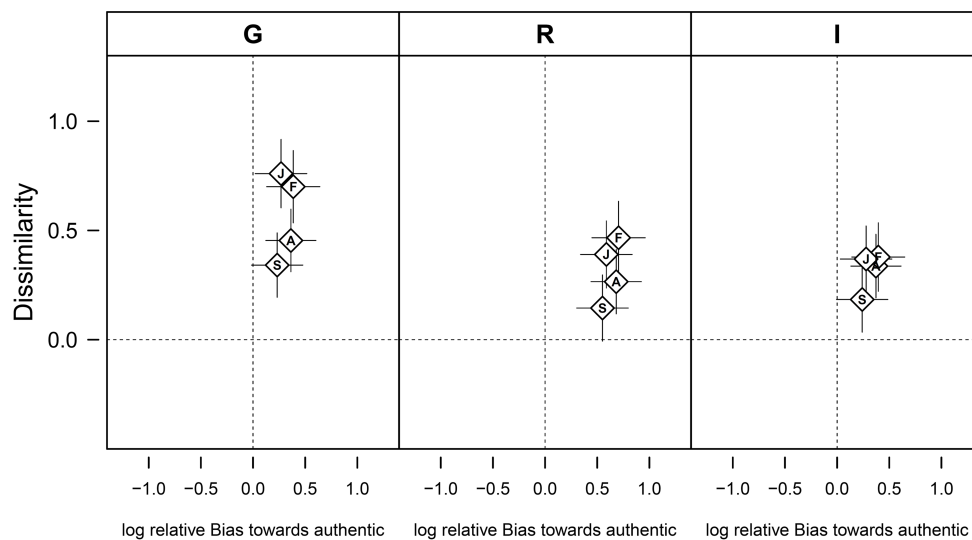


Fig. 3.2 Discrimination of authentic and play-acted vocal expressions of emotions as assessed by choice theory. The discriminatory ability is described by the dissimilarity between authentic and play-acted stimuli (depicting how well the stimuli could be discriminated) and by the participants' relative bias (towards authentic responses), which are plotted against each other. The figure shows how these parameters vary in dependence of cultural affiliation (G - Germany, R - Romania, I - Indonesia) and the intended emotional content (A - Anger, F - Fear, J - Joy, S - Sadness). Positive values on the x-axis indicate a bias towards preferentially choosing the response 'authentic', while higher dissimilarity values indicates a better discrimination ability. Data are given as means  $\pm$  95% confidence intervals.

The analysis of ratings using choice theory revealed that participants had a strong bias towards choosing the response 'authentic' in the authenticity ratings (Fig. 3.2), which may explain the higher recognition accuracy for authentic speech tokens. The post-hoc pair-wise comparisons between the participants of the different countries revealed a significantly greater bias in Romanians than Germans ( $z = 2.64, p = .045$ ; Fig. 3.2).



The overall mean dissimilarity of 0.40 implies a generally low discriminatory capability between authentic and play-acted vocal expressions of emotions (MacMillan and Creelman, 2005). Post-hoc tests revealed that German participants had a higher dissimilarity value and thus a better discriminatory ability than Romanian and Indonesian participants (German-Romanian:  $z=4.535$ ,  $p < .001$ ; German-Indonesian:  $z=4.590$ ,  $p < 0.001$ ).

### 3.4.2 Emotion Recognition

In total, the correct response rate in emotion ratings was 40.65% (SD = 6.41%), which is higher than a chance response rate of 25% resulting from a random selection of one of the four emotions. The emotion recognition ratings showed similar patterns in the three countries (Fig. 3.3). The GLMM analysis revealed that the rate of correct emotion recognition was influenced by Intended emotion, Stimulus authenticity and Country (see Table 3.1 for the results of the post-hoc analysis). Play-acted stimuli were recognized more accurately (42.78%) than authentic stimuli (38.52%). Specifically, play-acted anger was recognized more frequently than authentic anger and authentic sadness more than play-acted sadness. Authenticity did not significantly influence the emotion recognition rates for fear and joy. Concerning the four emotion categories, anger and sadness were on average recognized significantly more frequently than fear and sadness was recognized more frequently than joy. Finally, emotion recognition rates were significantly higher for German participants in comparison to Romanian and Indonesian participants, but not for Romanian participants in comparison to Indonesian participants (Table 3.1).

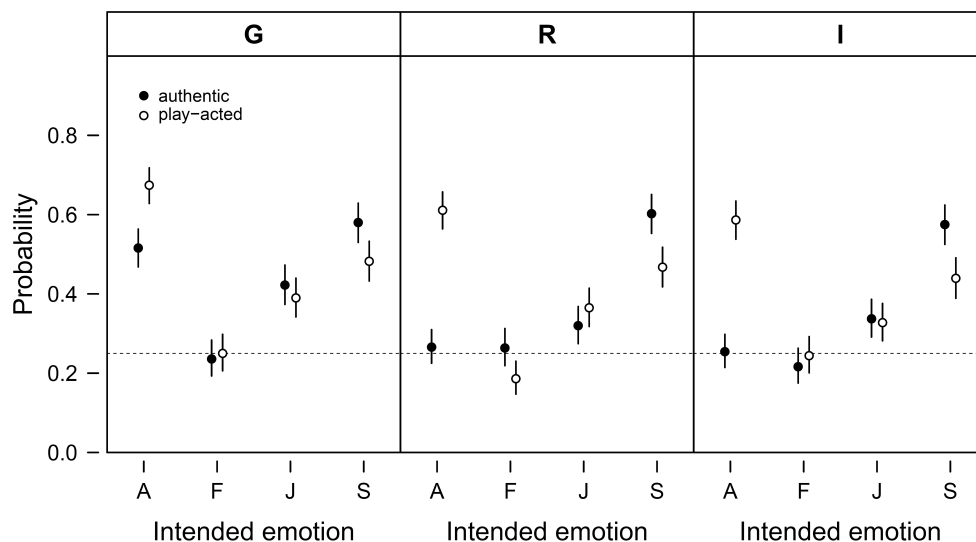


Fig. 3.3 Probability of correct emotion recognition. Given is the probability of correct emotion recognition with respect to the intended emotion (A - anger, F - fear, J - joy, S - sadness) and stimulus authenticity (authentic or play-acted). The data are split by cultural affiliation (G - Germany, R - Romania, I - Indonesia). Given are means and 95% confidence intervals. The probability of correct emotion recognition by chance is .25 as indicated by the dashed horizontal lines.

Table 3.1 Post-hoc tests of cultural affiliation, and stimulus-specific factors (stimulus authenticity, intended emotion) on the probability of correct emotion recognition

Linear Hypotheses	Estimate	Std. Error	z value	Pr(> z )	
Auth - Play == 0	-0.175602	0.046608	-3.768	0.00226	**
Germany - Romania == 0	0.291267	0.059652	4.883	< 0.001	***
Germany - Indonesia == 0	0.351577	0.059665	-5.893	< 0.001	***
Romania - Indonesia == 0	0.06031	0.06036	0.999	0.97849	
A - F == 0	1.22244	0.242372	5.044	<.001	***
A - J == 0	0.536029	0.233757	2.293	0.23193	
A - S == 0	-0.193133	0.233599	-0.827	0.99434	
J - F == 0	0.686411	0.247431	2.774	0.06781	.
S - F == 0	1.415573	0.247282	5.725	<.001	***
S - J == 0	0.729162	0.238845	3.053	0.02912	*
Auth - Play (A) == 0	-1.356013	0.090051	-15.058	<.001	***
Auth - Play (F) == 0	0.077003	0.105342	0.731	0.99776	
Auth - Play (J) == 0	-0.007027	0.088324	-0.08	1	
Auth - Play (S) == 0	0.583629	0.088031	6.63	<.001	***

Note: The p-values are adjusted for multiple testing. Auth = non-instructed. Play = instructed. A = anger. F = fear. J = joy. S = sadness. \*p < .05; \*\*p < .01; \*\*\*p < .001.

The response bias for emotion judgments was calculated with respect to cultural affiliation and stimulus authenticity. In all three countries participants showed a bias towards rating play-acted stimuli as angry (Fig 3.4). This bias was higher for German than for Romanian or Indonesian participants. German participants were also biased towards rating authentic stimuli as angry, while Romanian and Indonesian participants preferentially chose ‘sadness’ and were additionally biased against choosing ‘anger’ when rating authentic stimuli. There was no effect of authenticity or country of origin regarding the responses ‘joy’ and ‘fear’. Indonesian participants, whose bias against ‘joy’ was less distinct than for Romanian or German participants, were the only exception.

The outcome of the calculation of the dissimilarity values for all possible stimulus-response pairs during emotion ratings (including effects of country and stimulus authenticity) are shown in figure 3.5. There were few differences between authentic and play-acted emotional expressions and between the participants of the three countries. High dissimilarity values were found between anger and sadness, which indicates that these emotions could be distinguished easily. The very low dissimilarity values for the stimulus "fear" (see row "F" in the matrix plot in Fig. 3.5) indicate high confusion with the other emotion categories and reflect the low recognition rates for fear.

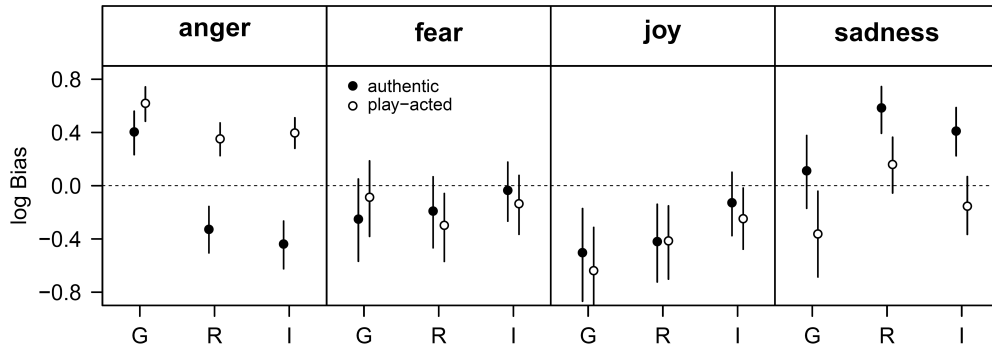


Fig. 3.4 Analysis of emotion recognition data by choice theory. Given is the log-transformed response bias for each of the four possible choices (anger, fear, joy, sadness) with respect to cultural affiliation (G - Germany, R - Romania, I - Indonesia). The filled and open symbols indicate the response bias for authentic and play-acted stimuli. Data are given as means and 95% uncertainty interval. In the absence of any bias, all four log-transformed bias values would be zero. Positive values indicate a bias towards choosing the response named in the headline, whereas a value below zero indicates a bias against choosing the respective response.

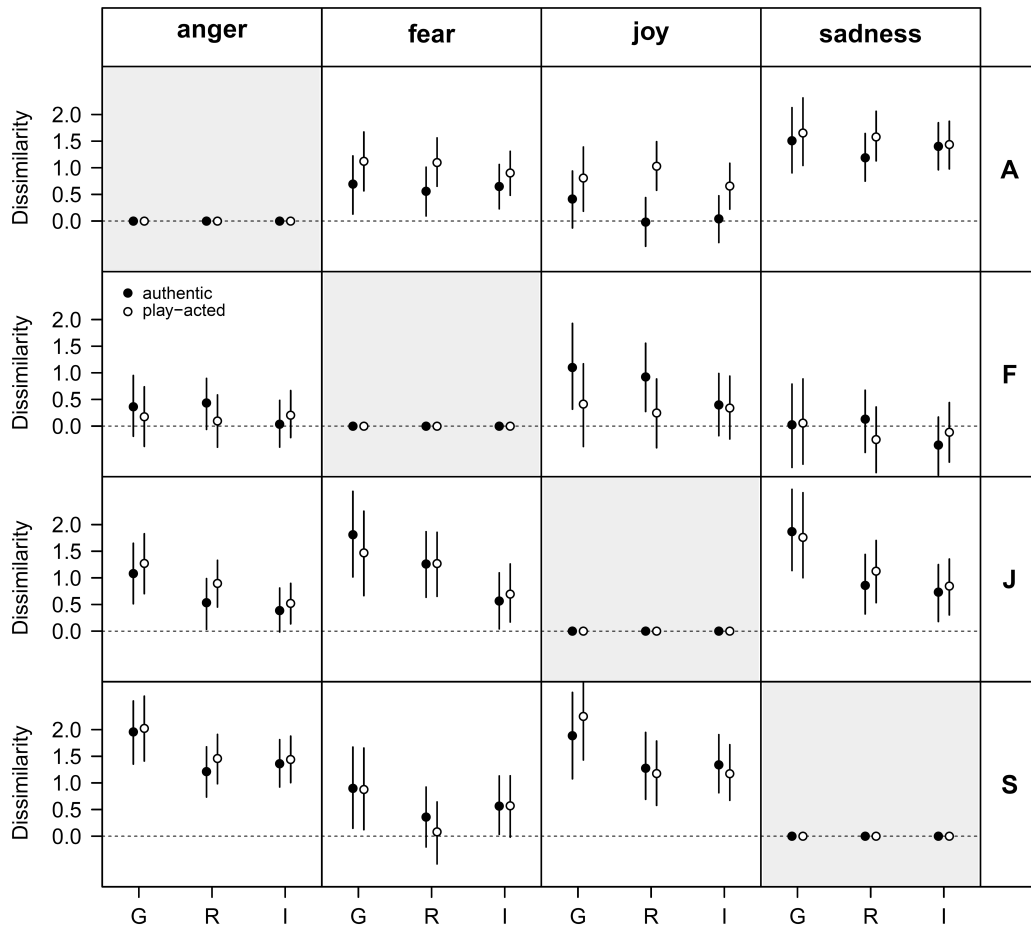


Fig. 3.5 Analysis of emotion recognition data using choice theory. Given is the dissimilarity for different pairs of emotion stimuli with respect to cultural affiliation (G - Germany, R - Romania, I - Indonesia). The rows and columns of this matrix plot indicate the four emotion stimuli (A - Anger, F - Fear, J - Joy, S - Sadness) and the four possible responses (anger, fear, joy, sadness), respectively. Filled and open symbols refer to authentic and play-acted conditions, respectively. Data are given as means and 95% uncertainty interval. The dissimilarity describes how well each stimulus (depicted by rows) is discriminated from each other stimulus (depicted by response columns).

### 3.5 Discussion

Participants in all three cultures had difficulty distinguishing between authentic (spontaneous) and play-acted (instructed) emotional expressions. The recognition of the expressed emotion also showed relatively low rates, but varied with respect to the emotion category and listener country of origin. Notably, the stimulus origin (authentic versus play-acted) had a clear effect on the recognition of vocal expressions of anger and sadness across all three cultures: anger was recognized more frequently when play-acted and sadness was recognized at higher rates when authentic, bolstering earlier findings for an independent German population (Drolet *et al.*, 2012). While these results are significant, it remains unclear what leads to this effect. It may be that play-acted anger is more exaggerated than spontaneously expressed anger, while sadness, in contrast, is more difficult to play-act. On the other hand, it may be that, overall, some stimulus feature makes play-acted stimuli more likely to be perceived as anger and spontaneous stimuli as sadness.

Regarding our initial hypotheses, we found support for the conjecture that play-acted anger was recognized with higher accuracy than authentic anger across cultures, possibly because of its stereotypical nature. For the other three categories, acting does not necessarily appear to be connected with a more exaggerated expression, which is contrary to previous results (Barkhuysen *et al.*, 2007; Laukka *et al.*, 2011). According to our results, play-acted expressions do not represent a socially learned code (Matsumoto *et al.*, 2009). Considering the similar interaction of emotion recognition and stimulus authenticity across the three cultures, our findings lend further support for the notion that emotion recognition is underpinned by human universals.

The fact that listeners of all three cultures were poor at discriminating between authentic and play-acted vocalizations shows that previous findings (Drolet *et al.*, 2012) are applicable cross-culturally. If emotional expressions are indicators for underlying states that may require behavioral responses by the observer (see for controversial discussion Barrett, 2011; Russell *et al.*, 2003), the ability to detect fake emotional expressions should be important and evolutionarily adaptive (Mehu and Scherer, 2012; Schmidt and Cohn, 2001). The inability to distinguish between play-acted and spontaneous expressions is, therefore, counterintuitive, but has also been found in previous studies (see for corresponding results Audibert *et al.*, 2008a; Ekman and O'Sullivan, 1991). People tend to believe in the truthfulness of a statement rather than mistrust it (Levine *et al.*, 1999; Zuckerman *et al.*, 1984). This effect, labeled as "truth bias", is reflected in our participants' bias to choose the answer "authentic" when asked about the encoding condition of the emotional expression. It may be that the social cost of ignoring an emotion in others (miss) or wrongly considering others to be deceivers

(false alarm) may make a bias towards believing in the authenticity of social signals advantageous (Ekman, 1996).

Besides the well documented in-group effect for German participants (Elfenbein and Ambady, 2002; Scherer *et al.*, 2001) in both emotion and authenticity recognition, cultural effects mainly became apparent in rating biases of emotions and not in recognition accuracy or dissimilarity. This has also been demonstrated by Sneddon and colleagues (2011), who showed that emotional stimuli were recognized similarly across different cultures, although the intensity ratings varied. Our initial hypothesis that Indonesian and Romanian participants exhibit a bias against negative emotions was, however, only partially supported. They had, in accordance to our hypothesis, a clear bias against selecting 'anger', but only for authentic stimuli. When listening to the spontaneous speech tokens, Indonesian and Romanian participants preferentially chose 'sadness'. No cultural difference was found for the selection of 'fear'. German participants showed a bias towards selecting 'anger' for both authentic and play-acted stimuli. According to the hypothesis that individualistic cultures are expected to reinforce the expression of negative emotions, German participants may have expected a higher likelihood of being confronted with expressions of anger based on their everyday experiences, regardless of the stimulus type presented. Conversely, the more collectivistic Romanian and Indonesian participants may have expected expressions of sadness to be more likely (see Matsumoto, 1989 for similar results). Thus, sadness seems to rank differently compared to anger and the lumping of all negative emotions in the context of response bias seems to be an oversimplification, which might also explain the absence of clear bias effects in previous studies (Elfenbein *et al.*, 2002; Sneddon *et al.*, 2011). Interestingly, the expected response bias against 'anger' for the Romanian and Indonesian participants is only present for authentic stimuli, which can be explained by stimulus-inherent features of the play-acted speech tokens overriding the response bias (Elfenbein *et al.*, 2002; Wagner, 1993). The link between putative cultural biases requires stronger empirical investigations before firm conclusions can be drawn, in particular regarding limitations on the number and types of countries examined (with respect to language and cultural distance). However, our results demonstrate that the implicit effects of authenticity clearly derive from a complex interaction between stimulus-inherent features and cultural expectations about the likelihood of specific emotional expressions.

Due to the use of spontaneous emotional expressions taken from anonymous radio interviews, our study did not allow for a within-speaker design. We thus could not explicitly test whether individual differences in speaker expressivity affected the results. However, the large number of radio speakers and actors involved (more than generally seen in comparable studies) allowed us to minimize the influences of such effects. Additionally, the recognition rates of fear and joy were quite low compared to previous studies on vocal expressions of emotions (Van Bezooijen *et al.*, 1983; Pell and Kotz, 2011;

Scherer *et al.*, 2001). This is interesting, taking into account that not only the spontaneous emotions, for which a low recognition would have been predicted, but also the play-acted ones, revealed recognition rates near chance levels. In contrast to standard methodology, we did not use exaggerated emotional expressions, preselected speech tokens, or emotional outbursts in a word or two (Van Bezooijen *et al.*, 1983; Pell *et al.*, 2009; Scherer *et al.*, 2001). Actors were provided with longer transcripts (several sentences) to portray emotionally to ensure situations as similar to the authentic recordings as possible. It seems unlikely that specifically these professional actors were unable to encode joy or fear considering that this has been done by laypersons and inexperienced actors before (Van Bezooijen *et al.*, 1983; Pell *et al.*, 2009). In particular, the low recognition rates for joy and fear at or close to chance levels might reveal interesting facts about emotional expressions overall. The inability to recognize fear may indicate that fear is less clear in segments of longer speech samples than previously thought. In fact, we believe that the low recognition rates overall is what made the discovery of the interaction with authenticity, as well as the differences in the response bias, possible. Clearly further work in this direction is needed to understand the relevance of emotion recognition research to day-to-day life. Nevertheless, the cross-cultural results revealed that spontaneous and play-acted emotional expressions are recognized similarly across cultures, indicating that both the recognition of play-acted and spontaneous emotional expressions rest on a similar universal basis. Furthermore, our results emphasize the importance of rating response biases, especially regarding more ambiguous expressions such as those taken from spontaneous situations.

### 3.6 Conclusion

Combining all results, this study supports the view that emotion recognition rests on a complex interplay between human universals and cultural specificities. On the one hand, we found the same pattern of recognition and the same implicit effects of encoding conditions across cultures; on the other hand, cultural differences became evident in distinct biases. In addition, although the low recognition of encoding conditions would appear to argue for acted stimuli in vocal research, the implicit effects on emotion recognition seen here indicate that the design of future studies on vocal emotion recognition must take this variation in stimulus characteristics into account.

### **3.7 Acknowledgements**

This research was funded by the German BMBF (Bundesministerium für Bildung und Forschung) within the collaborative research group “Interdisziplinäre Anthropologie”. We thank Jeanette Freynik for aid with conducting the experiments and Annika Grass for valuable comments on the manuscript.





## Chapter 4

### Study III: Explicit authenticity and stimulus features interact to modulate BOLD response induced by emotional speech

Matthis Drolet<sup>1,3</sup>, Ricarda I. Schubotz<sup>2,3</sup> & Julia Fischer<sup>1</sup>

<sup>1</sup> Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany

<sup>2</sup> Biological Psychology, University of Münster, Münster, Germany

<sup>3</sup> Minerva Group Motor Cognition, Max Planck Institute for Neurological Research, Köln, Germany

Cognitive, Affective, & Behavioral Neuroscience, 1–12 (2013)

10.3758/s13415-013-0151-0

#### Author Contributions

MD, RS and JF designed the study; MD collected the data; MD analysed the data; MD, RS and JF wrote the manuscript; all authors discussed the results and commented on the manuscript

### 4.1 Abstract

Context has been found to profoundly affect the recognition of social stimuli and correlated brain activation. The present study was designed to determine whether knowledge about emotional authenticity influences emotion recognition expressed through speech intonation. Participants classified emotionally expressive speech in an fMRI experimental design as sad, happy, angry, or fearful. For some trials, stimuli were cued as either authentic or play-acted to manipulate participants' top-down belief about authenticity, and these labels were presented both congruently and incongruently to the emotional authenticity of the stimulus. Contrasting authentic versus play-acted stimuli in uncued trials indicated that play-acted stimuli spontaneously up-regulate activity in auditory cortex and regions associated with emotional speech processing. In addition, a clear interaction effect of cue and stimulus authenticity showed up-regulation in the posterior superior temporal sulcus and the anterior cingulate cortex, indicating that cue had an impact on the perception of authenticity. In particular, when a cue indicating an authentic stimulus was followed by a play-acted stimulus, additional activation occurred in the temporoparietal junction, probably pointing to increased load on perspective-taking in such trials. While actual authenticity has a significant impact on brain activation, individual belief about stimulus authenticity can additionally modulate the brain response to differences in emotionally expressive speech.

## 4.2 Introduction

Perception of social stimuli, such as facial expressions, gestures, and vocal expressions of emotions, is a key component in human daily life. These stimuli may vary strongly in their content and ambiguity and frequently require the integration of multiple stimulus modalities (Ethofer, 2006; Hornak *et al.*, 1996; Kreifelts *et al.*, 2007; Regenbogen *et al.*, 2012; Scherer and Ellgring, 2007). Much research in this area focuses on the variation in stimulus quality and bottom-up influence on perception, while not addressing top-down modulation (Banse and Scherer, 1996; Pell *et al.*, 2009; Scherer, 1991). Context may, however, have a profound top-down influence on the recognition of social stimuli (Nelson and Russell, 2011; Pell, 2005). Albert and colleagues (2010) showed that stimuli portrayed with positive emotional context induced greater activation during response inhibition in an ERP experimental design, while Schirmer (2010) showed that valence of neutral words in memory is influenced by the affective prosody present during encoding, but does not affect retrieval ability. Much of this research is focused on the bottom-up effect of perceiver emotional state (Blanchette and Richards, 2010) or stimulus emotional content (Barrett and Kensinger, 2010; Chung and Barch, 2011), while, conversely, the influence of higher cognition on emotion perception has yet to be explored in detail.

To examine the overall integration of both top-down and bottom-up processes Teufel and colleagues proposed the Theory of Social Perception as a conceptual framework (Teufel *et al.*, 2010). Central to this framework is Theory of Mind (Premack and Woodruff, 1978), with its role in interpreting social stimuli by ascribing thoughts and intentions to other individuals (Hooker *et al.*, 2008; Mier *et al.*, 2010; O'Brien *et al.*, 2011). The brain network responsible for these processes includes the medial prefrontal cortex (mPFC), posterior superior temporal sulcus (pSTS), temporoparietal junction (TPJ), and retrosplenial cortex (Frith and Frith, 2003). The importance of the integration of emotion and ToM can be seen in research on empathy (Shamay-Tsoory, 2011; Vollm *et al.*, 2006) and affective theory of mind (Mier *et al.*, 2010). However, here too the influence of ToM on the perception of emotional stimuli remains largely unexamined.

Drolet and colleagues (2012) showed that authenticity can affect the perception of the EMOTION expressed. We contrasted speech produced without external instruction (“authentic”) with that produced by professional actors after instruction (“play-acted”). The ability to feign emotions and its relation to theory of mind is known from play behavior (Rakoczy *et al.*, 2004; Schmidt *et al.*, 2008) and deceit (DePaulo, 2003; Grezes *et al.*, 2006, 2004). As the detection of play-acting involves the perception of the speaker’s intention, the perception of acted emotions is an ideal case of the interaction of emotion perception and higher cognition. In addition, while professional acting has

major social and economic influence through entertainment and research, its effect on perception has not been thoroughly examined (Goldstein, 2009, 2011).

The present study was designed to examine effects of assumed authenticity on emotion perception by labeling stimuli with contextual information. The aim was to determine whether assumptions about the authenticity of emotional speech intonation (prosody) bias the recognition of the expressed emotion. The effect of emotional prosody on behavioral responses is well known (Buchanan *et al.*, 2000; Jaywant and Pell, 2011; Scherer, 1991) and brain activation correlated with its perception is an active area of research (Kotz *et al.*, 2012; Vigneau, 2006; Wildgruber *et al.*, 2002) Drolet *et al.* (2012) showed that anger was recognized best when play-acted, while sadness was recognized best when authentic, and, using Choice Theory (Kornbrot, 1978; Luce, 1959), we found a bias, such that authentic stimuli were more often categorized as sad and play-acted ones were more often categorized as angry, overall. BOLD response analyses showed that both active discrimination of AUTHENTICITY and authentic stimuli (versus play-acted) up-regulated activation in the ToM network (medial prefrontal, retrosplenial and temporoparietal cortices). Accordingly, we suggested that ToM network activity is induced in the listener by differences in authenticity of vocal emotional expression. What remained unclear, however, was whether this effect was purely stimulus-driven (“bottom-up”) or additionally modulated top-down by assumptions about stimulus authenticity. In the former case, emotional intonation itself may change the listener’s perception of stimulus EMOTION directly. In the latter, acoustic differences may signal AUTHENTICITY, the perception of which engages ToM as top-down modulation of perception in the listener, which then influences emotion categorization. Therefore, the study presented here was designed to determine whether the hypothesis of the top-down effects of stimulus AUTHENTICITY could be corroborated.

To do so, we used a modified paradigm in which participants only categorized stimuli by EMOTION. To manipulate top-down modulation, we labeled stimuli on-screen as authentic or play-acted before playback, both congruently and incongruently to the stimulus. If the effect of AUTHENTICITY on emotion categorization is purely a bottom-up effect of differences in emotional expression, these labels should have no effect on recognition rates. If, however, this effect is induced top-down, labels should alter or even override the stimulus-induced AUTHENTICITY effect. More specifically, authentic cues should decrease recognition of anger and increase recognition of sadness (Drolet *et al.*, 2012) irrespective of actual stimulus authenticity. Furthermore, due to the previously shown involvement of ToM, we expected authentic cues to up-regulate activity in the medial prefrontal cortex, the retrosplenium and the temporoparietal junction (ToM: Frith and Frith, 2003) if the effects were induced top-down. However, if AUTHENTICITY affects emotion recognition bottom-up, we expected authentic stimuli to up-regulate BOLD as found by Drolet and colleagues (2012), as well as regions associated with emotional prosody perception, including the orbital and inferior frontal, superior temporal, and inferior parietal

cortices (as reviewed by Schirmer and Kotz, 2006). Activation in the cingulate cortex was also predicted due to its involvement in both emotion perception and ToM processes (Bach *et al.*, 2008; Buchanan *et al.*, 2000; Wildgruber *et al.*, 2005).

## 4.3 Methods

### 4.3.1 Participants

18 female participants (20-30 years, mean 24 years, right-handed, German mother-tongue) were contacted through the Cologne MPI database for fMRI experiments. Only individuals without neurological or psychological complications (including the use of psychiatric medication) were selected. Participants were informed about the potential risks of magnetic resonance imaging and screened by a physician. They gave informed consent before participating and were remunerated. The Ethics Commission of Cologne University's Faculty of Medicine approved the experimental standards and procedure and data were handled pseudonymously with identity blind to the experimenter.

### 4.3.2 Stimulus Selection

Original recordings (mono wave format; sample rate of 44.1kHz) were selected from German radio archives. These were interviews with individuals talking in an emotional fashion (anger, fear, joy or sadness) about a highly charged ongoing or recollected event (e.g. the death of a child, winning a lottery, in rage about injustice, or threatened by a current danger). EMOTION was ascertained through verbal content and context described in recording summaries. Staged settings (e.g. talk-shows) and scripted interviews were excluded. Of the 80 speech tokens 35 were made outdoors and varied in their noise surroundings, but were of good quality with minimal background noise. To ensure inference-free verbal content, naïve subjects rated text-only transcripts. Recordings with EMOTION recognized better than chance were replaced to ensure semantic content would not indicate emotional content. The original set consisted of 20 samples per EMOTION (half male, half female; total of 80 recordings by 78 speakers; mean  $1.75s \pm 1.00SD$ ; range 0.36–4.06s). The average amplitude of all stimuli was equalized with Avisoft SASLab Pro Recorder Version 4.40 (Avisoft Bioacoustics, Berlin, Germany).

Actors from Germany performed the play-acted stimuli (42 actors each replicated a maximum of 3 recordings of equivalent emotional content). The actors were told to express the respective text and EMOTION in their own way, using only the text, identified context, and EMOTION (the segment to be used as stimulus was not indicated and the actors never heard the original recording). Play-acted recording environment was varied by location while avoiding excessive background noise (mean

1.76s±1.02SD; range 0.38–4.84s). The average amplitude of all stimuli was equalized with Avisoft SASLab Pro Recorder Version 4.40 (Avisoft Bioacoustics, Berlin, Germany).

#### 4.3.3 Trial and Stimulus Presentation

NBS Presentation (Neurobehavioral Systems, Inc., Albany, California) controlled trial sequence and the timing of each experimental run. Each participant completed n=178 trials (limited by the number of available stimuli with no stimulus repetition). N=144 stimuli were used for the emotion judgment task (EJ). One-third (n=48) were not cued (no context information was provided), one-third were cued as authentic, and one-third were cued as play-acted. Of the cued trials, half of the authentic and play-acted cues were congruent to the stimulus presented in the trial, while the others were incongruent. For emotion judgments, four responses were possible: anger, sadness, happiness, fear (presented in German as: “Wut”, “Trauer”, “Freude”, “Angst”). Cues were presented above the response options as authentic or played (“echt” or “spiel” respectively). The remaining 30 trials were used to implement 2 independent control tasks: 18 empty trials with pink-noise playback, and 16 age task trials (with 8 authentic and 8 play-acted stimuli not used in experimental trials) in which participants had to determine the age of the speaker. Speaker age determination was selected as a behavioral control task due to its similarity to the emotion judgment task in relying on prosodic information (Linville, 1985; Schötz, 2007). Participants were tasked to select the age closest to the age of the speaker (20, 30, 40 or 50).

Trial type and stimulus type were pseudo-randomized to reduce any systematic effects that could have otherwise occurred with simple randomization. Each participant was shown a button sequence on-screen (800x600 pixel video goggles: NordicNeuroLab, Bergen, Norway) complementary to the response box layout (10x15x5 cm gray plastic box with a row of 4 black plastic buttons). All buttons were assigned a possible response for the emotion judgment and age categorization tasks. The maximal line-of-sight angle for visual information was kept under 5 degrees to minimize eye movement. (Fig. 4.1)

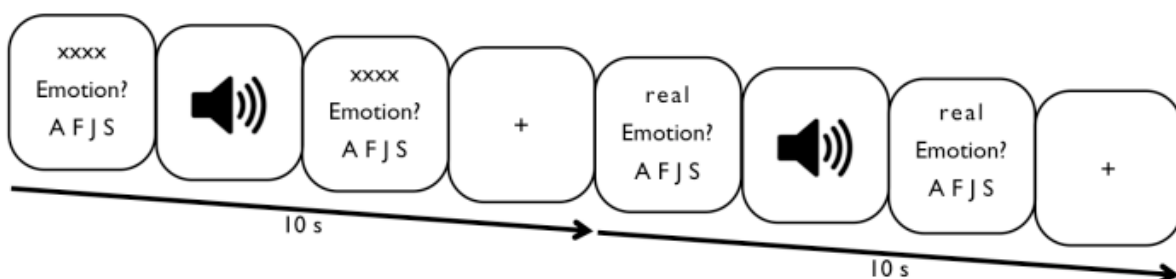


Fig. 4.1 Experimental trial sequences for uncued and cued emotion judgment trials with durations (in seconds). A F J S represent the four possible emotion responses (forced choice design: anger, fear, joy, and sadness, respectively). xxxx indicates display content during uncued trials.

#### 4.3.4 Experimental Procedure

Participants were fitted with headphones for audio playback (NNL: NordicNeuroLab, Bergen, Norway) after they were placed in a supine position on the fMRI table. Imaging was performed with a 3T Siemens MAGNETOM TrioTim (Cologne, Germany) system equipped with a standard birdcage head coil. Participants were placed with their four fingers (excluding thumb) positioned on the response buttons of the response box. Formfitting cushions were used to prevent head, arm, and hand movements. Twenty-four axial slices (210 mm field of view; 64 x 64 pixel matrix; 4 mm thickness; 1 mm spacing; in-plane resolution of 3 x 3 mm) parallel to the bicommissural line (AC–PC) and covering the whole brain were acquired using a single-shot gradient EPI sequence (2000 msec repetition time; 30 msec echo time; 90° flip angle; 1.8 kHz acquisition bandwidth) sensitive to BOLD contrast. In addition to functional imaging, 24 anatomical T1-weighted MDEFT images (Norris, 2000; Ugurbil *et al.*, 1993) were acquired. In a separate session, high-resolution whole-brain images were acquired from each participant to improve the localization of activation foci using a T1-weighted 3D-segmented MDEFT sequence covering the whole brain. Functional data were mapped onto this 3D average using the 2D anatomical images made immediately following the experiment. Including a visual and auditory test before the experiment (duration of 1 min), one experimental run lasted approximately 45 minutes.

#### 4.3.5 Behavioral Statistical Analysis

Recognition accuracy was analyzed using the R Statistical Package v2.15 (R Development Core Team, 2008). The generalized linear model was implemented to determine the best model fit for recognition rates using the `glmer` function from the `lme4` package with binomial error structure and logit link function (GLMM; Baayen *et al.*, 2008). The basic model examined the effect of the independent variables (stimulus EMOTION, stimulus AUTHENTICITY, and CUE type) on the response variable (whether EMOTION rating was correct or not) with participant and speaker repetition included as random effects (`glmer ( Rating ~ Emotion x Authenticity x Cue + ( 1|Participant ) + ( 1|Repetition )`)). First, the overall influence of each categorical independent variable was examined by way of a likelihood ratio test (LRT) with a Chi-squared distribution ( $\chi^2$ ). Models with and without the effect of interest were compared using the ANOVA function, indicating the overall variable effect (main effect and interactions). If a significant effect was found, models were examined for interactions (main effects are ignored). If no interaction was found, main effects of the respective independent variables on the response variable were examined. The effects of each categorical independent variable or, if appropriate, interactions between variables, are indicated by  $\chi^2$  values, degrees of freedom and effect significance (p-value). Mean behavioral data are presented as reaction times (in seconds) as well as the respective probability of correct responding for each EMOTION, AUTHENTICITY and CUE category with 95% confidence intervals.

### 4.3.6 Functional MRI Statistical Analysis

After motion correction using Siemens rigid-body registration protocol (München, Germany), the functional data were processed using the software package LIPSIA v1.5.0 (Lohmann *et al.*, 2001). This software package is available under the GNU General Public License ([www.cbs.mpg.de/institute/](http://www.cbs.mpg.de/institute/)). To correct for temporal offset between the slices acquired in one image a cubic-spline interpolation was employed. Low-frequency signal changes and baseline drifts were removed using a temporal high-pass filter set for each scanned participant dependent on the pseudo-randomized design (filter frequency range: 1/75-85 Hz). Spatial smoothing was performed with a Gaussian filter of 5.65 mm FWHM ( $\sigma = 2.4$ ). To align the functional data slices with a 3D stereotactic coordinate reference system, a rigid linear registration with six degrees of freedom (three rotational, three translational) was applied. The rotational and translational parameters were acquired on the basis of the MDEFT slices to achieve an optimal match between these slices and the individual 3D reference dataset. The MDEFT volume dataset with 160 slices and 1-mm slice thickness was standardized to the Talairach stereotactic space (Talairach and Tournoux, 1988). The rotational and translational parameters were subsequently transformed by linear scaling to a standard size. The resulting parameters were then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system, thus generating output data with a spatial resolution of 3 x 3 x 3 mm (27 mm<sup>3</sup>).

Statistical evaluation was based on a least-squares estimation using the general linear model for serially auto-correlated observations (Friston *et al.*, 1998; Worsley and Friston, 1995). The design matrix, with trials classified by all levels of the three main factors EMOTION category, CUEING type, and stimulus AUTHENTICITY, was generated with a delta function, convolved with the hemodynamic response function (gamma function). Each trial in the design matrix was identified by onset time and stimulus length, while speaker repetition was included as REG No INT to prevent this from influencing the statistical analysis. Brain activations were analyzed time-locked to recording onset and the analyzed epoch was individually set for each trial to the duration of the respective stimulus (mean 1.75s, range 0.35-4.84s). The model equation, including the observation data, design matrix, and error term, was convolved with a Gaussian kernel of dispersion 5.65 sec FWHM to account for temporal autocorrelation (Worsley and Friston, 1995). In the following, contrast images (i.e. beta value estimates of the raw-score differences between specified conditions) were generated for each participant. As all individual functional datasets were aligned to the same stereotactic reference space, the single-subject contrast images were entered into a second-level random effects analysis for each of the contrasts. One-sample t-tests (two-sided) were employed for the group analyses across the contrast images of all participants that indicated whether observed differences between conditions were significantly distinct from zero. The t-values were subsequently transformed into z-scores. To correct for false-positive results, first an initial voxel-wise z-threshold was set to  $z = 2.58$  (P



< 0.01; two-sided). In a second step, the results were corrected for multiple comparisons at the cluster level using cluster-size and cluster-value thresholds obtained by Monte Carlo simulations at a significance level of  $P < 0.05$  (i.e. the reported activations are significantly activated at  $P < 0.05$ ).

BOLD contrasts were computed based on task and stimulus categories. Trials with authentic stimulus playback were contrasted against trials with play-acted stimulus playback. First, a contrast within only non-cued trials examined the hypothesis-driven main effect of AUTHENTICITY, independent of CUEING. Second, the main effect of CUEING was examined by contrasting cued versus uncued trials. Then, to determine how CUEING modulates authenticity-induced activation, the interaction effect of CUEING and AUTHENTICITY was examined using a 2x2 factorial design. Trials were labeled according to both stimulus authenticity and cue type, producing a set of 4 experimental trial categories to be analyzed (1. authentic-stimulus authentic-cue; 2. play-acted-stimulus authentic-cue; 3. authentic-stimulus play-acted-cue; 4. play-acted-stimulus, play-acted-cue). In this design the contrast to indicate activation due to interaction of authenticity and cue is [authentic-stimulus authentic-cue > play-acted-stimulus authentic-cue] > [authentic-stimulus play-cue > play-acted-stimulus play-cue]. This contrast shows the effect cue validity has on the effect of stimulus authenticity. This simultaneously represents a contrast of incongruent and congruent trials, so that the interaction effects of stimulus and cue can be understood as the effects of cue congruency on stimulus perception. To determine the direction of the interaction effects found in this contrast, authentic stimuli were contrasted against play-acted stimuli in trials with authentic and play-acted cues individually as post-hoc t-tests. Due to the complexity of the effects of interest in this hypothesis-driven analysis, small volume correction was performed using a restricted search volume (spheres with a 20-voxel radius), as well as cluster-size and cluster-value thresholds obtained by Monte Carlo simulations. The volumes of the individual spheres, each centered at a previously published coordinate (mPFC: -1 -56 33; TPJ: -51 -60 26 & 54 -49 22 ; retrosplenium: -3 51 20) based on the review by van Overwalle & Baetens (2009; Wurm *et al.*, 2011) for ToM and the review by Vogt (2005) for the Cingulate Cortex, were summed to calculate the alpha level of each individual activation cluster within the a priori hypothesized regions of the ToM network. Therefore, analyses simultaneously included all voxels from the coordinates from all hypotheses.

## 4.4 Results

### 4.4.1 Behavioral results

First, the effect of CUEING was examined as an independent variable in the model (three levels: no cue, authentic cue, play-acted cue). This model, including all four EMOTION categories (anger, fear, joy, sadness), stimulus AUTHENTICITY (authentic, play-acted), and CUEING, indicated that CUEING had no

significant effect on emotion categorization ( $\chi^2(15)=16, p>0.10$ ). As the effect of AUTHENTICITY on emotion recognition was predicted only for anger and sadness stimuli, we examined a reduced model that contained only those two emotions to ensure that any potentially weak effects would not be overlooked. Overall, however, this model also indicated that CUEING had no influence on recognition either in interaction with AUTHENTICITY and EMOTION (model including 3-way interaction:  $\chi^2(6)=8.43, p>0.10$ ) or EMOTION only (cueing and emotion:  $\chi^2(6)=5.91, p>0.10$ ). While effect sizes are different between CUEING categories, none of these differences are significant and are unlikely to be systematic (Fig. 4.2). Categorization rates and reaction times are presented in Table 4.1, split by trial type.

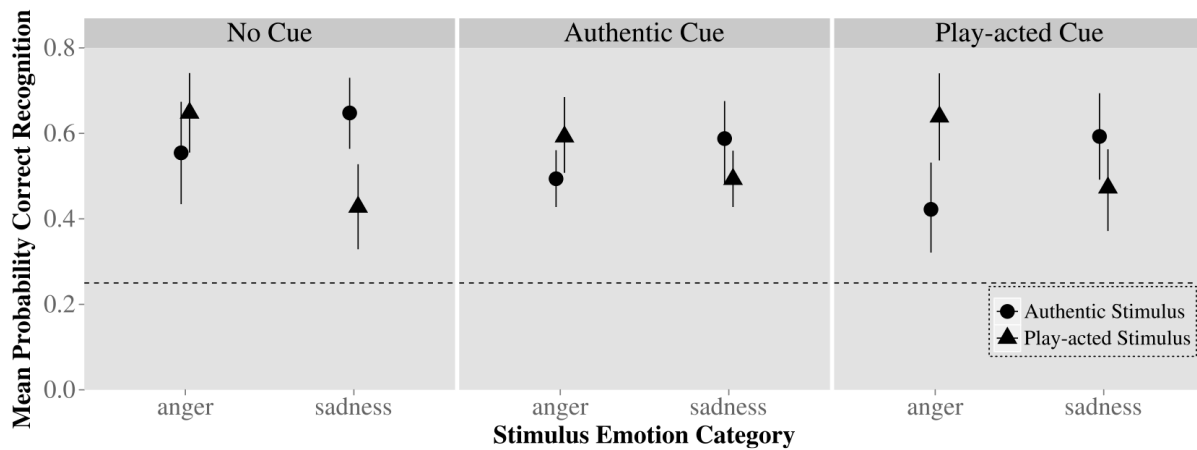


Fig. 4.2 Behavioral results from fMRI for judgments on anger and sadness stimuli, separated by cueing type. Actual emotion of recordings on x-axis, probability of recognition of emotion on y-axis. Chance level of recognition is indicated by the dotted line. Circles indicate recognition rates for authentic stimuli, while triangles indicate recognition rates for play-acted stimuli, along with 95% confidence intervals.

Table 4.1 Emotion categorization accuracy and reaction time for cue type

Recording Type		No Cue		Authentic Cue		Play-acted Cue	
		Recognition	RT	Recognition	RT	Recognition	RT
Authentic	Anger	0.56 ± 0.26	2.92 ± 0.55	0.49 ± 0.15	2.92 ± 0.51	0.42 ± 0.23	2.96 ± 0.60
	Fear	0.14 ± 0.14	3.06 ± 0.72	0.20 ± 0.13	3.14 ± 0.69	0.27 ± 0.15	3.17 ± 0.48
	Joy	0.35 ± 0.22	3.07 ± 0.50	0.40 ± 0.21	3.19 ± 0.68	0.30 ± 0.18	3.20 ± 0.70
	Sadness	0.65 ± 0.19	3.16 ± 0.62	0.59 ± 0.21	3.17 ± 0.75	0.59 ± 0.23	3.22 ± 0.73
Play-acted	Anger	0.65 ± 0.21	3.04 ± 0.72	0.59 ± 0.20	2.56 ± 0.57	0.64 ± 0.22	2.54 ± 0.34
	Fear	0.20 ± 0.21	2.48 ± 0.57	0.19 ± 0.13	3.23 ± 0.60	0.18 ± 0.13	3.35 ± 0.48
	Joy	0.32 ± 0.18	3.39 ± 0.51	0.32 ± 0.17	3.11 ± 0.61	0.33 ± 0.17	3.33 ± 0.54
	Sadness	0.43 ± 0.23	3.03 ± 0.65	0.49 ± 0.15	3.25 ± 0.73	0.47 ± 0.22	3.35 ± 0.56

Note: Recognition rates for emotions (Recognition) and reaction times (RT) for the respective categories. Rows split by actual stimulus authenticity and columns segregated by cue type. Emotion trials required a determination of the emotional content of the recordings.

To ensure that the current results are otherwise consistent with previous findings, effects of stimulus AUTHENTICITY on emotion categorization only (not split by CUEING category) were analyzed using the generalized linear mixed model. Overall, subjects detected anger correctly more often when play-acted than when authentic, whereas they detected sadness correctly more often when authentic than when play-acted (significant interaction EMOTION  $\times$  AUTHENTICITY ( $\chi^2(3)=28.63$ ,  $p<0.001$ ); Fig. 4.3). Subjects detected fear only near chance levels for both authentic and play-acted recordings. Post-hoc analyses indicated that recognition of joy was significantly better than fear ( $p<0.05$ ), but was not affected by AUTHENTICITY ( $p>0.10$ ).

We also modeled reaction times statistically to determine whether EMOTION, AUTHENTICITY, or CUEING influence behavior other than emotion categorization. Analyses indicated a significant main effect of stimulus EMOTION ( $\chi^2(3)=76.11$ ,  $p<0.01$ ), but none of stimulus AUTHENTICITY ( $\chi^2(1)=0.08$ ,  $p>0.10$ ) or CUEING ( $\chi^2(2)=2.02$ ,  $p>0.10$ ) on reaction times. Considering the average RTs that were measured from participant responses, it is unlikely that any differences in BOLD effects reported in the following were due to variability in participant concentration level or stimulus difficulty.

#### 4.4.2 FMRI results

First, stimulus AUTHENTICITY was contrasted only within uncued trials to determine the main effect of spontaneous modulation by AUTHENTICITY (whole-brain analysis). Activation loci with significantly greater activation during trials without CUEING [play-acted stimuli > authentic stimuli] included transverse temporal gyrus (TTG), posterior superior temporal sulcus (pSTS), temporoparietal junction (TPJ), fusiform gyrus, inferior frontal gyrus (IFG), superior frontal gyrus (SFG) and middle frontal gyrus (MFG) (Fig. 4.4, Table 4.2). The direct contrast cued versus uncued trials indicated no significant differences in activation between these two conditions.

The third analysis step examined the interaction effect between stimulus AUTHENTICITY and CUEING (small-volume analysis). As stated, this contrast applied a 2x2 design with stimulus authenticity versus cue validity to determine the effect of cueing interacting with stimulus authenticity over all trials. The stimulus by cue interaction contrast revealed activation in pSTS, TPJ and ACC (Fig. 4.5; Table 4.3).

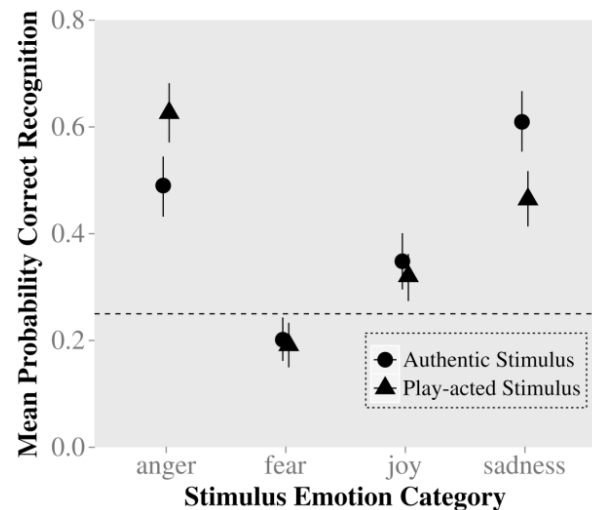


Fig. 4.3 Behavioral results from fMRI for emotion judgments. Actual emotion of recordings on x-axis, probability of recognition of emotion on y-axis. Chance level of recognition is indicated by the dotted line. Circles indicate recognition rates for authentic stimuli, while triangles indicate recognition rates for play-acted stimuli, along with 95% confidence intervals.

**No Cue: Play-acted > Authentic Stimuli**

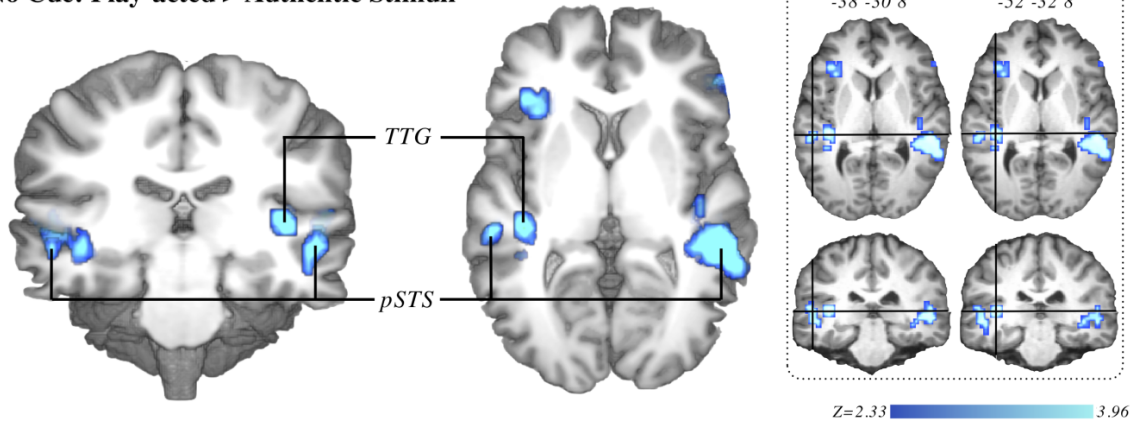
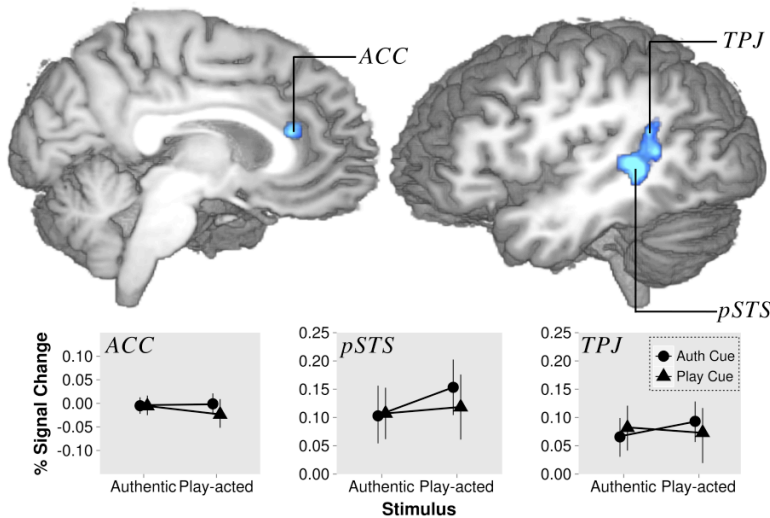


Fig. 4.4 Brain activation correlates of experimental tasks. Group-averaged ( $n = 18$ ) statistical maps of significantly activated areas for play-acted versus authentic stimuli. Corrected for false-positive results with an initial voxel-wise  $Z$ -threshold set to  $Z = 2.58$  ( $P < 0.01$ , two-sided), and cluster-size and cluster-value thresholds obtained by Monte Carlo simulations at a significance level of  $P < 0.05$ . Activation was mapped onto the best average subject 3-D anatomical map. Left: anterior view coronal section through TTG. Middle: top view sagittal section through TTG. Right: Two sections through TTG with crosshairs placed on pSTS (left) and TTG (right) in the left hemisphere. pSTS posterior superior temporal sulcus; TTG transverse temporal gyrus.

**A) Interaction: Cueing x Stimulus Authenticity**



**B) Play-Acted > Authentic Stimuli**

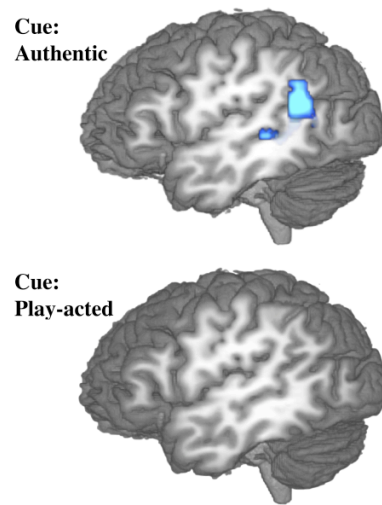


Fig. 4.5 Brain activation correlates of experimental tasks. Group-averaged ( $n = 18$ ) statistical maps of significantly activated areas. A) Interaction effects of cue and stimulus authenticity Signal change figures descriptively indicate the nature of the interaction based on the BOLD time-course from labeled regions in the z-map. B) Contrasts of play-acted versus authentic stimuli during trials cued as authentic (top) or play-acted (bottom) to determine significance of interaction sources.

Finally, to examine these interaction effects in more detail, authentic stimuli were contrasted against play-acted stimuli post-hoc, in trials with authentic and play-acted cues individually. For trials with authentic cues this revealed significantly increased activation for play-acted stimuli in temporoparietal junction (TPJ) and superior temporal sulcus (pSTS). Conversely, the same contrast performed for trials with only play-acted cueing did not show any significant activation (Fig. 4.5; Table

4.3). This empty z-map pointed to a relatively low power of the ACC effect revealed in the interaction of cue x stimulus authenticity. When inspecting this t-test with lowered statistical threshold, we found ACC to be (descriptively) more engaged for authentic as compared to play-acted stimuli when cueing was "play acted".

Table 4.2 Direct effects of stimulus authenticity (play-acted stimuli > authentic stimuli)

Area	BA	Hemisphere	Talairach coordinates			Z
			x	y	z	
<i>No Cue</i>						
TTG	41	L	-38	-31	6	4.06
IFG	45	R	49	32	0	3.42
Anterior Insula		L	-32	26	6	3.74
pSTS	21	R	49	-34	6	4.10
	22	L	-50	-28	0	3.63
Fusiform Gyrus	37	R	43	-49	-12	3.63
mSFG	8	R	4	17	51	3.85
	6	L	-5	-1	54	3.51
MFG	6	L	43	2	39	3.57

Anatomical specification, Talairach coordinates and maximum Z value of local maxima ( $p < 0.05$ , corrected). TTG transverse temporal gyrus; IFG inferior frontal gyrus; pSTS posterior superior temporal sulcus; mSFG medial superior frontal gyrus; MFG middle frontal gyrus.

Table 4.3 Effects of stimulus authenticity during cued trials

Area	BA	Hemisphere	Talairach coordinates			Z
			x	y	z	
<i>Cueing X Authenticity (Interaction)</i>						
pSTS	22	L	-47	-39	6	2.63
TPJ	39	L	-47	-52	12	3.08
ACC	24	R	7	26	18	3.48
<i>Play-acted Stimuli &gt; Authentic Stimuli</i>						
<i>Authentic Cue</i>						
pSTS	22	L	-47	-43	3	2.55
TPJ	39	L	-50	-55	23	3.08

Anatomical specification, Talairach coordinates and maximum Z value of local maxima ( $p < 0.05$ , corrected). TPJ temporoparietal junction; pSTS posterior superior temporal sulcus; ACC anterior cingulate cortex.

## 4.5 Discussion

In this fMRI study, we investigated whether information about authenticity of emotional speech influenced perception of the respective emotions in a top-down manner. We examined stimuli varying in their emotional content and authenticity for their bottom-up effects on behavior and brain activation, and assessed whether expectation (by cueing specific stimuli as either authentic or play-acted) had a top-down effect on emotion recognition and BOLD response. The behavioral data showed that the interaction effect of stimulus AUTHENTICITY and emotion recognition is equivalent to Drolet and colleagues (2012). While cueing stimuli either as authentic or play-acted did not influence emotion recognition itself, the results of the BOLD response analysis included both expected and surprising effects of CUEING. The lack of effects in the behavioral data makes a case for a purely bottom-up effect of authenticity on emotion recognition. While the bottom-up effects are clearly of more importance in these experimental tasks, the examination of brain activation correlates shows that additional information is not completely ignored, as discussed below.

A significant influence of AUTHENTICITY on brain activity was observed by contrasting authentic versus play-acted stimuli for trials without context cues (Fig. 4.4). This contrast yielded an increase in activity

in transverse temporal gyri (TTG) and posterior superior temporal sulcus (pSTS), extending into temporoparietal junction (TPJ) on the right for play-acted as opposed to authentic stimuli. The lack of modulation in the complete ToM network in this contrast was surprising. This difference to Drolet and colleagues (2012) indicates that the effects seen in that study were partly task-related. Therefore, the effect of stimulus AUTHENTICITY depends on individual expectation and focus, as modulated by task. This is discussed in more detail below. While there are hemispheric differences in the current results, these are difficult to interpret. First, hemispheric differences are an active area of research and several non-hypothesis driven reverse inferences could be valid in this context. Second, corrections for multiple comparisons required for fMRI BOLD analyses can exaggerate the influence of laterality differences, while such apparent differences represent non-statistical comparisons between p-values (differences between p-values may not be significant). Therefore, here these effects will not be discussed in detail.

The modulation in TTG indicates an influence on early cortical auditory processing (Arnott *et al.*, 2004), likely due to differences in acoustic features between authentic and play-acted stimuli (Jürgens *et al.*, 2011). Therefore, the direct influence of AUTHENTICITY on activation appears to be a spontaneous response to contextual differences between stimuli, particularly when subjects are not told about authenticity nor have to explicitly determine authenticity. As this effect is stimulus-based, it must be due to some acoustical difference between the categories of stimuli. Jürgens and colleagues (2011) found that a specific acoustic property, the contour or variability of the fundamental frequency (F0), is higher in play-acted than authentic stimuli. A future study would be needed to determine whether F0 contour indeed modulates activity in TTG in this case.

TPJ is heavily involved in the representation of mental states and beliefs (Samson *et al.*, 2004; Saxe, 2006). As such, both these regions are essential to the network activated for ToM tasks. Drolet and colleagues (2012) showed that TPJ is modulated by AUTHENTICITY when subjects categorize by authenticity explicitly, which requires perception of intention. Therefore, it is likely that the effect seen in this study is also related to representation of mental states (Samson *et al.*, 2004). While Beaucousin and colleagues (2007) found that left TPJ was not preferentially activated for emotional content, and other studies have found right TPJ to be more prominent in the perception of emotional prosody (Baum and Pell, 1999), left TPJ does appear to be relevant when a combination of mental state perception and emotion recognition is required, including potential context-relevant causes for those states, as hypothesized by Hervé and colleagues (2012).

Several studies have shown involvement of pSTS for mentalising about intentions in social situations (Frith and Frith, 2008) and representing human actions and relevant contexts (Saxe, 2006; Truett *et al.*, 2000). In particular, when contrasting tasks of cognitive versus affective ToM, Sebastian and

colleagues (2012) found that only pSTS, the temporal poles and precuneus were consistently activated for both, indicating a specific network overlap in these regions for ToM and affective perception. However, another potential influence must be considered, in particular due to the simultaneous modulation in TTG. While primary auditory processing occurs in TTG, further perception of lower-frequency activity, such as phonology or intonation, has been shown to correlate with activity in lateral temporal gyrus (Beaucousin *et al.*, 2007; Wildgruber *et al.*, 2005). Here the activation of pSTS may be related to the perception of lower-frequency changes. As mentioned above, Jürgens (2011) found that F0 contour, a low-frequency variation in intonation, shows the strongest correlation with AUTHENTICITY. While these two approaches to explaining pSTS activation may indicate some common process at work in both (Redcay, 2008), in this study they are not mutually exclusive. Differences in intonation may lead to initial differential activation bottom-up, activating pSTS even further when differences in intention or cause of action are perceived.

To examine top-down effects, trials were cued to examine whether the influence of explicitly stated authenticity, and hence expectation, would have a similar effect to those of actual AUTHENTICITY as discussed above. Cues were presented both congruently and incongruently. Cues appear to have had little effect on the categorization results, while they were nevertheless incorporated in stimulus perception and influenced BOLD response. To determine whether CUEING modulated brain activity during emotion perception, we examined the effect of the interaction of CUE and stimulus AUTHENTICITY. This contrast indicated that incongruent trials (authentic EMOTION following the announcement of play-acted EMOTION and vice versa) induced increased activation in pSTS and TPJ (Fig. 4.5A). This interaction effect was found to be due to authentically cued play-acted emotions as opposed to authentically cued authentic emotions, as shown in Figure 4.5B. In addition, the interaction contrast indicated increased activation in anterior cingulate cortex (ACC; Fig. 4.5A). Activation in ACC was also seen when contrasting authentic versus play-acted stimuli cued as play-acted, but this effect did not survive statistical correction and hence cannot be unambiguously interpreted at this point. However, based on its presence in the full interaction contrast, it can be suggested that ACC is slightly sensitive for incongruently cued emotional stimuli. This parallels studies that show that activation in this region correlates with stimulus conflict (MacDonald *et al.*, 2000) and uncertainty (Kéri *et al.*, 2004).

The interaction and post-hoc contrasts (Fig. 4.5) show that increased activation in TPJ and pSTS is due to a very specific set of circumstances. A prior belief of an authentic expression, with subsequent perception of a fake or play-acted one, recruits the area comprising TPJ and pSTS. As mentioned above, this stands in contrast to the ACC effect that did not reach significance in the t-test. This difference may indicate that an invalidly cued play-acted stimulus is much more salient than an invalidly cued authentic one (Fig. 4.5B). This salience may be due to the greater day-to-day relevance

of believed authentic expressions shown not to be, as in the case of deception detection (Burgoon *et al.*, 2008), which is known to recruit both TPJ and ACC (Grezes *et al.*, 2006, 2004). On the other hand, Drolet and colleagues (2012) found increased activation due to authentic stimuli during explicit perception of authenticity, and argued that this may be due to the greater importance of authentic emotional content. A similar effect could be involved in this case, with authentic cues perceived as more important in emotion perception. In either case, cues and stimulus features can induce activation in parts of the TOM network, and while this does not necessarily translate into changes in explicit emotion recognition, in a more natural day-to-day context such additional information may be integrated with the multitude of other contextual and multimodal information processed in parallel (Brück *et al.*, 2011; Rilliard *et al.*, 2009; Scherer and Ellgring, 2007). Regenbogen and colleagues (2012) showed that changes in the information content from individual modalities can affect both ToM network activation and various stages of emotion recognition, including empathy. In other words, the information from various sources, with explicit authenticity representing one such component, can affect brain activation individually, with the potential to affect behavior through stimulus and modality integration, even if this effect was not enough for the current experimental task.

Play-acted cues clearly induced less differential activation overall, indicating that this type of cue did not increase attention to the stimuli. Neither TTG, pSTS, nor TPJ were shown to be activated in this contrast. As opposed to the activation correlated with authentic cues, these trials did not induce greater attention nor increase mentalising or perspective-taking, either in the congruent or incongruent conditions. The difference between the influences of authentic and play-acted cues clearly shows that authentic cues influence perception more than play-acted cues, but also indicates how participants were able to ignore these cues and base their responses mostly on the perception of the stimuli.

The lack of TTG and pSTS modulation by AUTHENTICITY during cued trials was an additional unexpected finding. Activation in TTG and pSTS was induced by the stimuli spontaneously when no cue was given, but abolished when attention was directed to AUTHENTICITY. While it remains unclear why specifically this activation is affected, top-down influences on early auditory activation are known (Friederici, 2012; Todorovic *et al.*, 2011), and as unexpected stimulus properties can have an effect on BOLD response (Osnes *et al.*, 2012), the differences between authentic and play-acted stimuli may have a greater effect when participants are not explicitly told about AUTHENTICITY and are not expecting any contextual differences between stimuli.

Instruction differences would also explain why early auditory processing did not appear to be affected in Drolet and colleagues (2012), as participants were tasked with explicitly perceiving and



categorizing stimuli as either authentic or play-acted. This would have caused participants to expect differences in context between stimuli, and the explicit nature of the task additionally influenced perception during emotion categorization in that study. In particular, this also appears to reverse the direction or effect of stimulus authenticity. While an explicit discrimination of authenticity puts more weight on authentic stimuli (also increasing activation for such stimuli), spontaneous reaction to authenticity appears to be up-regulated by play-acted stimuli. While this, and the potential influence of acoustical properties, would have to be corroborated by further research, the current results show that TPJ recruitment by differences in AUTHENTICITY is spontaneous, while mPFC and retrosplenial cortex were not affected due to the lack of explicit authenticity recognition as in Drolet and colleagues (2012).

#### **4.6 Concluding Remarks**

While the interaction of emotion categorization and AUTHENTICITY is already known from our previous work, the current results indicate that this effect can occur automatically and much earlier in cortical processing than previously shown. Without explicit perception or CUEING, AUTHENTICITY modulates early stimulus-related activation and, potentially, already induces deeper cognitive processing of the nature of the stimuli. The acoustical differences due to authenticity and the increased activation induced by play-acted stimuli indicate that, in a bottom-up manner, increased F0 contour (indicative of play-acted stimuli) induces spontaneous up-regulation in the BOLD response for regions involved in speech processing for both early auditory processing and later phonology-related activation and mentalising. While an authentic cue leads subjects to reassess their perception of the presented stimulus in cases of conflict, a play-acted cue is perceived as less relevant and reduces differential activation overall and, when cue and stimulus are perceived to be in conflict, stimulus perception simply overrides the cue, as indicated by the behavioral data. Therefore, while explicit authenticity information does not appear to have an impact on explicit behavior as measured in this study, modulation in brain activation indicates that it may nevertheless modulate our perception of emotion in real-life contexts.



## Chapter 5

### Study IV: Recognizing the authenticity of emotional expressions: F0 contour matters when you need to know

Matthis Drolet<sup>1,2</sup>, Ricarda I. Schubotz<sup>2,3</sup> & Julia Fischer<sup>1</sup>

<sup>1</sup> Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany

<sup>2</sup> Biological Psychology, University of Münster, Münster, Germany

<sup>3</sup> Minerva Group Motor Cognition, Max Planck Institute for Neurological Research, Köln, Germany

Submitted

#### Author Contributions

MD designed the study; MD collected the data; MD analysed the data; MD, RS and JF discussed the results and commented on the manuscript

### 5.1 Abstract

Authenticity of vocal emotion expression affects emotion recognition and brain activity in the so-called ToM network. Exploiting fundamental frequency contour (F0), which varies more in play-acted expressions than authentic ones, we examined whether contour affects not only authenticity but also emotion recognition. Moreover, we tested whether contour alters ToM activation and explored the role of task as a top-down modulator. The effects of contour on BOLD signal were analyzed by contrasting high and low contour stimuli within two fMRI studies implementing emotion and authenticity judgment tasks. Higher contour was found to up-regulate activation task-independently in the primary auditory cortex. Stimulus contour and task were found to interact in a network including medial prefrontal cortex, with an increase in BOLD signal for low-contour stimuli during explicit perception of authenticity and an increase for high-contour stimuli during explicit perception of emotion. Contour-induced BOLD effects appear to be purely stimulus-driven in early auditory and intonation perception, while being strongly task-dependent in regions involved in higher cognition.

## 5.2 Introduction

Emotion is one of the most salient categories of social stimuli. External expressions of emotion can be perceived from body language (Van den Stock *et al.*, 2007), the face (DeKosky *et al.*, 1980; Ekman *et al.*, 1990), or the voice (Scherer, 1991; Wurm *et al.*, 2001). These processes are now known to be strongly context-dependent, such that the interpretation of an emotional expression is influenced by the situation and environment in which that expression is produced or perceived (Barrett and Kensinger, 2010; Brück *et al.*, 2011; Schirmer *et al.*, 2006). Among these factors, a particularly relevant contextual modulator of emotion processing is recognition of speaker intention through authenticity of expression (Drolet *et al.*, 2012, 2013).

The ability to perceive another's intention is commonly called Theory of Mind (ToM; Premack and Woodruff, 1978), defined as the implicit or explicit attribution of mental states (e.g., desires, beliefs, and intentions) to others and self (Frith and Frith, 2012; Frith and Singer, 2008). ToM has been studied extensively, including its evolutionary roots (Povinelli and Preuss, 1995), its development (Rakoczy and Tomasello, 2006; Wimmer and Perner, 1983), and its every-day use in humans (Abraham *et al.*, 2008; Tomasello, 2003). For instance, German *et al.* (2004) found that differences in the intention of pretend play can modulate brain activity implicitly, indicating that context and intention of expressions are automatically decoded. Importantly, ToM has been shown to both interact with emotion perception and be influenced by it (Mier *et al.*, 2010; Ray and Zald, 2012). In particular, Blair and colleagues (2007) showed that active regulation of emotional distractors is required in cognitive tasks such as ToM. The influence of ToM on emotion recognition, however, has not been studied extensively.

Our previous work (Drolet *et al.*, 2012) showed that emotional authenticity affects emotion perception. Recordings of speech produced by professional actors after instruction to express a specific emotion, described as play-acted, were compared to speech produced without external instruction, described as authentic. The differences between these categories involve differences in speaker intention, so that their perception is an ideal case of the interaction between emotion perception and ToM. Explicit rating by authenticity induced BOLD response modulation in the ToM network (medial prefrontal, retrosplenial and temporoparietal cortices) more so than did emotion categorization, and authentic stimuli additionally up-regulated activation in an important component of the ToM network, medial prefrontal cortex (medial PFC). A subsequent study using these stimuli (Drolet *et al.*, 2013) required only an emotion categorization task. In that study, instead of authentic stimuli up-regulating medial PFC, play-acted stimuli up-regulated temporoparietal junction (TPJ), early auditory processing in transverse temporal gyrus (TTG), and early sentence perception in middle and

superior temporal gyri (MTG, STG). It therefore appeared that emotional authenticity affects brain activation task-dependently.

While these effects in early auditory as well as higher association cortices indicate bottom-up and top-down processes at work in the perception of authenticity of emotional content, some specific questions remained. First, while emotion is known to affect vocal expressions (Banse and Scherer, 1996; Williams and Stevens, 1972), the actual acoustic properties that led to the aforementioned bottom-up effects of authenticity were unclear. Jürgens and colleagues examined the acoustic correlates of authenticity and found that contour (variability of the fundamental frequency or F0) was significantly greater for play-acted recordings than for authentic. The variability in fundamental frequency is measured across the entire span of the vocal expression (see Fig. 5.1). So someone hearing an expression with higher variability may consider it more likely to be play-acted, while one with lower variability would be more likely to be authentic. Wiethoff et al. and colleagues (2008) showed that several acoustic parameters, including contour, correlated with activation in right superior temporal gyrus (STG), while a recent review by Vigneau and colleagues (2011) found that several superior temporal and lateral frontal areas are activated by phonological processing.

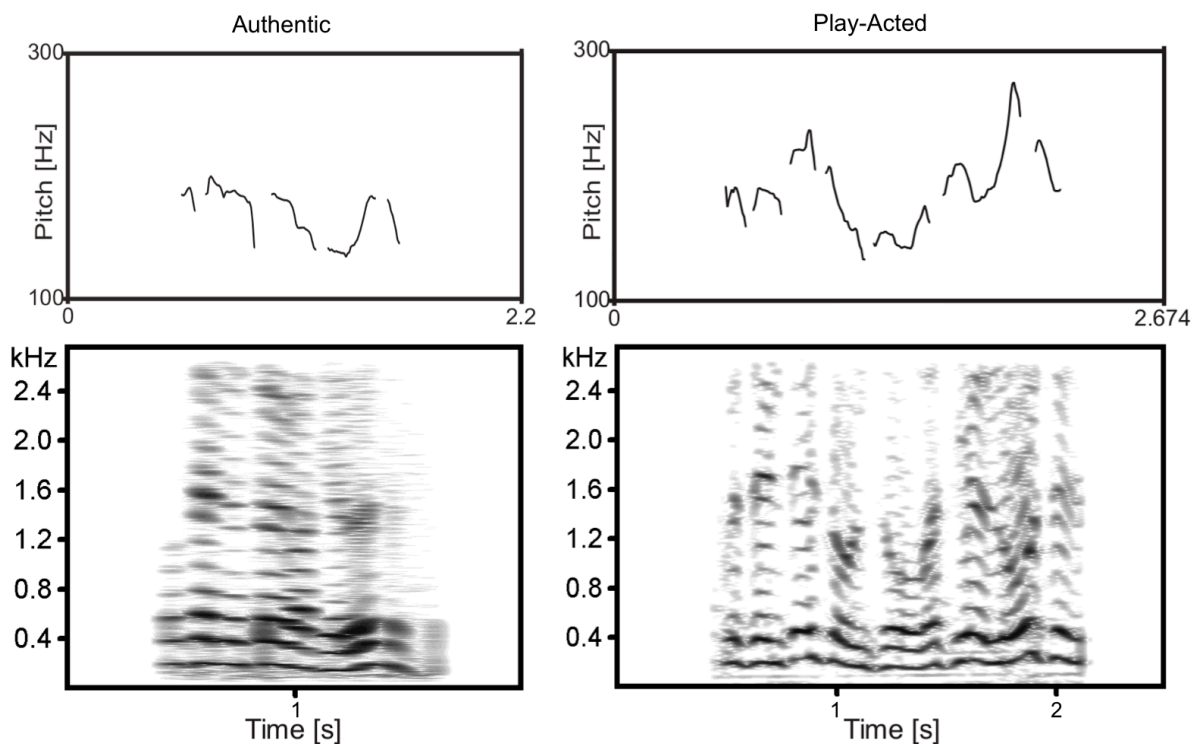


Fig. 5.1 F0 contour (top) and spectrogram (bottom) of two complementary example recordings. Examples are of the same text (“... Präsenz ist und so, was hier los ist...” translated: “...presence is and such, what is going on here...”) from the original authentic and play-acted versions.

Against this backdrop, we wished to determine whether F0 contour affects behavior and BOLD response when listening to emotional prosody and whether it is the acoustic variable that is

responsible for the authenticity effects seen in previous studies (Drolet *et al.*, 2012, 2013). Three main questions were addressed: (1) whether F0 contour influences the recognition of authenticity, (2) whether it influences early auditory and intonation processing and the ToM network during explicit categorization, and (3) whether it affects TPJ and medial PFC modulation differently depending on whether participants are rating emotional content or authenticity, due to the increase in BOLD signal for authentic stimuli during explicit recognition of authenticity and for play-acted stimuli during explicit recognition of emotion.

Subsequently we examined the influence of F0 contour on behavioral responses parametrically. Based on previous findings (Drolet *et al.*, 2012), low contour stimuli were predicted to be preferentially rated as sad or authentic and high contour stimuli as anger or play-acted. Regarding fMRI, three hypotheses were tested. First, the effect of contour was analyzed parametrically to determine what, if any, region responded to the entire measurable span of contour. Based on differing contour between play-acted and authentic stimuli (Jürgens *et al.*, 2011) and up-regulation in the primary auditory cortex by play-acted stimuli (Drolet *et al.*, 2012), we expected stimuli with higher contour to induce higher activation in primary auditory cortex, superior temporal gyrus and superior temporal sulcus. Second, BOLD effects of extremely high and low contour values were identified to directly compare recordings that, based on contour, have a high probability of being either authentic or play-acted. Based on our previous studies and Vigneau and colleagues (2011) and Wiethoff and colleagues (2008), STG, MTG, TPJ, medial PFC and lateral frontal areas were predicted to increase activation with increasing contour. Third, regions of interest (ROI) were examined in the previous contrast to determine whether task-requirements modulate the effect contour has on the BOLD response. ROIs were extracted from medial PFC and TPJ based on (Drolet *et al.*, 2013) with only the emotion task (i.e. with a relatively high power in that one task). These ROIs were then examined separately for the two tasks to examine the interaction of bottom-up contour effects and top-down task effects. Based on previous results we expected authentic stimuli to up-regulate BOLD during authenticity tasks, and play-acted stimuli to do so during emotion tasks.

## **5.3 Methods**

### *5.3.1 Stimulus Selection*

Original recordings (mono wave format; sample rate of 44.1kHz) were selected from German radio interviews of individuals talking in an emotional fashion (anger, fear, joy, sadness) about a highly charged ongoing or recollected event (e.g. the death of a child, winning a lottery, threatened by a current danger). Emotion was ascertained through verbal content and recording summaries. Staged and scripted settings were excluded. Of 80 speech segments, 35 were made outdoors but were of

good quality with minimal background noise. To ensure inference-free verbal content, text-only transcripts were rated by naïve subjects. Recordings with emotion recognized better than chance were replaced to ensure neutral semantic content. The original set consisted of 80 recordings by 78 speakers (half male, half female; mean  $1.75s \pm 1.00$  SD; range 0.36–4.06s).

Play-acted stimuli were performed by actors from Germany (42 actors each replicated a maximum of 3 recordings of equivalent emotional content), who were told to express each text in their own way, using the transcripts, summaries, and emotion (stimulus segments were not indicated and actors never heard the originals). Recording environment was varied while minimizing background noise, with 30 of 80 made outdoors (mean  $1.76s \pm 1.02$  SD; range 0.38–4.84s). Average amplitudes of all stimuli were equalized with Avisoft SASLab Pro Recorder 4.40 (Berlin, Germany). The final stimulus set consisted of 20 samples of joy and sadness, 22 samples of anger and 18 samples of fear, both for authentic and play-acted sets.

### 5.3.2 Participants

Study 1: 24 female participants (mean 24 years old; range 20-30 years; right-handed; German mother-tongue), without a history of neurological or psychological complications (including the use of psychiatric medication), were selected and contacted using the Cologne MPI database for fMRI experiments. Participants were informed about the potential risks of magnetic resonance imaging and screened by a physician. They gave informed consent before participating and were paid afterwards. The experimental standards were approved by the local ethics committee and data were handled pseudonymously. Study 2: Selection criteria were identical to Experiment 1, with 18 female participants selected (20-30 years, mean 24 years, right-handed, German mother-tongue).

### 5.3.3 Trial and Stimulus Presentation

Study 1: The program NBS Presentation (Neurobehavioral Systems, Inc., Albany, California) controlled the trial structure, timing and order of each experimental run. Each run (one per participant) included 178 trials, of which 72 were used for an emotion judgment task and 72 for an authenticity judgment task. In addition, 2 control tasks were included: 16 word detection trials, in which participants had to count occurrences of the word “und” (“and”), and 18 empty trials with pink-noise playback. For emotion judgments four responses were possible: anger, sadness, happiness, fear (presented in German as: “Wut”, “Trauer”, “Freude”, “Angst”), while for authenticity judgment responses were authentic (“echt”) and play-acted (“theater”; described to participants beforehand as “gespielt”, i.e. play-acted). To minimize eye movement the maximal line-of-sight angle for visual information was kept under 5 degrees. Trial type and stimulus type pseudo-randomizations were performed using conan (UNIX shell script: Max-Planck Institute (MPI) for Neurology in Leipzig, Germany) to reduce any systematic effects that could have otherwise occurred with simple randomization. Each participant



was shown a button sequence on-screen (800x600 pixel video goggles: NordicNeuroLab, Bergen, Norway) complementary to the response box layout (10x15x5 cm gray plastic box with a row of 4 black plastic buttons). For both emotion judgment and word detection all buttons were assigned a possible response. For authenticity judgment only the 2 leftmost buttons were used. Study 2: Software, hardware, and trial and stimulus randomization were identical to Experiment 1, but only the emotion judgment task was applied (144 trials). One-third ( $n=48$ ) were not cued (no authenticity information was provided), one-third were cued as authentic, and one-third were cued as play-acted. Cueing was congruent half the time and was presented above the response options as authentic or play-acted (“echt” or “spiel” respectively). The remaining 30 trials were used to implement 2 independent control tasks: 18 empty trials with pink-noise playback and 16 age task trials in which participants had to determine the age of the speaker (20, 30, 40 or 50).

#### 5.3.4 Experimental Procedure

Study 1: Participants were fitted with headphones for audio playbacks (NNL: NordicNeuroLab, Bergen, Norway) after they were placed in a supine position on the fMRI table. Imaging was performed with a 3T Siemens MAGNETOM TrioTim (Cologne, Germany) system equipped with a standard birdcage head coil. Participants were placed with their four fingers (excluding thumb) positioned on the response buttons of the response box. Formfitting cushions were used to prevent head, arm, and hand movements. Twenty-two axial slices (210 mm field of view; 64 x 64 pixel matrix; 4 mm thickness; 1 mm spacing; in-plane resolution of 3 x 3 mm) parallel to the bicommissural line (AC–PC) and covering the whole brain were acquired using a single-shot gradient EPI sequence (2000 msec repetition time; 30 msec echo time; 90° flip angle; 1.8 kHz acquisition bandwidth) sensitive to BOLD contrast. In addition to functional imaging, 22 anatomical T1-weighted MDEFT images (Norris, 2000; Ugurbil *et al.*, 1993) were acquired. In a separate session, high-resolution whole-brain images were acquired from each participant to improve the localization of activation foci using a T1-weighted 3D-segmented MDEFT sequence covering the whole brain. Functional data were mapped onto this 3D average using the 2D anatomical images made immediately following the experiment. Including a visual and auditory test before the experiment, one experimental run lasted approximately 45 minutes. Study 2: The general procedure was identical to experiment 1. Twenty-four axial slices were acquired using a single-shot gradient EPI sequence sensitive to BOLD contrast.

#### 5.3.5 Behavioral Statistics

To determine whether contour affects participant responses, the responses to both emotion and authenticity tasks from study 1 were examined as to whether labeled recordings differed significantly in contour values. The generalized linear model was implemented (R Statistical Package v2.15; R Development Core Team, 2008) to determine the best model fit for response rates with the glmer

function from the lme4 package using restricted maximum likelihood estimation with binomial error structure and logit link function (Baayen *et al.*, 2008). The basic model examined the effect contour has on responses with participant included as a random factor [glmer (Response~Contour+(1|Subject),family=binomial,REML=FALSE)]. Post-hoc statistics were applied using a likelihood ratio test (LRT) with a Chi-squared distribution ( $\chi^2$ ) to examine the effect on each emotion and were corrected for multiple comparisons.

### 5.3.6 Functional MRI Statistics

After motion correction using Siemens rigid-body registration protocol (München, Germany), the functional data were processed using the software package LIPSIA v1.5.0 (Lohmann *et al.*, 2001). This software package is available under the GNU General Public License ([www.cbs.mpg.de/institute/](http://www.cbs.mpg.de/institute/)). To correct for temporal offset between the slices acquired in one image a cubic-spline interpolation was employed. Low-frequency signal changes and baseline drifts were removed using a temporal high-pass filter set for each scanned participant dependent on the pseudo-randomized design (filter frequency range: 1/75-85 Hz). Spatial smoothing was performed with a Gaussian filter of 5.65 mm FWHM ( $\sigma = 2.4$ ). To align the functional data slices with a 3D stereotactic coordinate reference system, a rigid linear registration with six degrees of freedom (three rotational, three translational) was applied. The rotational and translational parameters were acquired on the basis of the MDEFT slices to achieve an optimal match between these slices and the individual 3D reference dataset. The MDEFT volume dataset with 160 slices and 1-mm slice thickness was standardized to the Talairach stereotactic space (Talairach and Tournoux, 1988). The rotational and translational parameters were subsequently transformed by linear scaling to a standard size. The resulting parameters were then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system, thus generating output data with a spatial resolution of 3 x 3 x 3 mm (27 mm<sup>3</sup>).

Two design matrices were applied. In the first, contour (variation in fundamental frequency; measured as standard deviation) was modeled parametrically to examine its correlation with brain activation. The design matrix contained authentic and play-acted stimulus trials in each of the first two event types respectively, with an amplitude vector of one. The second and third event types each also contained authentic and play-acted stimulus trials with an amplitude vector corresponding to the specific stimulus's contour value. The last event type in the design matrix, null-events, was assigned an amplitude value of one. This analysis was performed for the data from Drolet and colleagues (2012 [with authenticity and emotion task]) and Drolet and colleagues (2013 [only with emotion task]).

To further substantiate the parametric analysis, an additional step was taken. Using a second design, stimuli with extreme values of contour were contrasted directly. To do so, trials were preselected based on their respective recording's contour value and grouped as high or low contour trials within the design matrix. Since we know authenticity correlates with contour (Jürgens *et al.*, 2011), a simple block selection of upper and lower contour stimuli would lead to selections with unequal numbers of, and unbalanced mean contour values for, authentic and play-acted stimulus categories. Therefore, individual stimuli were excluded from or included in each group to ensure equal numbers of authentic and play-acted stimuli (15 of each) and equivalent average contour values (Table 5.1). High and low contour trial groups were contrasted for emotion task and authenticity task trials individually (trials outside the low and high contour groups were included as a single regressor of no interest). Subsequently, a conjunction of these two contrasts was used to examine regions activated in both tasks, while an exclusive disjunction was applied to uncover regions more highly activated in one task but not in the other.

Table 5.1 Mean values of high and low stimulus groups by contour (standard deviation of F0)

		Authentic (Hz)	Play-acted (Hz)	Statistics
Study 1	Low Contour	7.03 ± 3.14	8.53 ± 3.00	t = 2.07, p > 0.2
	High Contour	48.44 ± 13.83	56.93 ± 22.39	t = 2.04, p > 0.1
Study 2	Low Contour	6.78 ± 3.00	8.27 ± 2.99	t = 2.06, p > 0.2
	High Contour	48.89 ± 15.24	56.36 ± 20.75	t = 2.05, p > 0.1

*Mean contour (measured as standard deviation of fundamental frequency) and standard deviation of the selected high and low value recordings for each authenticity category and study. Statistics represent t-tests of the differences between the contour values by authenticity as applied to BOLD contrasts.*

Statistical evaluation was based on a least-squares estimation using the general linear model for serially auto-correlated observations (Friston *et al.*, 1998; Worsley and Friston, 1995). Both designs were generated with a delta function, convolved with the hemodynamic response function (gamma function). Each trial in the design matrix was identified by its onset time and stimulus length, while speaker repetition was included to prevent this from influencing the statistical analysis. Brain activations were analyzed time-locked to recording onset and the analyzed epoch was set individually for each trial to the duration of the respective stimulus. The model equation, including the observation data, design matrix, and error term, was convolved with a Gaussian kernel of dispersion 5.65 sec FWHM to account for temporal autocorrelation (Worsley and Friston, 1995). In the following, contrast images (i.e. beta value estimates of the raw-score differences between specified conditions) were generated for each participant. As all individual functional datasets were aligned to the same stereotactic reference space, the single-subject contrast images were entered into a second-level Bayesian analysis (Neumann and Lohmann, 2003). In the approach by Neumann and Lohmann

(2003), posterior probability maps and maps of the effect size for the effects of interest in groups of subjects are calculated based on the resulting least-squares estimates of parameters for the GLM. The output of the Bayesian second-level analysis is a probability map showing the probability for the contrast to be larger than zero. For visualization and listed activation peaks a threshold of 99.5% and minimum cluster size of 100 voxels was applied to the probability maps.

Finally, the ROI analysis was performed on the high versus low contour contrast to determine the relative influence of contour dependent on task instruction. ROIs were selected from medial PFC (x,y,z of Talairach space: 1,44,22; -2,47,24) bilaterally, and TPJ (-53,-52,32), within activation clusters in the second study that implemented only an emotion task (i.e. with a relatively high power in that one task). Activation was then extracted from these ROIs for the two tasks in the first study separately. The generalized least-squares model was implemented (R Statistical Package v2.15; R Development Core Team, 2008) to determine the best model fit for recognition rates with the gls function from the nlme package using restricted maximum likelihood estimation and compound symmetry for paired data (Pinheiro and Bates, 2009). The model examined the effects contour and task have on BOLD response (GLM beta values) with participant included as correlation factor: [gls (BOLD ~ Contour \* Task, correlation= corCompSymm (form =~ 1|Subject), data = D)]. Results are indicated using the interaction effect degrees of freedom, t-value, p-value, and Spearman's rank correlation coefficient (rho).

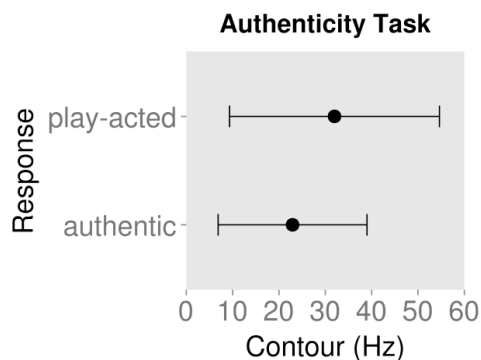


Fig. 5.2 Mean contour (measured as standard deviation of fundamental frequency) and standard deviation for responses to recognition of auth-enticity task.

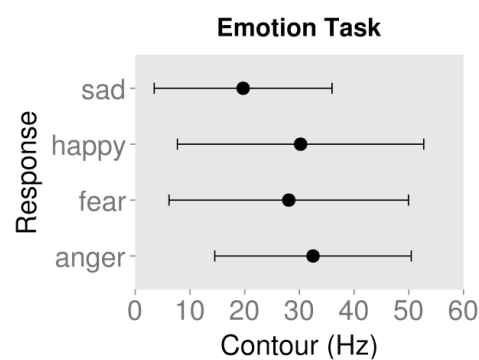


Fig. 5.3 Mean contour (measured as standard deviation of fundamental frequency) and standard deviation for responses to recognition of emotion task.

### 5.4 Results

Figures 5.2 and 5.3 show the mean ± standard deviation of contour values for each response type for the emotion and authenticity tasks respectively. Examining the behavioral effects of contour indicated that contour values for “authentic” responses were significantly lower than for “play-acted” responses ( $\chi^2(1)=88.3$ ,  $p<0.001$ ; Fig. 5.2). In addition, contour values for “sad” responses were

significantly different from “anger” ( $\chi^2(1)=768.0, p<0.001$ ), “fear” ( $\chi^2(1)=1037.9, p<0.001$ ), and “joy” ( $\chi^2(1)=1089.6, p<0.001$ ), with stimuli labeled as “sad” being significantly lower in contour than the other emotions (Fig. 5.3).

The parametric analysis of brain activation with contour value was performed on the data for both studies. The parametric analysis of the first study was performed both across and split by task type, however, neither of these contrasts indicated any significant activation. Within the second study (which implemented only the emotion task), this analysis did produce significant activation in TTG [-47,-16,6] and inferior frontal sulcus [IFS: -41,-1,39; -38,23,24] (Fig. 5.4).

Due to reduced contour effects for the emotion task in study 1, whole-brain contour-induced activations during authenticity and emotion trial could only be compared between studies 1 and 2 (not within study 1). Both a conjunction and an exclusive disjunction were performed on these two contrasts. The conjunction (identifying regions similarly modulated in both tasks) revealed activation in the precentral (PrG) and postcentral gyri (PoG) on the left, the dorsal anterior cingulate cortex bilaterally (dACC) extending into the left pre-supplementary motor area (pre-SMA), the left transverse temporal gyrus (TTG), the left inferior occipital gyrus (IOG), and the lateral hemispheres of the cerebellum. The disjunction (regions differentially modulated in either task) revealed activation in the medial prefrontal cortex bilaterally (dmPFC) extending into anterior cingulate cortex (ACC) bilaterally, the left middle frontal cortex (MFG), the left inferior parietal lobule (IPL), the right inferior parietal sulcus (IPS), and left middle temporal gyrus (MTG) (Table 5.2; Fig. 5.5).

Table 5.2 Effects of stimulus authenticity during cued trials

Area	BA	Hemisphere	Talairach coordinates		
			x	y	z
Conjunction					
SMA	6	L	-5	6	48
dACC	32	L	-8	21	36
	32	R	1	21	36
PrG	6	L	-29	-9	54
PoG	3	L	-41	-28	51
TTG	42	L	-53	-15	6
IOG	19	L	-47	-78	0
	37	L	-44	-63	-3
Cerebellum	Crus I	L	-41	-54	-21
Cerebellum	V	R	13	-51	-12
Disjunction					
mPFC	10	L	-11	47	12
	9	R	1	41	24
ACC	32	L	-5	38	21
	32	R	4	41	3
MFG	8	L	-38	24	45
IPL	40	L	-56	-48	42
IPS	40	R	49	-36	42
MTG	21	L	-65	-24	-9

*Anatomical specification, Brodmann area (BA) Talairach coordinates and maximum Z value of local maxima (99.5% Bayesian probability). aMCC anterior middle cingulate; MFG middle frontal gyrus; SPL superior parietal lobule; IPL inferior parietal lobule; FG fusiform gyrus; IOG inferior occipital gyrus; sgACC subgenual anterior cingulate; mPFC medial prefrontal; SFG superior frontal gyrus; STG superior temporal gyrus; TTG transverse temporal gyrus; pSTS posterior superior temporal sulcus; TPJ temporoparietal junction; pSTS posterior superior temporal sulcus; TPJ temporoparietal*

**Parametric Analysis of Contour**

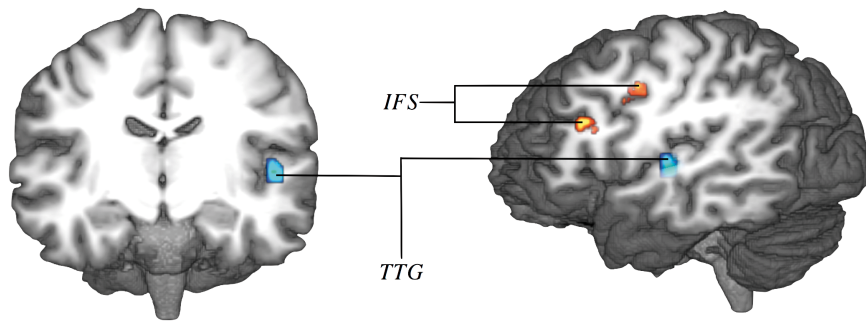


Fig. 5.4 Brain activation correlates of emotion experimental task of study 2 (Drolet et al., 2013). Group-averaged ( $n = 18$ ) statistical maps of significantly activated areas for parametric effect of contour (higher contour in blue, lower contour in red). Posterior probability maps with a threshold of 99.5% and 100 voxel minimum size. Activation was mapped onto the best average fit subject 3-D anatomical map. Left: anterior coronal section through middle temporal lobe. Right: left view sagittal section through temporal lobe. IFS inferior frontal sulcus; TTG transverse temporal gyrus.

**Low > High Contour (Authenticity Task & Emotion Task)**

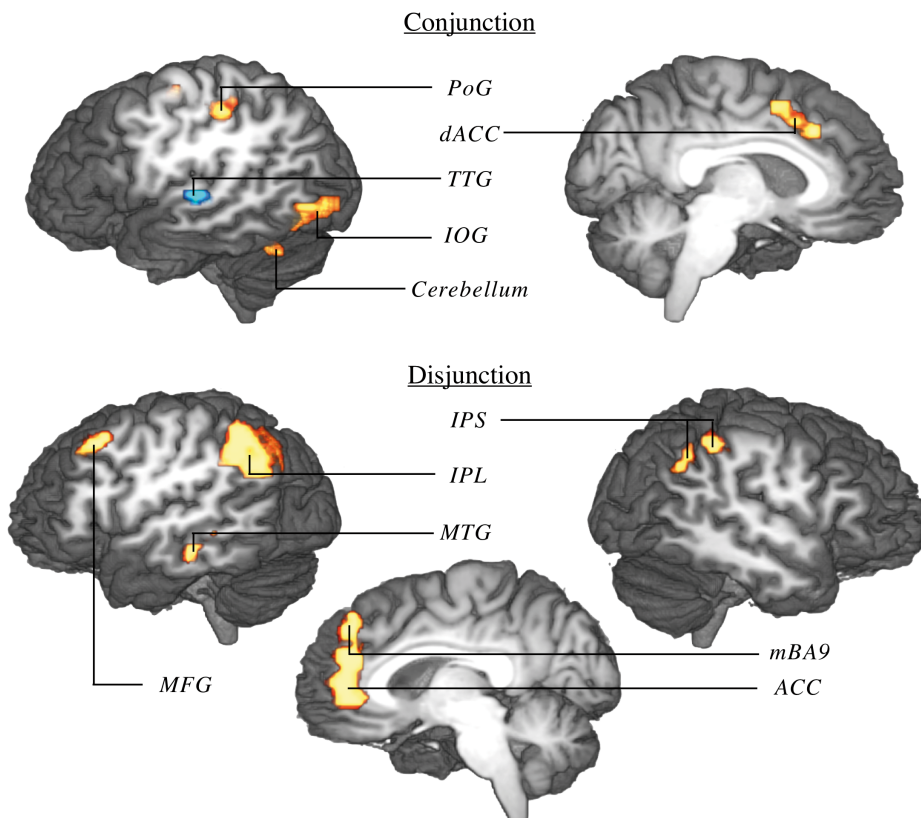


Fig. 5.5 Brain activation correlates of emotion and authenticity experimental tasks. Group-averaged ( $n = 18$ ) statistical maps of significantly activated areas for parametric effect of contour (higher contour in blue, lower contour in red). Posterior probability maps with a threshold of 99.5% and 100 voxel minimum size. Activation was mapped onto the best average fit subject 3-D anatomical map. Top: left view sagittal section. Bottom: left view sagittal section through midline. PoG postcentral gyrus; dACC dorsal anterior cingulate cortex; TTG transverse temporal gyrus; IOG inferior occipital gyrus; IPS inferior parietal sulcus; IPL inferior parietal lobule; MTG middle temporal gyrus; MFG middle frontal gyrus.

Finally, the ROI analysis performed on the data from the first study, split by task, using coordinates extracted from the second study, showed a clear interaction effect of task and contour. While activation in TTG and STG remained the same (as indicated above), bilateral activation in medial PFC appeared to reverse between the two tasks. Of these tendencies, the activation in right medial PFC reached significance, such that the interaction of contour and task was significant ( $t(84)=3.26$ ,  $p<0.01$ ,  $\rho=0.604$ ; Fig. 5.6).

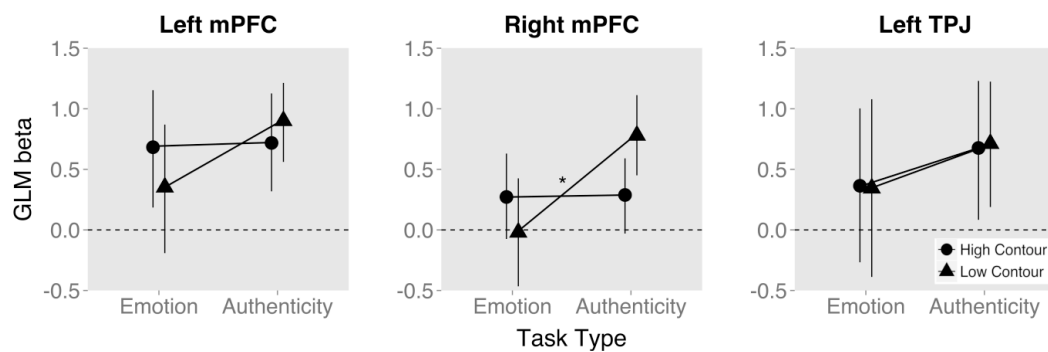


Fig. 5.6 Mean and 95% confidence interval of BOLD signal measure (GLM beta value) by task type and trial stimulus fundamental frequency contour value. Measures extracted from three regions of interest within study 1 selected from study 2.

## 5.5 Discussion

The purpose of this study was to examine whether the variance in the fundamental frequency (F0 contour) during emotionally expressive speech affects behavior and brain activation when listening to such stimuli, and whether F0 contour is the acoustic variable responsible for the authenticity effects seen in previous studies (Drolet *et al.*, 2012). In addition, we were interested in whether the expected F0 contour modulation of brain activation is task-dependent, such that low and high contour may each be selectively relevant for specific tasks, based on differential mPFC activation in studies 1 and 2.

Independent of the task, high contour stimuli (which are more likely to be play-acted than authentic) induced BOLD modulation in the left primary auditory cortex. This finding suggests that contour has a strong effect on activation even at a very early stage in perception and is the source of the authenticity effects seen in the primary auditory cortex previously (Drolet *et al.*, 2013). The primary auditory cortex is the first part of the cortex in which acoustic differences modulate activation (Arnott *et al.*, 2004; Celesia, 1976) and is even active for simple forms of auditory information (Binder *et al.*, 2000). In particular, frequencies are mapped tonotopically in TTG (Pantev *et al.*, 1995; Romani *et al.*, 1982). Thus, a greater variability of F0 frequencies stimulates more neurons, leading to greater BOLD response. The behavioral results demonstrate that contour also affects responses in both experimental tasks. While the range of contour values is quite large and their corresponding effects are subtle, “authentic” and “sad” responses were associated with lower contour values than “play-

acted” or all other emotions, respectively. The early integration of contour as seen through BOLD modulation in early auditory processing can, therefore, affect task-related behavior.

On the other extreme, task-independent activation by low contour stimuli (which are more likely to be authentic than play acted) included dorsal ACC, extending into pre-SMA. While activation in ACC had been seen in previous work (Drolet *et al.*, 2012, 2013), the cingulate cortex had not specifically been discussed in association with low contour stimuli. The current results may be due to this region’s involvement as an interface between emotion and cognition (Allman *et al.*, 2001). Wittforth and colleagues found that this region is activated more for emotional content incongruent to expectation. As discussed above, the recordings used here represent an overlap of both emotive and cognitive factors in the form of emotional prosodic content and authenticity and the differences in authenticity may be perceived as incongruence to the expectation of authentic emotion.

However, this does not explain why lower contour would preferentially recruit this region. While high contour stimuli are more likely to be play-acted than low contour stimuli are to be authentic, the difference between the lowest contour values of authentic and play-acted stimuli is only 1.31 Hz, while the difference between the highest values is 22.96 Hz (data from Jürgens *et al.*, 2011). As such, it is not only the absolute value of mean contour between authentic and play-acted stimuli that differs, but also how distinct the contour distribution is. Activation in ACC may, therefore, increase for low contour stimuli due to the greater ambiguity of authentic and play-acted stimuli at lower contour values.

While ACC activation was previously attributed to general response conflict (Ridderinkhof *et al.*, 2004), it now appears that ACC activation is more specific. Difficult tasks recruit the posterior medial PFC overall. This activation can be segregated anatomically into medial BA 8, supplementary motor area (SMA/BA6), and ACC (BA24/32). Volz and colleagues (Volz *et al.*, 2005) suggested that while medial BA 8 is recruited in the presence of complex tasks and difficult rules related to task completion, activation in medial SMA and dorsal ACC are related to difficulty in distinguishing stimuli while making a selection in the presence of clear task requirements. This differentiation has been termed decision conflict versus response conflict, respectively. Pochon and colleagues (Pochon *et al.*, 2008) subsequently suggested that while activation in medial SMA is related to competing motor responses, posterior ACC is rather activated for difficulty categorizing stimuli according to task (but clear) rules and, thereby, the integration of personal preferences. Such decision conflict is generally associated with ambiguous stimuli (Braver *et al.*, 2001; Carter and van Veen, 2007), as the emotion stimuli used here. In addition, the combined network of activation of ACC with SMA/PrG and PoG/IOG is during top-down attention control (Hopfinger *et al.*, 2000). Therefore, low-contour stimuli



required greater attention and perceptual focus than high-contour stimuli in order to respond because the authenticity of low-contour stimuli is more ambiguous and difficult to categorize.

The task-based conjunction also revealed that for both emotion and authenticity judgments activation increased in the cerebellum. Of that activation, the localization in the right cerebellum is related to motor activation in the PrG, indirectly associated with the response conflict apparent in the cerebral activation, with lateralization appropriate for the exclusively right-handed response task. Up-regulation of Crus I in the left cerebellum, however, is likely not related to motor activation. The posterior cerebellar lobes are involved in both cognitive and emotive processes (Schmahmann and Caplan, 2006), including being necessary for extended verbal working memory (Ravizza *et al.*, 2006). As Desmond and colleagues (1997) originally hypothesized, this part of the cerebellum could be recruited for expectancy-induced conflict related to phonological processing (though this is in no way its exclusive role), so that Crus I activation may increase due to the need to hold prosodic information in working memory while processing the more ambiguous low contour stimuli.

The disjunction, on the other hand, showed that medial BA 9 (bilaterally extending into anterior ACC) was differentially activated between the tasks. Activation increased for low contour values during the authenticity task, while they did so for high contour values during the emotion task of study 2. Activation in medial PFC and TPJ was previously seen in study 1 (Drolet *et al.*, 2012) as part of the ToM network (Frith and Frith, 2003), suggesting that contour is the source of the authenticity effects seen in mPFC previously (Drolet *et al.*, 2013).

However, task appears to greatly influence how contour is processed. Authentic stimuli appear to be more important during the authenticity task than during the emotion task. As previously hypothesized (Drolet *et al.*, 2012), when trying to perceive authenticity the context of the recording may be more important, and since there are more previous experiences and, therefore, more potential contexts associated with authentic stimuli, these have the potential to induce greater activation. This was additionally corroborated by way of a direct comparison of values from emotion and authenticity tasks using ROIs from study 2 within study 1 (focusing specifically on TPJ and mBA9). Within these previous coordinates right medial PFC activation was significantly different. While medial PFC is sensitive to task-relevant stimulus features, TPJ does not appear to be differentially activated by an interaction of task and contour.

The rest of the network uncovered by the disjunction is part the top-down attention control network. Hopfinger and colleagues (2000) found that the SFG, parietal, and occipital gyrus increase activation during top-down shifting of attention, potentially due to increased working memory load. While almost the complete network seen in that study was found to be active in the disjunction, activation in IPL and IPS was particularly strong, likely due its role as a hub for the integration of various

stimulus features during focus is amplified here due to the inclusion of emotion (Klasen *et al.*, 2011) and speech intonation (Gandour *et al.*, 2003; Krishnan *et al.*, 2004) as important task and stimulus features. Importantly, the fact that this network was uncovered through a disjunction shows that, while this task-related attention occurs toward low-contour authentic stimuli for the authenticity task (where the greater number of potential contexts possible for authentic stimuli requires more processing), the emotion task induces greater processing of high contour play-acted stimuli. While the latter may simply be due to the bottom-up effect of high contour activation in the primary auditory cortex, the perception of play-acting may also be perceived as a form of pretense, inducing increased activation to ensure that the correct emotion was perceived relative to the context at hand. The involvement of medial PFC in detection of pretense has been documented before (German *et al.*, 2004; Shany-Ur *et al.*, 2012).

These and previous results point to an intriguing functional difference between TPJ and mPFC. TPJ was specifically recruited for the authenticity task and when a stimulus was explicitly cued as either authentic or play-acted, but did not differentiate stimuli differently based on task (Drolet *et al.* 2012, 2013). In other words, while activated top-down by task requirements, bottom-up modulation remained constant in TPJ. Medial PFC, however, was up-regulated either by authentic (low contour) or play-acted (high contour) stimuli dependent on task requirements. Therefore, mPFC became sensitive to those stimuli of greatest importance for a particular task. Stimuli with a greater likelihood of being play-acted may be treated as a form of pretense, in which case the stimulus is deemed as requiring more attention and is processed in more detail to ensure that the correct emotion is recognized during explicit emotion recognition. During authenticity perception, on the other hand, the number of potential contexts in which the emotion could have been produced increases mPFC activation as part of the ToM network. While the potential reasons remain speculative, clearly contour is processed differently based on task requirements in the medial prefrontal cortex and the attention control network, and while TPJ is activated by those task requirements, it does so independent of the stimulus features perceived.

## 6. General Discussion

The overarching concept of this project was that of top-down versus bottom-up cognitive processing and how these interact and lead to behavior associated with the recognition of emotional expressions. I specifically focused on the perception of emotional prosody and was additionally interested in advancing the discussion of cultural versus universal emotion expression. The authentic/play-acted dichotomy of emotional expressions provided a platform for testing the relevant interactions of emotion and cognition.

### 6.1 Brain Activation

#### 6.1.1 Bottom-up Stimulus-driven effects

The stimulus-driven effect of authenticity and correlated contour modulation in TTG clearly shows the bottom-up influence of these stimulus properties. Cortical auditory processing begins in TTG (Binder *et al.*, 1994; Celesia, 1976) and is coded tonotopically for the span of frequencies available to human hearing (Pantev *et al.*, 1995; Romani *et al.*, 1982). It is likely due to this tonotopic organization that an increase in contour also increases activation: greater F0 variability means a greater number of frequencies in the stimuli, which means more neurons firing and a corresponding increase in the BOLD response (Buxton *et al.*, 2004; Ogawa *et al.*, 1993). It is the combination of the last two studies that makes this result so compelling. **Studies III** and **IV** indicated that TTG is activated independent of either task or supplementary cueing, additionally strengthening the evidence for stimulus-driven activation due to specific auditory features. The similarity in the localization of this activation between the authenticity contrast in **study III** and the parametric contour analysis in **study IV** provide additional evidence for the role of contour as acoustic correlate to authenticity. It cannot be excluded, however, that some top-down attention influenced this activation. While **studies III** and **IV** are quite clear, **study I** did not find significant modulation in TTG in the authentic versus play-acted contrast during either task. While this could “simply” be an issue of signal strength due to the presence of two different tasks, there is the possibility of some top-down modulation due to the authenticity task (not required in **study III**). It has been found that activation in early sensory cortical regions like TTG can be modulated based on task demands (Fritz *et al.*, 2007; Todorovic *et al.*, 2011). Overall, however, this activation appears to be primarily bottom-up and tuned to the stimulus.

Of the network known to be activated for emotional prosody, the region from middle to superior temporal gyri was most affected by authenticity- and contour-based contrasts throughout these studies. This is in keeping with the involvement of these parts of the temporal cortex in speech

perception overall (Mazoyer *et al.*, 1993). In fact, Vigneau and colleagues (2006; 2011) found that phonemes, the basic units of speech sounds that combine to make up meaningful units and overall prosody, preferentially activate superior portions of the temporal cortex. While activation peaked significantly in TTG when contrasting authentic and play-acted stimuli in the current studies, in uncorrected contrasts from both fMRI experiments this activation extended beyond the primary and secondary auditory cortices, such that bottom-up influences of prosody overlap with phoneme processing in the superior temporal gyrus (Vigneau *et al.*, 2011).

Based only on these results it is not clear what the activation in STS signifies. Its activation is confounded by the concept of pSTS as a component of the ToM network. Research examining perception of emotional prosody has consistently found a strong effect in pSTS (Ethofer, 2006; Ethofer *et al.*, 2006). Interestingly, this activation was found when participants are explicitly tasked with recognition of emotional prosody and not in more simple contrasts of emotional versus non-emotional intonation. When only examining acoustic features of emotional prosody, activation again occurred more dorsally within the superior temporal gyrus (Wiethoff *et al.*, 2008). This area in the superior temporal gyrus has been suggested to be an “emotional voice area” due to its consistent activation by emotional intonation (Ethofer *et al.*, 2012). This strongly suggests that authenticity and contour modulation in pSTS was related to social cognition as initially hypothesized instead of more basic perception of emotion, which does not require extensive ToM abilities. Nevertheless, pSTS is spontaneously recruited by acoustic cues involved in the authentic/play-acted dichotomy. This implicit effect of authenticity shows that even such subtle contextual cues influencing perception of the speaker are attended to. While potential bottom-up activation of ToM processes is a potential explanation, the specific role of the STS may essentially be as a hub for social stimuli from various domains, recruiting other regions as required (Lahnakoski *et al.*, 2012). This would explain the wide variety of stimuli that can activate this region, all with a common theme of social perception, including ToM (Frith and Frith, 2003), biological motion (Puce and Perrett, 2003) and, the most relevant here, emotion recognition in face and voice (Hagan *et al.*, 2009; Jacob *et al.*, 2012; Peelen *et al.*, 2010).

Based on a preliminary study it was assumed that there would be an effect of authenticity at least on anger and sadness. However, while emotion recognition was quite poor, some emotion-specific effects in the BOLD response were expected in addition to the authenticity effects discussed so far. However, **study I** clearly showed that any potential between-emotion effects were even more subtle than the authenticity effects. No emotion-based contrast produced clear activation for any particular emotion. While emotion recognition-specific effects have been found elsewhere, the activation of specific regions or networks related to the recognition of expressions of individual emotions has been rather inconsistent between studies (Barrett, 2011). Such effects may require more sensitive

analysis techniques, such as multivariate pattern analysis (Ethofer *et al.*, 2009). However, such studies also make use of preselected and prototypical expressions. Effects of more subtle real-world stimuli on brain activation may require even more sensitive methodologies to detect the nevertheless above-chance recognition rates seen in this project.

### 6.1.2 Top-Down Task-Driven Effects

The primary network expected to be activated by authenticity was the ToM network. While the entire ToM network was activated for explicit authenticity categorization, mPFC and TPJ were most consistently activated by stimulus authenticity throughout the experiments, as seen in **studies III** and **IV**. Considering lateralization, the left TPJ was activated more than the right. This is in line with findings that left TPJ is preferentially recruited when emotional-context is relevant to social cognition (Burnett and Blakemore, 2009; Herve *et al.*, 2013). For the paradigm involved here, then, it is the interaction of ToM and emotion perception that led to left TPJ activation. Its role in the ToM network was shown in **study I** for explicit authenticity categorization and further modulation by authenticity cueing in **study III**. However, this involvement with top-down modulation through task or cue differs from the pSTS activation, which appeared independent of either cueing or task. Common to both the authenticity task and cueing, however, is that participants' attention was explicitly directed towards authenticity. Thus, TPJ is recruited when attention to speaker intention is induced top-down, whether this involves explicit mentalising or not. This parallels findings from Young and colleagues (Young *et al.*, 2010) on implicit mentalising and perspective-taking in TPJ.

Preferential activation by authenticity during authenticity categorization implies that mPFC, in association with TPJ, is activated for this type of task due to the requirement to perceive the speaker's intentional state (internally triggered or play-acted expression). **Study I** showed that mPFC is not recruited for simple emotion recognition when compared to authenticity perception. This may explain the lack of an effect of cueing on emotion categorization within **study III**. While the inclusion of the authenticity cues can conflict with stimulus perception in TPJ as seen in that study, this did not influence emotion labeling because mPFC was not explicitly recruited. In fact, **study IV** showed that during emotion recognition mPFC activation is reversed, increasing for play-acted instead of authentic stimuli, with no effect of cue in **study III**. It has been hypothesized that one role of the mPFC is to integrate and maintain contextual information and to compare with memory from the hippocampus (Polyn and Kahana, 2008). Here mPFC is activated by authenticity to maintain this information for during mentalising, but this integration is only active when explicit ToM is required. During the emotion task this did not appear to be necessary and, while the cues in **study III** may be decoded and learned, they do not affect the task at hand.

Initially hypothesized from the data in **study III**, the differences in activation between the tasks indicates that explicit perception of authenticity and emotion occur in parallel but do not directly influence each other. Nevertheless, authenticity does affect emotion categorization. Thus, authenticity can simultaneously influence our perception of the authenticity of a vocal expression while also affecting our perception of the encoded emotion, without any awareness that these processes are occurring in parallel. Frith and Frith (2008) reviewed a series of studies examining such implicit and explicit processes in social cognition. From these data they suggested that implicit and explicit processes can be either complementary or direct opposites, but often without direct interaction between the two. Most importantly, explicit processes seem to rarely directly affect implicit processes. Considering the automatic and spontaneous nature of emotion perception (Eimer and Holmes, 2007; Laukka *et al.*, 2011; Neumann and Strack, 2000) and other social cognition-relevant features in general (Engell *et al.*, 2007; German *et al.*, 2004) it may be that these abilities have evolved this way to ensure rapid recognition of emotional information without potential interference and slowed reaction times from further processing of contextual cues (Albert *et al.*, 2010; Drevets, 1998).

Based on the study by Jürgens and colleagues (2011) it appeared likely that the effect of authenticity discussed above was due to contour variability. **Study III** could not confirm a top-down influence on emotion perception. It therefore remained uncertain whether the effect of contour on emotion perception was due to learned associations with authenticity or, more simply, a direct effect on the emotional component of intonation. The most important indication from brain activation that contour does not simply make ambiguous stimuli “sound like” a particular emotion, is the activation of the attention network within **study IV**. The part of the attention network modulated spontaneously by low contour (independent of task) was previously shown to be specific to bottom-up processing of the target stimulus (Hopfinger *et al.*, 2000). Authentic stimuli induced higher cognitive load, including greater working memory load as seen in the parallel activation of mPFC and left lateral cerebellum (Knolle *et al.*, 2011; Koelsch *et al.*, 2009). The top-down attention network recruited mPFC for explicit perception of authenticity, with increased processing of authentic stimuli held in working memory. During both emotion and authenticity categorization, spontaneous activation by authentic stimuli activated a network of regions previously associated with stimulus-induced attention in both the visual (Hopfinger *et al.*, 2000) and auditory domains (Downar *et al.*, 2000; Westerhausen *et al.*, 2010). However, during the emotion task this did not lead to recruitment of other cognitive processes such as ToM. While authentic stimuli activated this stimulus-induced attention network, this did not influence mPFC activation, which was up-regulated by play-acted stimuli, likely due to the same bottom-up effect as in TTG. While authenticity is processed bottom-up and parallel to emotional content, these two sources of information are not integrated during

emotion categorization. Instead, emotion perception relies on learned associations for rapid stimulus perception. The previously assumed to be bottom-up activation by authentic stimuli in mPFC from **study I** was likely a carry-over effect due to the randomized trial order, such that participants were still thinking about authenticity when an emotion task occurred immediately after an authenticity task. While information about contour is decoded, it does not spontaneously recruit ToM, instead relying on bottom-up processes during emotion categorization. However, spontaneous activation of the attention network during such categorization indicates that authenticity differences may be retained in memory to improve perception of similar emotional expressions in the future.

## 6.2 Behavior

### 6.2.1 Authenticity Effect

While there were slight differences in recognition rates, the overall pattern remained the same throughout all four studies. The behavioral task of emotion recognition was repeated with three separate groups of German participants. The three groups were each presented with slightly different situations in which to perform the tasks. **Study I** required fast response times, single stimulus playbacks, and the interfering nature of the fMRI methodology (small space, loud background noise). **Study II** was performed in a regular office environment, at a desk with a computer. Participants were able to replay the stimuli several times and there was no time limit for responding. Finally, the experiment described in **study III** was similar to **study I** due to the fMRI setting, but participants were explicitly told whether a stimulus was authentic or play-acted for two-thirds of the recordings. While such widely differing settings were not a specific goal of this project, they do show how robust the effect of authenticity is.

The behavioral results were surprising in several ways. First and foremost was the lack of any clear differentiation by authenticity in fear and joy stimuli. Fear, in particular, was recognized only at chance levels throughout all the experiments described, which was surprising considering previous findings related to encoding of fear in the voice (Pell *et al.*, 2009; Sauter *et al.*, 2010; but see Effenbein and Ambady, 2002) and the relevance as information related to potential threat (Ethofer, 2006). While recordings were selected based on emotionally expressive speech, selection of stimulus segments had to be based on neutral verbal content. This, combined with the fact that stimuli were not preselected based on high recognition, is likely a major reason for the low ratings overall. However, this is not a methodological problem *per se*. Most studies make use of such idealized and pretested stimuli for empirical reasons, and experiments involving prototypical expressions of emotion are important since many of the findings to date would not have been possible otherwise. However, the current results clearly show that those previous findings do not exactly portray day-to-

day social communication. Thus, stimulus sets similar to the one used here will be needed to fully understand such interactions.

Further, while acting is enough to create an approximation of vocal emotion expression, the effects of authenticity are distinct. The effects seen here show, for the first time, that it is currently not possible to remove all indications of acting from the voice, even with professional training. Combined with a previous study on the effects of authenticity in facial expressions (Wilting *et al.*, 2006) this shows that it may be true for acting overall. With the authenticity ratings so close to chance levels, this effect is also clearly implicit. This means that using explicit authenticity ratings of stimuli would not be enough to select appropriate stimuli with respect to the potential influence of authenticity. In addition, the lack of an effect of explicitly labeling stimuli by authenticity indicates how persistent and automatic this is. Depending on the experimental paradigm, authenticity could have a significant effect that is not easily controlled for.

The first two studies additionally included choice theoretical analyses (Luce, 1959). While hit rates can provide a wealth of information about emotion perception, sensitivity and bias values can provide information about what leads to these rates. Sensitivity indicates how well stimulus categories can be discerned from one another independent of bias. Bias values indicate whether participants tend to prefer particular categories. Participants had a bias towards rating authentic expressions as sadness and play-acted expressions as anger. There was no bias for or against labeling expressions as fear, while there was a general tendency for all participants to not choose joy. The lack of bias and low recognition rates for fear provide more evidence that this emotion is not well encoded in emotional prosody. The bias against selecting joy, however, was a more surprising finding. It may be that a lack of risk in making a type 1 error with happy stimuli reduces the need to select this category, since it is less likely that a negative consequence would be incurred from making a false assumption. This is in contrast to the other negative emotions, where mislabeling could lead to more negative consequences (e.g. ignoring an expression of anger). This also shows that perfect recognition of emotional content is not a prerequisite for studying emotional expressions. Very high recognition rates due to stimulus preselection based on hit rates would have made it impossible to examine the effect of authenticity. While many studies require higher emotion recognition rates (e.g. to determine emotion-specific regions or networks in the brain), examining interactions with other factors, including those present in more natural settings, may benefit from not using “ideal” stimulus sets.

Completely unambiguous emotion perception may not be necessary or even advantageous. For example, auditory emotion is seldom the only piece of information available. Vocal expressions usually complement other sources of information, such as facial expressions (Rigoulot and Pell, 2012)



and each can greatly influence the perception of the other (De Gelder and Vroomen, 2000). It may be that such inter-sensory redundancy (Campanella and Belin, 2007) makes implicit information about authenticity in the voice less problematic since further information can be acquired from subsequent vocal expressions, verbal content, and the other senses. In addition, if a complete picture is not restricted by the limitations of a single modality, then implicit contextual information may not require recruitment of explicit resources. Nevertheless, learning from such implicit information could be useful to future interactions. Chun and Jiang (2003) provide experimental evidence of implicit information from visual contextual cues persisting in memory for at least a week. While acted behavior may not represent a serious enough breach in expectation at the moment of emotion recognition, learning over time that an individual may not express authentic emotions may influence responses towards that individual in the future.

### 6.2.2 Cultural Effect

With participants in Romania (limited exposure to German) and Indonesia (no exposure to German) similarly affected by authenticity, the major difference found between cultures concerned the amount of bias against choosing anger for authentic stimuli. While the implications regarding potential differing cultural backgrounds was discussed in that chapter, I return to this topic here to further address the question of universal versus cultural emotions. The very similar recognition rates for the emotions between all three cultures indicates that there is some level of universality to emotional expressions as implicated by Elfenbein and Ambady (2002) and Scherer and colleagues (Scherer *et al.*, 2001). However, the differences in biases between cultures show that a combination of universals and social learning best explains these data. The implicit nature of the effect of authenticity on emotion categorization shows that conscious perception by the individual is not required, instead relying on previously learned distinctions that directly affect expectations, potentially due to differences between individualistic and collectivistic cultures (Harada *et al.*, 2010; Matsumoto *et al.*, 1998; Oyserman *et al.*, 2002). In the effect of authenticity we therefore find a combination of innate emotion recognition and the effect of culture.

A combination of universal and cultural influences in emotion expression could provide a framework for contemporary discussions on the matter. After almost a century of direct comparisons of emotion expressions between and within cultures, the points of contention have changed very little. While proponents of the socio-cultural view of emotion expressions gather more evidence for their point of view (Jack *et al.*, 2012) and defend it strongly (Jack *et al.*, 2013) against counter-arguments (Sauter and Eisner, 2013), even today many studies are uncovering further evidence for universals in both facial (Elfenbein and Ambady, 2003) and vocal expressions of emotion (Sauter *et al.*, 2010). Based in part on the findings of the cross-cultural study presented here, I suggest that the truth must lie

somewhere in between. The recognition rates of emotional intonation between the cultures are too similar to be entirely learned, but the variation in authenticity-induced biases shows that modification due to social environment must be occurring as well. If neither theory of the expression of emotions can entirely explain the findings supporting the other, then the concept of complex interactions of culture and innate emotional expressions is likely to be the only plausible explanation. While theories have been further developed to account for diverging findings (Elfenbein and Ambady, 2002; Scherer *et al.*, 2011), the overall search for real-world examples of both continues unabated. Considering the historical evidence for both culture-specific and universal emotions as described in the introduction, changing the research focus towards the types and occurrences of interactions between these two categories may provide a much more fruitful framework for developing future hypotheses on the subject.

### 6.3 Conclusions

#### 6.3.1 Top-down versus Bottom-up Processing

Since the inception of this research project, interest in context-dependent effects on social cognition (Koban *et al.*, 2012; Li *et al.*, 2012) and emotion recognition in particular (Albert *et al.*, 2010; Barrett and Kensinger, 2010; Chung and Barch, 2011; Gruber and McDonald, 2012; Jaeger and Rugg, 2012) has increased dramatically. The results presented here indicate two important factors paralleling other findings. First, effects on psychological and behavioral experimentation related to real-world scenarios have yet to be identified. The implicit and automated processes involved in the perception of authenticity show that many stimulus and context details can be processed in parallel, affect main task performance, and may indicate implicit learning of social stimuli (bottom-up attention network in **study IV**). The integration of context has been found to influence emotion perception in other studies and explicit emotion perception can even increase encoding of context (Barrett and Kensinger, 2010). Tracy and Prehn (2011) found that contextual information can influence how the emotional state of the speaker is perceived, even if the spoken stimulus itself is identical. In addition, psychiatric research has shown that problems with context encoding can vary greatly, with schizophrenic patients showing reduced dissociation of social and nonsocial contexts (Chung and Barch, 2011), while adults with Asperger syndrome show deficits in implicit social context encoding that can be improved upon through explicit descriptions (Baez *et al.*, 2012). Second, the differential effect that task had on authenticity-induced activation shows not only that vocal emotion perception BOLD response can be strongly influenced by secondary stimulus features, but that these effects are strongly task-dependent as well. Task is known to greatly influence activation in PFC in particular through its role in integration, attention, and decision-making (Buschman and Miller, 2007; Chadick

and Gazzaley, 2011; Gazzaley and Nobre, 2012), but the overlap of both implicit and explicit effects should be of particular interest for research so strongly influenced by relatively small effects.

### 6.3.2 *The Nature of Emotion Expression*

While mentioned only briefly in the introduction, the dichotomy between feeling and emotion may provide an explanation for the behavioral effect of and the implicit BOLD response to authenticity. This dichotomy differentiates between the physiological/behavior-triggering processes that occur implicitly (emotion) and the experience of emotion that requires some level of conscious perception (feeling). While the question of how to distinguish between emotion and feeling and how this influences emotion expression is ongoing (Damasio *et al.*, 2000; Goldie, 2009; Russell *et al.*, 2003), it is possible that either both or one of these is missing in the generation of the play-acted samples. This could be due to the apparently limited ability to directly control all aspects of vocalizations (Juslin and Laukka, 2001; Munhall *et al.*, 2009; Scherer, 1986). In particular, this provides a paradigm for experimentation regarding the necessity of an internal emotional and/or feeling state for the expression of emotion. The fact that expressions can be acted seems to indicate that such a state is not directly necessary. However, if the effect of authenticity on perception of expressions can be found to be associated with the presence or absence of a specific internal state, then the relationship between actual internal states of emotion and external expressions of emotion would become clearer.

If the effect of authenticity is related in some way to the internal state of the speaker, then this also has implications for acting and empirical research. The spontaneous reaction to authenticity shows that its perception is due to previously learned associations (likely stored as procedural knowledge) and that this is perceived across cultural boundaries. Whatever the underlying cause, authenticity can be perceived, and this must be taken into account in future studies and may lead to findings not available to research using only acted expressions. For example, the cross-cultural biases due to authenticity may provide an explanation for diverging evidence for universal and culturally-specific emotions. Depending on acting style actors may, to a greater or lesser extent, induce the emotional state they are expected to act (Goldstein, 2009). However, acted expressions could show even greater cultural effects if acting methodologies vary between countries, above and beyond the cultural differences in emotion expression in daily life. For example, the biases in emotion recognition may be remnants of certain in-group effects seen in other studies (Elfenbein and Ambady, 2002; Hart *et al.*, 2000; Scherer *et al.*, 2011). In fact, there is some evidence that in-group effects may be a construct of acted expressions (Matsumoto *et al.*, 2009). If cross-cultural emotion perception can be influenced by differences in expression and by acting, then this would produce very complex interactions not yet controlled for in emotion expression research. Not only could this

explain the differences in findings of universal (Elfenbein and Ambady, 2002) and culture-specific (Jack *et al.*, 2012) emotion expressions, it would also explain the patterns seen in recent studies of facial expressions. Jack and colleagues (2012) present what they deem to be clear indications of culture-specific recognition of emotions. However, Sauter and colleagues (2013) argue that despite the differences, the results also show universal effects. The complex interaction of emotion recognition, culture, and individual-specific acting styles could explain these discrepancies, leaving room for innate and learned emotional expressions to interact.

### 6.3 Outlook

Regarding the stimulus set used in this project, one of the main goals for future research will be to examine whether other sets of similarly acquired speech recordings produce similar results. While the effects on brain activation are quite clear, both these and the behavioral data would benefit from comparisons with other stimulus sources (different interview types, television, etc.). There are various factors that may influence findings. For example, the authentic speech context still involved an interview setting. With the development of electronic and online media over the past decade there is now an abundance of sources available over the internet that may more closely resemble “real-world” conditions.

For acted stimuli, a more specific focus on the effect of acting styles and individual styles of the actors themselves would be of particular interest. In fact, Jürgens and colleagues (in prep) have collected data showing that acted stimuli performed by non-professionals are recognized more similarly to authentic stimuli than the professionally play-acted ones used here. There are likely to be several different questions of this nature that remain to be explored.

For the bottom-up and top-down processes evident from the fMRI results, further research should go into the automatic up-regulation by authentic stimuli during emotion perception. It would be particularly interesting how the wide network spontaneously recruited by authenticity and contour influences later perception. This may require examinations of more long-term effects of cueing, for example, or the application of newly developed techniques, such as multivariate pattern analysis.

I have shown that a secondary cognitive feature (authenticity context) can have a direct impact on emotion recognition. However, the exact cause of the behavioral patterns could only be deduced from the current results, such as the possible influence of acting styles or the effect of learning on the perception of the acoustic correlates of authenticity. Examining this further would clarify not only the current results, but could lead to a greater understanding of when and how culture influences emotion recognition.

The results presented here and the suggestions for future study implicate the importance of examining the complex interactions of processes in the brain. While many individual elements have been addressed, the currently open and active questions will require a deeper understanding of the capacities of the brain for integration. The interactions between innate and learned emotion expressions, emotion and cognition in brain and in behavior, and top-down and bottom-up processing streams, will lead us to a better understanding of the brain and overall human behavior.



---

## 7. References

- Abraham, A., Werning, M., Rakoczy, H., von Cramon, D., & Schubotz, R. (2008) Minds, persons, and space: An fMRI investigation into the relational complexity of higher-order intentionality. *Conscious. Cogn.*, 17, 438–450.
- Adolphs, R. (1999) Intact recognition of emotional prosody following amygdala damage. *Neuropsychologia*, 37(11), 1285–1292.
- Adolphs, R. (2009) The social brain: neural basis of social knowledge. *Annu. Rev. Psychol.*, 60, 693–716.
- Adolphs, R., Damasio, H., & Tranel, D. (2002) Neural systems for recognition of emotional prosody: A 3-D lesion study. *Emotion*, 2(1), 23–51.
- Adolphs, R., Tranel, D., & Damasio, H. (2001) Emotion recognition from faces and prosody following temporal lobectomy. *Neuropsychology*, 15(3), 396–404.
- Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. R. (1995) Fear and the human amygdala. *J. Neurosci.*, 15(9), 5879–5891.
- Agresti, A. (2007) *An introduction into categorical data analysis*. Wiley, New Jersey.
- Albert, J., López-Martín, S., & Carretié, L. (2010) Emotional context modulates response inhibition: Neural and behavioral data. *NeuroImage*, 49(1), 914–921.
- Alexander, W. H. & Brown, J. W. (2011) Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.*, 14, 1338–1344.
- Allman, J., Hakeem, A., Erwin, J., Nimchinsky, E., & Hof, P. (2001) The anterior cingulate cortex: The Evolution of an Interface between Emotion and Cognition. *Ann Ny Acad Sci*, 935, 107–117.
- Anolli, L. (1997) The Voice of Deception: Vocal Strategies of Naive and Able Liars. *J. Nonverbal Behav.*, 21(4), 259–284.
- Apperly, I. A. (2008) Beyond Simulation–Theory and Theory–Theory: Why social cognitive neuroscience should use its own concepts to study ‘theory of mind’. *Cognition*, 107(1), 266–283.
- Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004) Assessing the auditory dual-pathway model in humans. *Neuroimage*, 22(1), 401–408.
- Audibert, N., Aubergé, V., & Rilliard, A. (2008a) Acted vs. spontaneous expressive speech: perception with inter-individual variability. In, *Programme of the Workshop on Corpora for Research on Emotion and Affect*, p. 23.
- Audibert, N., Aubergé, V., & Rilliard, A. (2008b) How we are not equally competent for discriminating acted from spontaneous expressive speech. *Speech Pros.*, p. 693.

## References

---

- Averill, J. R. (1980) A constructivist view of emotion. In, Plutchik,R. and Kellermann,H. (eds), *Emotion, Theory, Research and Experience*. Academic Press, New York, p. 305–339.
- Awad, M., Warren, J., Scott, S., Turkheimer, F., & Wise, R. (2007) A common system for the comprehension and production of narrative speech. *J. Neurosci.*, 27(43), 11455-11464.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.*, 59(4), 390–412.
- Bach, D., Grandjean, D., Sander, D., Herdener, M., Strik, W., & Seifritz, E. (2008) The effect of appraisal level on processing of emotional prosody in meaningless speech. *Neuroimage*, 42(2), 919–927.
- Bachevalier, J. & Meunier, M. (2005) The neurobiology of social-emotional cognition in nonhuman primates. In, Easton,A. and Emery,N. (eds), *The Cognitive Neuroscience of Social Behaviour*. Psychology Pr, East Sussex, p. 19–57.
- Baez, S., Rattazzi, A., Gonzalez-Gadea, M. L., Torralva, T., Vigliecca, N. S., Decety, J., et al. (2012) Integrating intention and context: assessing social cognition in adults with Asperger syndrome. *Front. Hum. Neurosci.*, 6, 302.
- Banse, R. & Scherer, K. (1996) Acoustic Profiles in Vocal Emotion Expression. *J. Pers. Soc. Psychol.*, 70 (3), 614–636.
- Bänziger, T. & Scherer, K. (2007) Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus. *Lect. Notes Comput. Sci.*, 4738, 476–487.
- Barkhuysen, P., Krahmer, E., & Swerts, M. (2007) Cross-modal perception of emotional speech.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985) Does the autistic child have a ‘theory of mind’? *Cognition*, 21(1), 37–46.
- Barrett, L. F. (2011) Was Darwin Wrong About Emotional Expressions? *Curr. Dir. Psychol. Sci.*, 20(6), 400–406.
- Barrett, L. F. & Kensinger, E. A. (2010) Context Is Routinely Encoded During Emotion Perception. *Psychol. Sci.*, 21(4), 595–599.
- Bates, D. (2005) Fitting linear mixed models in R using the lme4 package. *R News*, 5, 27–30.
- Baum, S. & Pell, M. (1999) The neural bases of prosody: Insights from lesion studies and neuroimaging. *PAPH*, 13(8), 581–608.
- Beaucousin, V., Lacheret, A., Turbelin, M., Morel, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007) FMRI study of emotional speech comprehension. *Cereb. Cortex*, 17(2), 339.
- Bechara, A., Damasio, H., & Damasio, A. (2000) Emotion, decision making and the orbitofrontal cortex. *Cereb. Cortex*, 10(3), 295.



- Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008) The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behav. Res. Methods*, 40(2), 531–539.
- Van Bezooijen, R., Otto, S. A., & Heenan, T. A. (1983) Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *J. Cross-Cult. Psychol.*, 14, 387–406.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., et al. (2000) Human Temporal Lobe Activation by Speech and Nonspeech Sounds. *Cereb. Cortex*, 10(5), 512 – 528.
- Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997) Human Brain Language Areas Identified by Functional Magnetic Resonance Imaging. *J. Neurosci.*, 17(1), 353 – 362.
- Binder, J. R., Rao, S. M., Hammeke, T. A., Yetkin, F. Z., Jesmanowicz, A., Bandettini, P. A., et al. (1994) Functional magnetic resonance imaging of human auditory cortex. *Ann. Neurol.*, 35(6), 662–672.
- Blair, K. (2007) Modulation of emotion by cognition and cognition by emotion. *Neuroimage*, 35(1), 430–440.
- Blanchette, I. & Richards, A. (2010) The influence of affect on higher level cognition: A review of research on interpretation, judgement, decision making and reasoning. *Cogn. Emot.*, 24(4), 561.
- Bliss, T. V. P. & Collingridge, G. L. (1993) A synaptic model of memory: long-term potentiation in the hippocampus. *Nature*, 361(6407), 31.
- Bolton, T. (1902) A biological view of perception. *Psychol. Rev.*, 9(6), 537–548.
- Braver, T., Barch, D., Gray, J., Molfese, D., & Snyder, A. (2001) Anterior cingulate cortex and response conflict: effects of frequency, inhibition and errors. *Cereb. Cortex*, 11(9), 825.
- Brück, C., Kreifelts, B., & Wildgruber, D. (2011) Emotional voices in context: a neurobiological model of multimodal affective information processing. *Phys. Life Rev.*, 8(4), 383–403.
- Bryant, G. & Barrett, H. (2008) Vocal emotion recognition across disparate cultures. *J. Cogn. Cult.* 8, 1 (2), 135–148.
- Buchanan, T., Lutz, K., Mirzazade, S., Specht, K., Shah, N., Zilles, K., et al. (2000) Recognition of emotional prosody and verbal components of spoken language: an fMRI study. *Cogn. Brain Res.*, 9 (3), 227–238.
- Buckner, R., Andrews-Hanna, J., & Schacter, D. (2008) The Brain's Default Network: Anatomy, Function, and Relevance to Disease. *Ann. N. Y. Acad. Sci.*, 1124(1), 1–38.
- Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008) Cognitive Biases and Nonverbal Cue Availability in Detecting Deception. *Hum. Commun. Res.*, 34(4), 572–599.

## References

---

- Burnett, S. & Blakemore, S. (2009) Functional connectivity during a social emotion task in adolescents and in adults. *Eur. J. Neurosci.*, 29(6), 1294–1301.
- Buschman, T. J. & Miller, E. K. (2007) Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. *Science*, 315(5820), 1860–1862.
- Bush, G., Luu, P., & Posner, M. I. (2000) Cognitive and emotional influences in anterior cingulate cortex. *Trends Cogn. Sci.*, 4(6), 215–222.
- Buxton, R. B., Uludağ, K., Dubowitz, D. J., & Liu, T. T. (2004) Modeling the hemodynamic response to brain activation. *NeuroImage*, 23, S220–S233.
- Call, J. & Tomasello, M. (1998) Distinguishing intentional from accidental actions in orangutans (*Pongo pygmaeus*), chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *J. Comp. Psychol.*, 112(2), 192.
- Campanella, S. & Belin, P. (2007) Integrating face and voice in person perception. *Trends Cogn. Sci.*, 11(12), 535–543.
- Cannon, W. B. (1931) Again the James-Lange and the thalamic theories of emotion. *Psychol. Rev.*, 38(4), 281.
- Cannon, W. B. (1927) The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory. *Am. J. Psychol.*, 39(1/4), 106–124.
- Carter, C. & van Veen, V. (2007) Anterior cingulate cortex and conflict detection: An update of theory and data. *Cogn. Affect. Behav. Neurosci.*, 7(4), 367–379.
- Celesia, G. G. (1976) Organization of auditory cortical areas in man. *Brain*, 99(3), 403–414.
- Chadick, J. Z. & Gazzaley, A. (2011) Differential coupling of visual cortex with default or frontal-parietal network based on goals. *Nat. Neurosci.*, 14(7), 830–832.
- Chun, M. M. & Jiang, Y. (2003) Implicit, long-term spatial contextual memory. *J. Exp. Psychol. Learn. Mem. Cogn.*, 29(2), 224–234.
- Chung, Y. S. & Barch, D. M. (2011) The effect of emotional context on facial emotion ratings in schizophrenia. *Schizophr. Res.*, 131(1-3), 235-241.
- Colibazzi, T., Posner, J., Wang, Z., Gorman, D., Gerber, A., Yu, S., et al. (2010) Neural systems subserving valence and arousal during the experience of induced emotions. *Emotion*, 10(3), 377–389.
- Corbetta, M. & Shulman, G. L. (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.*, 3(3), 201.
- Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., et al. (2001) Emotion recognition in human-computer interaction. *Signal Process. Mag. Ieee*, 18(1), 32–80.
- Damasio, A. (1994) *Descartes' Error: Emotion, Reason and the Human Brain*. HarperCollins, New York.

- Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L. ., Parvizi, J., et al. (2000) Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat. Neurosci.*, 3(10), 1049–1056.
- Darwin, C. (1872) *The expression of emotions in man and animals*. John Murray, London.
- Davidoff, J. (2001) Language and perceptual categorisation. *Trends Cogn. Sci.*, 5(9), 382–387.
- Davitz, J. (1964) *The communication of emotional meaning*. McGraw-Hill, New York.
- DeKosky, S., Heilman, K., Bowers, D., & Valenstein, E. (1980) Recognition and discrimination of emotional faces and pictures. *Brain Lang.*, 9(2), 206–214.
- DePaulo, B. M. (2003) Cues to Deception. *Psychol. Bull.*, 129(1), 74–118.
- Desmond, J. E., Gabrieli, J. D. E., Wagner, A. D., Ginier, B. L., & Glover, G. H. (1997) Lobular Patterns of Cerebellar Activation in Verbal Working-Memory and Finger-Tapping Tasks as Revealed by Functional MRI. *J. Neurosci.*, 17(24), 9675–9685.
- Dolcos, F., LaBar, K. S., & Cabeza, R. (2004) Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fMRI study. *NeuroImage*, 23(1), 64–74.
- Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2000) A multimodal cortical network for the detection of changes in the sensory environment. *Nat. Neurosci.*, 3(3), 277–283.
- Drevets, W. C. (1998) Reciprocal Suppression of Regional Cerebral Blood Flow during Emotional versus Higher Cognitive Processes: Implications for Interactions between Emotion and Cognition. *Cogn. Emot.*, 12(3), 353–385.
- Drolet, M., Schubotz, R. I., & Fischer, J. (2012) Authenticity affects the recognition of emotions in speech: behavioral and fMRI evidence. *Cogn. Affect. Behav. Neurosci.*, 12(1), 140–150.
- Drolet, M., Schubotz, R. I., & Fischer, J. (2013) Explicit authenticity and stimulus features interact to modulate BOLD response induced by emotional speech. *Cogn. Affect. Behav. Neurosci.*, 1–12.
- Dunbar, R. I. (1998) The social brain hypothesis. *Brain*, 9, 10, 178-190.
- Eibl-Eibesfeldt, I. (1968) [On the ethology of the human greeting behavior. I. Observations on Bali natives Papuans and Samoans along with comparative remarks]. *Z. Tierpsychol.*, 25(6), 727-744.
- Eimer, M. & Holmes, A. (2007) Event-related brain potential correlates of emotional face processing. *Neuropsychologia*, 45(1), 15–31.
- Ekman, P. (1992) An argument for basic emotions. *Cogn. Emot.*, 6(3-4), 169–200.
- Ekman, P. (1996) Why don't we catch liars? *Soc. Res.*, 63, 801–817.
- Ekman, P., Davidson, R., & Friesen, W. (1990) The Duchenne smile: Emotional expression and brain physiology II. *J. Pers. Soc. Psychol.*, 58(2), 342–353.

## References

---

- Ekman, P., Friesen, W., & O'Sullivan, M. (1988) Smiles when lying. *J. Pers. Soc. Psychol.*, 54(3), 414–420.
- Ekman, P., Friesen, W., & Scherer, K. (1976) Body movement and voice pitch in deceptive interaction. *Semiotica*, 16, 23–27.
- Ekman, P. & Friesen, W. V. (1971) Constants across cultures in face and emotion. *J. Pers. Soc. Psychol.*, 17, 124–129.
- Ekman, P. & O'Sullivan, M. (1991) Who can catch a liar? *Am. Psychol.*, 46, 913–920.
- Ekman, P. & Oster, H. (1979) Facial expressions of emotion. *Annu. Rev. Psychol.*, 30, 527–554.
- Ekman, P., Sorenson, E., & Friesen, W. (1969) Pan-Cultural Elements in Facial Displays of Emotion. *Science*, 164(3875), 86–88.
- Elfenbein, H. A. & Ambady, N. (2003) When familiarity breeds accuracy: cultural exposure and facial emotion recognition. *J. Pers. Soc. Psychol.*, 85(2), 276.
- Elfenbein, H. A., Beaupre, M., Levesque, M., & Hess, U. (2007) Toward a dialect theory: cultural differences in the expression and recognition of posed facial expressions. *Emotion*, 7, 131–146.
- Elfenbein, H. A., Mandal, M. K., Ambady, N., Harizuka, S., & Kumar, S. (2002) Cross-cultural patterns in emotion recognition: Highlighting design and analytical techniques. *Emotion*, 2(1), 75–84.
- Elfenbein, H. & Ambady, N. (2002) On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychol. Bull.*, 128(2), 203–235.
- Emery, N. (2004) *The evolution of social cognition. The Cognitive Neuroscience of Social Behaviour.* Psychology Press, New York.
- Engell, A. D., Haxby, J. V., & Todorov, A. (2007) Implicit Trustworthiness Decisions: Automatic Coding of Face Properties in the Human Amygdala. *J. Cogn. Neurosci.*, 19(9), 1508–1519.
- Ethofer, T. (2006) Impact of voice on emotional judgment of faces: An event-related fMRI study. *Hum. Brain Mapp.*, 27(9), 707–714.
- Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., et al. (2006) Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *Neuroimage*, 30(2), 580–587.
- Ethofer, T., Bretschner, J., Gschwind, M., Kreifelts, B., Wildgruber, D., & Vuilleumier, P. (2012) Emotional Voice Areas: Anatomic Location, Functional Properties, and Structural Connections Revealed by Combined fMRI/DTI. *Cereb. Cortex*, 22(1), 191–200.
- Ethofer, T., Van De Ville, D., Scherer, K., & Vuilleumier, P. (2009) Decoding of emotional information in voice-sensitive cortices. *Curr. Biol.*, 19(12), 1028–1033.
- Everhart, D., Demaree, H., & Shipley, A. (2006) Perception of Emotional Prosody: Moving Toward a Model That Incorporates Sex-Related Differences. *Behav. Cogn. Neurosci. Rev.*, 5(2), 92.

- Ferstl, E. & von Cramon, D. (2002) What Does the Frontomedian Cortex Contribute to Language Processing: Coherence or Theory of Mind? *NeuroImage*, 17(3), 1599–1612.
- Freedman, D. G. (1964) Smiling in blind infants and the issue of innate vs. acquired. *J. Child Psychol. Psychiatry*, 5(3-4), 171–184.
- Frick, R. (1985) Communicating emotion: The role of prosodic features. *Psychol. Bull.*, 97(3), 412–429.
- Friederici, A. D. (2012) The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn. Sci.*, 16(5), 262-268.
- Frijda, N. (1986) *The emotions: Studies in emotion and social interaction*. Cambridge University Press, Cambridge, England.
- Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., & Turner, R. (1998) Event-related fMRI: characterizing differential responses. *Neuroimage*, 7(1), 30–40.
- Frith, C. D. & Frith, U. (2012) Mechanisms of Social Cognition. *Annu. Rev. Psychol.*, 63(1), 287–313.
- Frith, C. & Frith, U. (2008) Implicit and Explicit Processes in Social Cognition. *Neuron*, 60(3), 503–510.
- Frith, C. & Frith, U. (2007) Social cognition in humans. *Curr. Biol.*, 17(16), 724–732.
- Frith, C. & Singer, T. (2008) The role of social cognition in decision making. *Philos. Trans. R. Soc. B Biol. Sci.*, 363(1511), 3875–3886.
- Frith, U. & Frith, C. (2003) Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. B Biol. Sci.*, 358(1431), 459–473.
- Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007) Auditory attention — focusing the searchlight on sound. *Curr. Opin. Neurobiol.*, 17(4), 437–455.
- Gallagher, H. & Frith, C. (2003) Functional imaging of ‘theory of mind’. *Trends Cogn. Sci.*, 7(2), 77–83.
- Gandour, J., Wong, D., Dziedzic, M., Lowe, M., Tong, Y., & Li, X. (2003) A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. *Hum. Brain Mapp.*, 18(3), 149–157.
- Gazzaley, A. & Nobre, A. C. (2012) Top-down modulation: bridging selective attention and working memory. *Trends Cogn. Sci.*, 16(2), 129–135.
- De Gelder, B. & Vroomen, J. (2000) The perception of emotions by ear and by eye. *Cogn. Emot.*, 14(3), 289–311.
- Gendron, M. & Barrett, L. (2009) Reconstructing the Past: A Century of Ideas About Emotion in Psychology. *Emot. Rev.*, 1(4), 316–339.
- George, M. S., Parekh, P. I., Rosinsky, N., Ketter, T. A., Kimbrell, T. A., Heilman, K. M., et al. (1996) Understanding emotional prosody activates right hemisphere regions. *Arch. Neurol.*, 53(7), 665.
- German, T., Niehaus, J., Roarty, M., Giesbrecht, B., & Miller, M. (2004) Neural correlates of detecting pretense: Automatic engagement of the intentional stance under covert conditions. *J. Cogn. Neurosci.*, 16(10), 1805–1817.

## References

---

- Gilboa, A., Winocur, G., Grady, C., Hevenor, S., & Moscovitch, M. (2004) Remembering our past: functional neuroanatomy of recollection of recent and very remote personal events. *Cereb. Cortex*, 14(11), 1214.
- Gobl, C. & Ni Chasaide, A. (2003) The role of voice quality in communicating emotion, mood and attitude. *Speech Commun.*, 40, 189–212.
- Goldie, P. (2009) Getting Feelings into Emotional Experience in the Right Way. *Emot. Rev.*, 1(3), 232–239.
- Goldstein, T. (2011) Correlations Among Social-Cognitive Skills in Adolescents Involved in Acting or Arts Classes. *Mind Brain Educ.*, 5(2), 97–103.
- Goldstein, T. (2009) Psychological perspectives on acting. *Psychol. Aesthet. Creat. Arts*, 3(1), 6–9.
- Goldstein, T. R. & Bloom, P. (2011) The mind on stage: why cognitive scientists should study acting. *Trends Cogn. Sci.*, 15, 141–142.
- Goudbeek, M. & Scherer, K. (2010) Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *J. Acoust. Soc. Am.*, 128(3), 1322–1336.
- Grandjean, D., Sander, D., & Scherer, K. (2008) Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization. *Conscious. Cogn.*, 17(2), 484.
- Grezes, J., Berthoz, S., & Passingham, R. (2006) Amygdala activation when one is the target of deceit: Did he lie to you or to someone else? *Neuroimage*, 30(2), 601–608.
- Grezes, J., Frith, C., & Passingham, R. (2004) Brain Mechanisms for Inferring Deceit in the Actions of Others. *J. Neurosci.*, 24(24), 5500–5505.
- Gruber, A. J. & McDonald, R. J. (2012) Context, emotion, and the strategic pursuit of goals: interactions among multiple brain systems controlling motivated behavior. *Front. Behav. Neurosci.*, 6, 50.
- Gur, R. C., Gunning-Dixon, F., Bilker, W., & Gur, R. E. (2002) Sex differences in temporo-limbic and frontal brain volumes of healthy adults. *Cereb. Cortex*, 12(9), 998.
- Hagan, C. C., Woods, W., Johnson, S., Calder, A. J., Green, G. G., & Young, A. W. (2009) MEG demonstrates a supra-additive response to facial and vocal emotion in the right superior temporal sulcus. *Proc. Natl. Acad. Sci.*, 106(47), 20010.
- Haggard, P. (2008) Human volition: towards a neuroscience of will. *Nat Rev Neurosci*, 9(12), 934–946.
- Hall, J. & Schmid Mast, M. (2007) Sources of accuracy in the empathic accuracy paradigm. *Emotion*, 7(2), 438–446.
- Hammerschmidt, K. & Jürgens, U. (2007) Acoustical correlates of affective prosody. *J. Voice*, 21(5), 531–540.

- Harada, T., Li, Z., & Chiao, J. (2010) Differential dorsal and ventral medial prefrontal representations of the implicit self modulated by individualism and collectivism: An fMRI study. *Soc. Neurosci.*, 5(3), 257 – 271.
- Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000) Differential response in the human amygdala to racial outgroup vs ingroup face stimuli. *Neuroreport*, 11(11), 2351–2354.
- Herve, P.-Y., Razafimandimby, A., Jobard, G., & Tzourio-Mazoyer, N. (2013) A Shared Neural Substrate for Mentalizing and the Affective Component of Sentence Comprehension. *Plos One*, 8(1), e54400.
- Hervé, P.-Y., Razafimandimby, A., Vigneau, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2012) Disentangling the brain networks supporting affective speech comprehension. *NeuroImage*, 61(4), 1255–1267.
- Hietanen, J., Surakka, V., & Linnankoski, I. (1998) Facial electromyographic responses to vocal affect expressions. *Psychophysiology*, 35(5), 530–536.
- Hoekert, M., Vingerhoets, G., & Aleman, A. (2010) Results of a pilot study on the involvement of bilateral inferior frontal gyri in emotional prosody perception: an rTMS study. *Bmc Neurosci.*, 11(1), 93.
- Hofstede, G. (1980) *Cultures's consequences*. Sage, Beverly Hills, CA.
- Hofstede, G. (1996) The Nation-state as a source of common mental programming: Similarities and differences across Eastern and Western Europe. In, Gustavsson, S. and Lewin, L. (eds), *The Future of Nation State - Essays on cultural pluralism and political integration*. Routledge, London, New York, p. 2–20.
- Hollan, D. (2000) Constructivist Models of Mind, Contemporary Psychoanalysis, and the Development of Culture Theory. *Am. Anthr.*, 102(3), 538–550.
- Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T., & D'Esposito, M. (2008) Mentalizing about emotion and its relationship to empathy. *Soc. Cogn. Affect. Neurosci.*, 3(3), 204.
- Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000) The neural mechanisms of top-down attentional control. *Nat. Neurosci.*, 3(3), 284.
- Hornak, J., Rolls, E. ., & Wade, D. (1996) Face and voice expression identification in patients with emotional and behavioural changes following ventral frontal lobe damage. *Neuropsychologia*, 34(4), 247–261.
- Hothorn, T., Bretz, F., & Westfall, P. (2008) Simultaneous inference in general parametric models. *Biom. J.*, 50, 346 – 363.
- Houston, B. K. & Holmes, D. S. (1975) Role Playing Versus Deception: The Ability of Subjects to Simulate Self-Report and Physiological Responses. *J. Soc. Psychol.*, 96(1), 91–98.

## References

---

- Hunt, W. (1941) Recent developments in the field of emotion. *Psychol. Bull.*, 38(5), 249–276.
- Izard, C. (2009) Emotion theory and research: Highlights, unanswered questions, and emerging issues. *Annu. Rev. Psychol.*, 60, 1–25.
- Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012) Facial Expressions of Emotion Are Not Culturally Universal. *Proc. Natl. Acad. Sci.*, 109(19), 7241–7244.
- Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2013) Reply to Sauter and Eisner: Differences outweigh commonalities in the communication of emotions across human cultures. *Proc. Natl. Acad. Sci.*, 110(3), E181–E182.
- Jacob, H., Kreifelts, B., Brück, C., Erb, M., Hösl, F., & Wildgruber, D. (2012) Cerebral integration of verbal and nonverbal emotional cues: Impact of individual nonverbal dominance. *NeuroImage*, 61(3), 738–747.
- Jaeger, A. & Rugg, M. (2012) Implicit effects of emotional contexts: An ERP study. *Cogn. Affect. Behav. Neurosci.*, 1–13.
- James, W. (1884) II. WHAT IS AN EMOTION? *Mind*, (34), 188.
- James, W. (1890) *The Principles of Psychology*, Vol. 2. Courier Dover Publications.
- Jaywant, A. & Pell, M. D. (2012) Categorical processing of negative emotions from speech prosody. *Speech Commun.*, 54(1), 1–10.
- Jürgens, R., Hammerschmidt, K., & Fischer, J. (2011) Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Front. Emot. Sci.*, 2, 180.
- Juslin, P. & Laukka, P. (2003) Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychol. Bull.*, 129(5), 770–814.
- Juslin, P. N. (2012) Are musical emotions invariant across cultures? *Emot. Rev.*, 4, 283–284.
- Juslin, P. N. & Laukka, P. (2001) Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion*, 1(4), 381–412.
- Kappas, A. & Poliakova, N. (2007) Judgments of the affective valence of spontaneous vocalizations: The influence of situational context. In Izdebski, K. (ed), *Emotions in the Human Voice, Volume 1: Foundations*. Plural Publishing Inc., San Diego, CA. p. 109–122.
- Kéri, S., Decety, J., Roland, P., & Gulyás, B. (2004) Feature uncertainty activates anterior cingulate cortex. *Hum. Brain Mapp.*, 21(1), 26–33.
- Klasen, M., Kenworthy, C. A., Mathiak, K. A., Kircher, T. T. J., & Mathiak, K. (2011) Supramodal Representation of Emotions. *J. Neurosci.*, 31(38), 13635–13643.
- Knolle, F., Schröger, E., Baess, P., & Kotz, S. A. (2011) The Cerebellum Generates Motor-to-Auditory Predictions: ERP Lesion Evidence. *J. Cogn. Neurosci.*, 1–9.



- Koban, L., Pourtois, G., Bediou, B., & Vuilleumier, P. (2012) Effects of social context and predictive relevance on action outcome monitoring. *Cogn. Affect. Behav. Neurosci.*, 12(3), 460-478.
- Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008) Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage*, 42(2), 998–1031.
- Kochunov, P., Mangin, J., Coyle, T., Lancaster, J., Thompson, P., Rivière, D., et al. (2005) Age-related morphology trends of cortical sulci. *Hum. Brain Mapp.*, 26(3), 210–220.
- Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009) Functional architecture of verbal and tonal working memory: An fMRI study. *Hum. Brain Mapp.*, 30(3), 859–873.
- Kompus, K., Hugdahl, K., Öhman, A., Marklund, P., & Nyberg, L. (2009) Distinct control networks for cognition and emotion in the prefrontal cortex. *Neurosci. Lett.*, 467(2), 76–80.
- Kornbrot, D. E. (1978) Theoretical and empirical comparison of Luce’s choice model and logistic Thurstone model of categorical judgment. *Percept. Psychophys.*, 24, 193–208.
- Kotz, S. A., Kalberlah, C., Bahlmann, J., Friederici, A. D., & Haynes, J. (2012) Predicting vocal emotion expressions from the human brain. *Hum. Brain Mapp.*, 10.1002/hbm.22041.
- Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007) Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *NeuroImage*, 37(4), 1445–1456.
- Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. A. (2004) Human frequency-following response: representation of pitch contours in Chinese tones. *Hear. Res.*, 189(1–2), 1–12.
- Krueger, F., Barbey, A., & Grafman, J. (2009) The medial prefrontal cortex mediates social event knowledge. *Trends Cogn. Sci.*, 13(3), 103–109.
- Lackner, J. (1974) Observations on the speech processing capabilities of an amnesic patient: several aspects of HM’s language function. *Neuropsychologia*, 12(2), 199.
- Lahnakoski, J. M., Glerean, E., Salmi, J., Jääskeläinen, I. P., Sams, M., Hari, R., et al. (2012) Naturalistic fMRI Mapping Reveals Superior Temporal Sulcus as the Hub for the Distributed Brain Network for Social Perception. *Front. Hum. Neurosci.*, 6, 233.
- Laukka, P. (2005) Categorical perception of vocal emotion expressions. *Emotion*, 5(3), 277–295.
- Laukka, P., Audibert, N., & Aubergé, V. (2012) Exploring the determinants of the graded structure of vocal emotion expressions. *Cogn. Emot.*, 26(4), 710–719.
- Laukka, P., Audibert, N., & Aubergé, V. (2007) Graded structure in vocal expression of emotion: What is meant by ‘prototypical expressions’? In, *WS Paralinguistic speech - between models and data*. Saarbrücken, p. 1–4.

## References

---

- Laukka, P., Neiberg, D., Forsell, M., Karlsson, I., & Elenius, K. (2011) Expression of affect in spontaneous speech: Acoustic correlates and automatic detection of irritation and resignation. *Comput. Speech Lang.*, 25(1), 84–104.
- Lazarus, R. (1991) *Emotion and Adaptation*. Oxford University Press, New York.
- LeDoux, J. E. (2000) Emotion Circuits in the Brain. *Annu. Rev. Neurosci.*, 23(1), 155–184.
- Levine, T., Park, H. S., & McCornack, S. (1999) Accuracy in detecting truths and lies: Documenting the ‘veracity effect’. *Commun. Monogr.*, 66(2), 125–144.
- Li, Y., Li, F., Chen, J., & Li, H. (2012) An ERP study on the understanding of the distinction between real and apparent emotions. *Neurosci. Lett.*, 529(1), 33–38.
- Linville, S. E. (1985) Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females. *J. Acoust. Soc. Am.*, 78(1), 40.
- Lohmann, G., Müller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., et al. (2001) Lipsia—a new software system for the evaluation of functional magnetic resonance images of the human brain. *Comput. Med. Imaging Graph.*, 25(6), 449–457.
- Luce, R. (1959) *Individual choice behavior*. Wiley, New York.
- Luce, R. D. (1963) A threshold theory for simple detection experiments. *Psychol. Rev.*, 70, 61–79.
- Lunn, D., Thomas, A., Best, N., & Spiegelhalter, S. (2000) WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Stat. Comput.*, 10, 325–337.
- MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000) Dissociating the Role of the Dorsolateral Prefrontal and Anterior Cingulate Cortex in Cognitive Control. *Science*, 288(5472), 1835–1838.
- MacMillan, N. A. & Creelman, C. D. (2005) *Detection Theory: A User’s Guide*. Lawrence Erlbaum Assoc Inc, London.
- Markus, H. R. & Kitayama, S. (1991) Culture and the self: Implications for cognition, emotion, and motivation. *Psychol. Rev.*, 98(2), 224–253.
- Mason, W. A. & Capitano, J. P. (2012) Basic emotions: A reconstruction. *Emot. Rev.*, 4, 238–244.
- Matsumoto, D. (1992) American-japanese cultural differences in the recognition of universal facial expressions. *J. Cross-Cult. Psychol.*, 2, 372–84.
- Matsumoto, D. (1989) Cultural influences on the perception of emotion. *J. Cross-Cult. Psychol.*, 20, 92–105.
- Matsumoto, D. & Hwang, H. S. (2011) Culture and emotion: The integration of biological and cultural contributions. *J. Cross-Cult. Psychol.*, 43, 91–118.
- Matsumoto, D., Olide, A., & Willingham, B. (2009) Is There an Ingroup Advantage in Recognizing Spontaneously Expressed Emotions? *J. Nonverbal Behav.*, 33(3), 181–191.

- Matsumoto, D., Seung Hee, Y., & Fontaine, J. (2008) Mapping expressive differences around the world: The relationship between emotional display rules and individualism versus collectivism. *J. Cross-Cult. Psychol.*, 39, 55–74.
- Matsumoto, D., Takeuchi, S., Andayani, S., Kouznetsova, N., & Krupp, D. (1998) The contribution of individualism vs. collectivism to cross-national differences in display rules. *Asian J. Soc. Psychol.*, 1, 147–165.
- Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., et al. (1993) The cortical representation of speech. *J. Cogn. Neurosci.*, 5(4), 467–479.
- McClure, E. (2000) A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychol. Bull.*, 126(3), 424–453.
- Mehu, M. & Scherer, K. R. (2012) A psycho-ethological approach to social signal processing. *Cognitive processing*, 13 (Suppl 2), 397–414.
- Mesulam, M. M. (1998) From sensation to cognition. *Brain*, 121(6), 1013–1052.
- Mier, D., Lis, S., Neuthe, K., Sauer, C., Esslinger, C., Gallhofer, B., et al. (2010) The involvement of emotion recognition in affective theory of mind. *Psychophysiology*, 47(6), 1028–1039.
- Mitchell, J. P. (2005) The false dichotomy between simulation and theory-theory: the argument's error. *Trends Cogn. Sci.*, 9(8), 363–364.
- Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., et al. (1996) A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, 383(6603), 812.
- Munhall, K. G., MacDonald, E. N., Byrne, S. K., & Johnsrude, I. (2009) Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *J. Acoust. Soc. Am.*, 125(1), 384–390.
- Nelson, N. L. & Russell, J. A. (2011) Putting motion in emotion: Do dynamic presentations increase preschooler's recognition of emotion? *Cogn. Dev.*, 26(3), 248–259.
- Neumann, J. & Lohmann, G. (2003) Bayesian second-level analysis of functional magnetic resonance images. *NeuroImage*, 20(2), 1346–1355.
- Neumann, R. & Strack, F. (2000) 'Mood contagion': The automatic transfer of mood between persons. *J. Pers. Soc. Psychol.*, 79(2), 211–223.
- Norris, D. (2000) Reduced power multislice MDEFT imaging. *J. Magn. Reson. Imaging*, 11(4), 445–451.
- O'Brien, M., Miner Weaver, J., Nelson, J. A., Calkins, S. D., Leerkes, E. M., & Marcovitch, S. (2011) Longitudinal associations between children's understanding of emotions and theory of mind. *Cogn. Emot.*, 25, 1074–1086.
- Ochsner, K. & Gross, J. (2005) The cognitive control of emotion. *Trends Cogn. Sci.*, 9(5), 242–249.

## References

---

- Ogawa, S., Menon, R. S., Tank, D. W., Kim, S. G., Merkle, H., Ellermann, J. M., et al. (1993) Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *Biophys. J.*, 64(3), 803–812.
- Olson, I., Plotzker, A., & Ezzyat, Y. (2007) The Enigmatic temporal pole: a review of findings on social and emotional processing. *Brain*, 130(7), 1718–1731.
- On defining emotions (2012) *Emot. Rev.*, 4.
- Osnes, B., Hugdahl, K., Hjelmervik, H., & Specht, K. (2012) Stimulus expectancy modulates inferior frontal gyrus and premotor cortex activity in auditory perception. *Brain Lang.*, 121(1), 65–69.
- Van Overwalle, F. (2011) A dissociation between social mentalizing and general reasoning. *NeuroImage*, 54(2), 1589–1599.
- Van Overwalle, F. (2009) Social cognition and the brain: A meta-analysis. *Hum. Brain Mapp.*, 30(3), 829–858.
- Van Overwalle, F. & Baetens, K. (2009) Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, 48(3), 564–584.
- Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002) Rethinking individualism and collectivism: evaluation of theoretical assumptions and meta-analyses. *Psychol. Bull.*, 128(1), 3.
- Panksepp, J. (1998) *Affective Neuroscience: The Foundations of Human and Animal Emotions*. Oxford University Press, USA.
- Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., et al. (1995) Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalogr. Clin. Neurophysiol.*, 94(1), 26–40.
- Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010) Supramodal Representations of Perceived Emotions in the Human Brain. *J. Neurosci.*, 30(30), 10127–10134.
- Pell, M. (2005) Prosody–face Interactions in Emotional Processing as Revealed by the Facial Affect Decision Task. *J. Nonverbal Behav.*, 29(4), 193–215.
- Pell, M. D. & Kotz, S. A. (2011) On the Time Course of Vocal Emotion Recognition. *Plos One*, 6(11), e27256.
- Pell, M., Monetta, L., Paulmann, S., & Kotz, S. (2009) Recognizing Emotions in a Foreign Language. *J. Nonverbal Behav.*, 33(2), 107–120.
- Pereira, C. (2000) Dimensions of emotional meaning in speech. *power*, 322(418), 114.
- Phan, K., Wager, T., Taylor, S., & Liberzon, I. (2002) Functional Neuroanatomy of Emotion: A Meta-Analysis of Emotion Activation Studies in PET and fMRI. *NeuroImage*, 16(2), 331–348.
- Pinheiro, J. & Bates, D. (2009) *Mixed-Effects Models in S and S-PLUS*. Springer.

- 
- Pochon, J.-B., Riis, J., Sanfey, A. G., Nystrom, L. E., & Cohen, J. D. (2008) Functional Imaging of Decision Conflict. *J. Neurosci.*, 28(13), 3468–3473.
- Polyn, S. & Kahana, M. (2008) Memory search and the neural representation of context. *Trends Cogn. Sci.*, 12(1), 24–30.
- Povinelli, D. & Preuss, T. (1995) Theory of mind: evolutionary history of a cognitive specialization. *Trends Neurosci.*, 18(9), 418–424.
- Premack, D. & Woodruff, G. (1978) Does the Chimpanzee Have a Theory of Mind? *Behav. Brain Sci.*, 1(04), 515–526.
- Prinz, J. (2004) Which emotions are basic? In, Evans, D. and Cruse, P. (eds), *Emotion, Evolution, and Rationality*. Oxford University Press, p. 69–87.
- Puce, A. & Perrett, D. (2003) Electrophysiology and brain imaging of biological motion. *Philos. Trans. R. Soc. B Biol. Sci.*, 358(1431), 435–445.
- R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rakoczy, H. (2009) Done wrong or said wrong? Young children understand the normative directions of fit of different speech acts. *Cognition*, 113(2), 205–212.
- Rakoczy, H. & Tomasello, M. (2006) Two-year-olds grasp the intentional structure of pretense acts. *Dev. Sci.*, 9(6), 557–564.
- Rakoczy, H., Tomasello, M., & Striano, T. (2004) Young Children Know That Trying Is Not Pretending: A Test of the ‘Behaving-As-If’ Construal of Children’s Early Concept of Pretense. *Dev. Psychol.*, 40(3), 388–399.
- Ravizza, S. M., McCormick, C. A., Schlerf, J. E., Justus, T., Ivry, R. B., & Fiez, J. A. (2006) Cerebellar damage produces selective deficits in verbal working memory. *Brain*, 129(2), 306–320.
- Ray, R. D. & Zald, D. H. (2012) Anatomical insights into the interaction of emotion and cognition in the prefrontal cortex. *Neurosci. Biobehav. Rev.*, 36(1), 479–501.
- Redcay, E. (2008) The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neurosci. Biobehav. Rev.*, 32(1), 123–142.
- Regenbogen, C., Schneider, D. A., Gur, R. E., Schneider, F., Habel, U., & Kellermann, T. (2012) Multimodal human communication — Targeting facial expressions, speech content and prosody. *NeuroImage*, 60(4), 2346–2356.
- Reiss, D. & Marino, L. (2001) Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. *Proc. Natl. Acad. Sci.*, 98(10), 5937–5942.
- Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004) The role of the medial frontal cortex in cognitive control. *Sci. Signal.*, 306(5695), 443.

## References

---

- Rigoulot, S. & Pell, M. D. (2012) Seeing Emotion with Your Ears: Emotional Prosody Implicitly Guides Visual Attention to Faces. *Plos One*, 7(1), e30740.
- Rilliard, A., Shochi, T., Martin, J.-C., Erickson, D., & Auberge, V. (2009) Multimodal Indices to Japanese and French Prosodically Expressed Social Affects. *Lang. Speech*, 52(2-3), 223–243.
- Roberson, D., Damjanovic, L., & Kikutani, M. (2010) Show and Tell: The Role of Language in Categorizing Facial Expression of Emotion. *Emot. Rev.*, 2(3), 255 –260.
- Rogan, M. T., Stäubli, U. V., & LeDoux, J. E. (1997) Fear conditioning induces associative long-term potentiation in the amygdala. *Nature*, 390(6660), 604.
- Rolls, E. (1999) The functions of the orbitofrontal cortex. *Neurocase Neural Basis Cogn.*, 5(4), 301.
- Romani, G., Williamson, S., & Kaufman, L. (1982) Tonotopic organization of the human auditory cortex. *Science*, 216(4552), 1339–1340.
- Roseman, I. J. (1984) Cognitive determinants of emotion: A structural theory. *Rev. Pers. Soc. Psychol.*, 5, 11-36.
- Ross, E. D. (2011) Affective prosody: What do comprehension errors tell us about hemispheric lateralization of emotions, sex and aging effects, and the role of cognitive appraisal. *Neuropsychologia*, 49(5), 866–877.
- Ross, E. D., Edmondson, J. A., Seibert, G. B., & Homan, R. W. (1988) Acoustic analysis of affective prosody during right-sided Wada test: a within-subjects verification of the right hemisphere's role in language. *Brain Lang.*, 33(1), 128–145.
- Russell, J. (1980) A circumplex model of affect. *J. Pers. Soc. Psychol.*, 39, 1161–1178.
- Russell, J. A., Bachorowski, J.-A., & Fernández-Dols, J.-M. (2003) Facial and Vocal Expressions of Emotion. *Annu. Rev. Psychol.*, 54(1), 329–349.
- Russell, J. & Barrett, L. (1999) Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *J. Pers. Soc. Psychol.*, 76(5), 805–819.
- Sabbagh, M. A. (2004) Understanding orbitofrontal contributions to theory-of-mind reasoning: Implications for autism. *Brain Cogn.*, 55(1), 209–219.
- Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004) Left temporoparietal junction is necessary for representing someone else's belief. *Nat. Neurosci.*, 7(5), 499–500.
- Sauter, D. A. & Eisner, F. (2013) Commonalities outweigh differences in the communication of emotions across human cultures. *Proc. Natl. Acad. Sci.*, 110(3), E180–E180.
- Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010) Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proc. Natl. Acad. Sci.*, 107(6), 2408 –2412.
- Saxe, R. (2006) Uniquely human social cognition. *Curr. Opin. Neurobiol.*, 16(2), 235–239.

- Schachter, S. & Singer, J. (1962) Cognitive, social, and physiological determinants of emotional state. *Psychol. Rev.* *69*(5), 379.
- Scheiner, E. & Fischer, J. (2011) Emotion expression - the evolutionary heritage in the human voice. In, Welsch, W. (ed), *Interdisciplinary anthropology: The continuing evolution of man*. Springer, Heidelberg & New York.
- Scherer, K. (2000) Psychological Models of Emotion. In, *The Neuropsychology of Emotion*. Oxford University Press.
- Scherer, K. (1986) Vocal affect expression: A review and model for future research. *Psychol. Bull.*, *99* (2), 143–165.
- Scherer, K. (1991) Vocal cues in emotion encoding and decoding. *Motiv. Emot.*, *15*(2), 123–148.
- Scherer, K., Banse, R., & Wallbott, H. (2001) Emotion inferences from vocal expression correlate across languages and cultures. *J. Cross-Cult. Psychol.*, *32*(1), 76.
- Scherer, K., Clark-Polner, E., & Mortillaro, M. (2011) In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *Int. J. Psychol.*, *46*(6), 401–435.
- Scherer, K. & Ellgring, H. (2007) Multimodal expression of emotion: Affect programs or componential appraisal patterns. *Emotion*, *7*(1), 158–171.
- Scherer, K. R. (1984) On the nature and function of emotion: a component process approach. In, Scherer, K. R. and Ekman, P. (eds), *Approaches to emotion*. Erlbaum, Hillsdale, NJ, p. 293–318.
- Scherer, K. R. (2013) Vocal markers of emotion: Comparing induction and acting elicitation. *Comput. Speech Lang.*, *27*(1), 40–58.
- Scherer, K. R. (2005) What are emotions? And how can they be measured? *Soc. Sci. Inf.*, *44*(4), 695–729.
- Scherer, K. R. & Wallbott, H. G. (1994) Evidence for universality and cultural variation of differential emotion response patterning. *J. Pers. Soc. Psychol.*, *66*, 310–328.
- Schilbach, L., Eickhoff, S. B., Rotarska-Jagiela, A., Fink, G. R., & Vogeley, K. (2008) Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the 'default system' of the brain. *Conscious. Cogn.*, *17*(2), 457–467.
- Schirmer, A. (2010) Mark My Words: Tone of Voice Changes Affective Word Representations in Memory. *Plos One*, *5*(2), e9080.
- Schirmer, A. & Kotz, S. (2006) Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn. Sci.*, *10*(1), 24–30.
- Schirmer, A., Kotz, S., & Friederici, A. (2005) On the role of attention for the processing of emotions in speech: Sex differences revisited. *Cogn. Brain Res.*, *24*(3), 442–452.

## References

---

- Schirmer, A., Lui, M., Maess, B., Escoffier, N., Chan, M., & Penney, T. (2006) Task and sex modulate the brain response to emotional incongruity in Asian listeners. *Emotion*, 6(3), 406–417.
- Schmahmann, J. D. & Caplan, D. (2006) Cognition, emotion and the cerebellum. *Brain*, 129(2), 290–292.
- Schmidt, K. L. & Cohn, J. F. (2001) Human facial expressions as adaptations: Evolutionary questions in facial expression research. *Am. J. Phys. Anthropol.*, 116, 3–24.
- Schmidt, M. E., Pempek, T. A., Kirkorian, H. L., Lund, A. F., & Anderson, D. R. (2008) The Effects of Background Television on the Toy Play Behavior of Very Young Children. *Child Dev.*, 79(4), 1137–1151.
- Schötz, S. (2007) Acoustic analysis of adult speaker age. *Speak. Classif.*, 88–107.
- Sebastian, C. L., Fontaine, N. M. G., Bird, G., Blakemore, S.-J., Brito, S. A. D., McCrory, E. J. P., et al. (2012) Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Soc. Cogn. Affect. Neurosci.*, 7(1), 53–63.
- Seymour, B. & Dolan, R. (2008) Emotion, Decision Making, and the Amygdala. *Neuron*, 58(5), 662–671.
- Shafer, A. T., Matveychuk, D., Penney, T., O’Hare, A. J., Stokes, J., & Dolcos, F. (2012) Processing of Emotional Distraction Is Both Automatic and Modulated by Attention: Evidence from an Event-related fMRI Investigation. *J. Cogn. Neurosci.*, 24(5), 1233–1252.
- Shamay-Tsoory, S. G. (2011) The neural bases for empathy. *Neuroscientist*, 17(1), 18–24.
- Shany-Ur, T., Poorzand, P., Grossman, S. N., Growdon, M. E., Jang, J. Y., Ketelle, R. S., et al. (2012) Comprehension of insincere communication in neurodegenerative disease: lies, sarcasm, and theory of mind. *Cortex J. Devoted Study Nerv. Syst. Behav.*, 48(10), 1329–1341.
- Shapiro, B. E. & Danly, M. (1985) The role of the right hemisphere in the control of speech prosody in propositional and affective contexts. *Brain Lang.*, 25(1), 19–36.
- Smith, C. A. & Ellsworth, P. C. (1985) Patterns of cognitive appraisal in emotion. *J. Pers. Soc. Psychol.*, 48, 813–838.
- Smith, J. (1982) Recognition models evaluated: A commentary on Keren and Baggen. *Percept. Psychophys.*, 31, 183–189.
- Smith, J. D., Shields, W. E., & Washburn, D. A. (2004) The comparative psychology of uncertainty monitoring and metacognition. *Behav. Brain Sci.*, 26(03), 317–339.
- Sneddon, I., McKeown, G., McRorie, M., & Vukicevic, T. (2011) Cross-cultural patterns in dynamic ratings of positive and negative natural emotional behaviour. *Plos One*, 6e14679.
- Söderling, B. (1959) The First Smile A Developmental Study. *Acta Paediatr.*, 47(s117), 78–82.



- Spreng, R., Mar, R., & Kim, A. (2009) The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *J. Cogn. Neurosci.*, 21(3), 489–510.
- Van den Stock, J., Righart, R., & de Gelder, B. (2007) Body expressions influence recognition of emotions in the face and voice. *Emotion*, 7(3), 487–494.
- Stoodley, C. J. & Schmahmann, J. D. (2009) Functional topography in the human cerebellum: a meta-analysis of neuroimaging studies. *Neuroimage*, 44(2), 489–501.
- Streeter, L. A. & et al (1977) Pitch changes during attempted deception. *J. Pers. Soc. Psychol.*, 35(5), 345–350.
- Sturtz, S., Ligges, U., & Gelman, A. (2005) R2WinBUGS: A package for running WinBUGS from R. *J. Stat. Softw.*, 12, 1–16.
- Talairach, J. & Tournoux, P. (1988) Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: an approach to cerebral imaging. Thieme, New York.
- Tanner, W. J. & Swets, J. (1954) A decision-making theory of visual detection. *Psychol. Rev.*, 61(6), 401–409.
- Teufel, C., Fletcher, P. C., & Davis, G. (2010) Seeing other minds: attributed mental states influence perception. *Trends Cogn. Sci.*, 14(8), 376–382.
- Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011) Prior Expectation Mediates Neural Adaptation to Repeated Sounds in the Auditory Cortex: An MEG Study. *J. Neurosci.*, 31(25), 9118 – 9123.
- Tomasello, M. (2003) What Makes Human Cognition Unique? From Individual to Shared to Collective Intentionality. *Mind Lang.*, 18(2), 121–147.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005) Understanding and sharing intentions: The origins of cultural cognition. *Behav. Brain Sci.*, 28(5), 706-707.
- Tracy, J. L. & Prehn, C. (2011) Arrogant or self-confident? The use of contextual knowledge to differentiate hubristic and authentic pride from a single nonverbal expression. *Cogn. Emot.*, 26(1), 14–24.
- Trimbitas, O., Lin, Y., & D., C. K. (2007) Arta de a cere scuze in cultura romaneasca: Use of apology in ethnic Romanian culture. *Hum. Commun.*, 10401–420.
- Truett, A., Puce, A., & McCarthy, G. (2000) Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.*, 4(7), 267–278.
- Tucker, D. M., Watson, R. T., & Heilman, K. M. (1977) Discrimination and evocation of affectively intoned speech in patients with right parietal disease. *Neurology*, 27(10), 947–947.

## References

---

- Ugurbil, K., Garwood, M., Ellermann, J., Hendrich, K., Hinke, R., Hu, X., et al. (1993) Imaging at high magnetic fields: initial experiences at 4 T. *Magn. Reson. Q.*, 9(4), 259–277.
- Vann, S. D., Aggleton, J. P., & Maguire, E. A. (2009) What does the retrosplenial cortex do? *Nat. Rev. Neurosci.*, 10(11), 792–802.
- Vigneau, M. (2006) Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *Neuroimage*, 30(4), 1414–1432.
- Vigneau, M., Beaucousin, V., Hervé, P.-Y., Jobard, G., Petit, L., Crivello, F., et al. (2011) What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing?: Insights from a meta-analysis. *NeuroImage*, 54(1), 577–593.
- Vogt, B. A. (2005) Pain and emotion interactions in subregions of the cingulate gyrus. *Nat. Rev. Neurosci.*, 6(7), 533–544.
- Vollm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., et al. (2006) Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *Neuroimage*, 29(1), 90–98.
- Volz, K., Schubotz, R., & von Cramon, D. (2005) Variants of uncertainty in decision-making and their neural correlates. *Brain Res. Bull.*, 67(5), 403–412.
- Wagner, H. (1993) On measuring performance in category judgment studies of nonverbal behavior. *J. Nonverbal Behav.*, 17(1), 3–28.
- Warren, G., Schertler, E., & Bull, P. (2008) Detecting Deception from Emotional and Unemotional Cues. *J. Nonverbal Behav.*, 33(1), 59–69.
- Watanabe, M. & Sakagami, M. (2007) Integration of cognitive and motivational context information in the primate prefrontal cortex. *Cereb. Cortex*, 17(suppl 1), i101.
- Weiskopf, N., Hutton, C., Josephs, O., & Deichmann, R. (2006) Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: A whole-brain analysis at 3 T and 1.5 T. *NeuroImage*, 33(2), 493–504.
- Weiskopf, N., Hutton, C., Josephs, O., Turner, R., & Deichmann, R. (2007) Optimized EPI for fMRI studies of the orbitofrontal cortex: compensation of susceptibility-induced gradients in the readout direction. *Magn. Reson. Mater. Phys. Biol. Med.*, 20(1), 39–49.
- Westerhausen, R., Moosmann, M., Alho, K., Belsby, S.-O., Hämäläinen, H., Medvedev, S., et al. (2010) Identification of attention and cognitive control networks in a parametric auditory fMRI study. *Neuropsychologia*, 48(7), 2075–2081.
- Wheaton, K. J., Thompson, J. C., Syngienotis, A., Abbott, D. F., & Puce, A. (2004) Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. *NeuroImage*, 22(1), 277–288.

- Wiethoff, S., Wildgruber, D., Kreifelts, B., Becker, H., Herbert, C., Grodd, W., et al. (2008) Cerebral processing of emotional prosody—influence of acoustic parameters and arousal. *NeuroImage*, 39(2), 885–893.
- Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006) Cerebral processing of linguistic and emotional prosody: fMRI studies. *Prog. Brain Res.*, 249–268.
- Wildgruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W., et al. (2004) Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cereb. Cortex*, 14(12), 1384.
- Wildgruber, D., Pihan, H., Ackermann, H., Erb, M., & Grodd, W. (2002) Dynamic brain activation during processing of emotional intonation: influence of acoustic parameters, emotional valence, and sex. *Neuroimage*, 15(4), 856–869.
- Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., et al. (2005) Identification of emotional intonation evaluated by fMRI. *Neuroimage*, 24(4), 1233–1241.
- Williams, C. E. & Stevens, K. N. (1972) Emotions and Speech: Some Acoustical Correlates. *J. Acoust. Soc. Am.*, 52(4B), 1238–1250.
- Wiltling, J., Kraemer, E., & Swerts, M. (2006) Real vs. acted emotional speech. In, *INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing, INTERSPEECH 2006 - ICSLP*, p. 805–808.
- Wimmer, H. & Perner, J. (1983) Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103–128.
- Worsley, K. & Friston, K. (1995) Analysis of fMRI time-series revisited - again. *Neuroimage*, 2(3), 173–181.
- Wurm, L., Vakoch, D., Strasser, M., Calin-Jageman, R., & Ross, S. (2001) Speech perception and vocal expression of emotion. *Cogn. Emot.*, 15(6), 831–852.
- Wurm, M., von Cramon, D., & Schubotz, R. (2011) Do we mind other minds when we mind other minds' actions? A functional magnetic resonance imaging study. *Hum. Brain Mapp.*, 32, 2141–2150.
- Young, L., Dodell-Feder, D., & Saxe, R. (2010) What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia*, 48(9), 2658–2664.
- Zuckerman, M., Koestner, R., Colella, M. J., & Alton, A. O. (1984) Anchoring in the detection of deception and leakage. *J. Pers. Soc. Psychol.*, 47, 301–311.

## 8. List of Figures and Tables

### 8.1 Figures

Fig. 1.1 Model of bottom-up and top-down processes .....	4
Fig. 2.1 Experimental trial sequences for emotion and authenticity judgment trials .....	18
Fig. 2.2 Behavioral results from fMRI for emotion judgments .....	22
Fig. 2.3 Behavioral results from fMRI for authenticity judgments .....	22
Fig. 2.4 Analysis of emotion recognition data in terms of choice theory .....	23
Fig. 2.5 Brain activation correlates of experimental tasks .....	25
Fig. 2.6 Brain activation correlates of stimulus authenticity.....	25
Fig. 3.1 Probability of correct authenticity recognition .....	42
Fig. 3.2 Discrimination of authentic and play-acted vocal expressions .....	42
Fig. 3.3 Probability of correct emotion recognition .....	43
Fig. 3.4 Analysis of emotion recognition data by bias .....	45
Fig. 3.5 Analysis of emotion recognition data by dissimilarity.....	45
Fig. 4.1 Experimental trial sequences for uncued and cued trials .....	56
Fig. 4.2 Behavioral results from fMRI split by cue .....	60
Fig. 4.3 Behavioral results from fMRI (not split by cue).....	61
Fig. 4.4 Brain activation correlates of authenticity .....	62
Fig. 4.5 Brain activation correlates of cue and stimulus interaction .....	62

---

Fig. 5.1 F0 contour and spectrogram .....	72
Fig. 5.2 Mean contour for responses to authenticity task .....	78
Fig. 5.3 Mean contour for responses to emotion task.....	78
Fig. 5.4 Brain activation correlates of emotion tasks of study 2 .....	80
Fig. 5.5 Brain activation correlates of emotion and authenticity tasks .....	80
Fig. 5.6 Mean and 95% confidence interval of BOLD signal measure.....	81

## 8.2 Tables

Table 2.1 Recognition Accuracy for Emotion and Authenticity Trials .....	21
Table 2.2 Reaction times by stimulus type .....	23
Table 2.3 Brain activation correlates of judging authenticity .....	23
Table 2.4 Brain activation correlates of stimulus authenticity.....	24
Table 3.1 Post-hoc tests .....	44
Table 4.1 Emotion categorization accuracy and reaction time .....	60
Table 4.2 Brain activation correlates of stimulus authenticity.....	63
Table 4.3 Brain activation correlates of stimulus authenticity during cued trials .....	63
Table 5.1 Mean values of high and low stimulus groups .....	77
Table 5.2 Brain activation correlates of stimulus authenticity during cued trials .....	79

## 9. Acknowledgments

I want to thank Julia Fischer for supervising my doctoral thesis and supporting me in my venture in this very broad research topic. In particular I want to thank Julia for her enthusiasm in working with me, even when I was just an email correspondence from Canada, and for the rapid integration into the lab group at the “Deutsches Primatenzentrum” all those years ago. Discussions, whether based on data or philosophical questions, were both analytical and accepting of novel ideas, while Julia’s scientific competence was invaluable in dealing with the inevitable roadblocks along the way.

I also want to thank Ricarda Schubotz for her support and advice towards making all this possible, right from the very beginning with my Diploma project. Ricarda’s practical and conceptual advice on all matters fMRI and the intricacies of the human brain were invaluable to my training and research as a neuroscientist. I was lucky enough to be a member of not one, but two great research teams in three separate institutions.

I thank the Neurosenses Scholarship initiative for providing me with the financial support needed to work on my PhD project. I would also like to thank Christiane Thiel and Jens Frahm for their support as members of my Thesis Committee.

I want to thank the whole Cognitive Ethology lab, not only for great discussions and support throughout my doctoral project, but in the years prior as well. I want to thank Rebecca, Laura, Vanessa, and Christian for the great discussions during the trips to the cafeteria and evening card games. Special thanks go to Tabitha, Gisela, Rebecca, Brandon and Barbara for all the so-necessary lunch breaks and, along with Philip and the whole PhD± group, for the diversity of social events needed to balance out work-life. I want to thank Nadine, Mechthild, and Ludwig for all the IT and bureaucratic help! Last but not least, I want to thank the Simenti team, including Peter, Urs, Annika, Matthias, Tanja and Adeelia for making me feel “like I was there”.

I want to thank the Motor Cognition and Biological Psychology teams, in Cologne and Münster respectively. I particularly want to thank Moritz Wurm, Anne Kühn, Marike Schiffer, and Katja Kornysheva for integrating me into this second great team and the intense discussions on all matters from neuroscience to politics. Of no less importance were all the methodological help on the fMRI and programming fronts! I also want to thank Christiane Ahlheim and Mari Hrkac who helped get me settled and integrated in the second-generation group in Münster.

I particularly want to thank my family. I want to thank my sisters, Målin and Hannah, for their caring support through thick and thin, and my parents, Imme and Jacques, for providing me with such amazing experiences throughout my life and, especially, for always challenging the *status quo*.



---



---

---