The impact of vocal expressions on the understanding of affective states in others

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

“Doctor rerum naturalium“

der Georg-August-Universität Göttingen

im Promotionsprogramm Biologie

der Georg-August University School of Science (GAUSS)

vorgelegt von

Rebecca Jürgens

aus Hildesheim

Göttingen, 2014

Betreuungsausschuss

1.Betreuerin: Prof. Dr. Julia Fischer, Kognitive Ethologie, Deutsches Primatenzentrum, Göttingen

2.Betreuer: Prof. Dr. Hannes Rakoczy, Biologische Entwicklungspsychologie, G.-E.-M.-I. Institut für Psychologie, Universität Göttingen

Anleiterin: Prof. Dr. Annekathrin Schacht, NWG Experimentelle Psycholinguistik, CRC Textstrukturen, Universität Göttingen

Mitglieder der Prüfungskommission

Referentin: Prof. Dr. Julia Fischer, Kognitive Ethologie, Deutsches Primatenzentrum, Göttingen

Korreferent: Prof. Dr. Hannes Rakoczy, Biologische Entwicklungspsychologie, G.-E.-M.-I. Institut für Psychologie, Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Annekathrin Schacht, NWG Experimentelle Psycholinguistik, CRC Textstrukturen, Universität Göttingen

Prof. Dr. Margarete Boos, Sozial-und Kommunikationspsychologie, G.-E.-M.-I. Institut für Psychologie, Universität Göttingen

Dr. Igor Kagan, Kognitive Neurowissenschaften, Deutsches Primatenzentrum, Göttingen

Dr. Bernhard Fink, Biologische Persönlichkeitspsychologie, G.-E.-M.-I. Institut für Psychologie, Universität Göttingen

Tag der mündlichen Prüfung: 24.11.2014

Table of Contents

I. Summary

Understanding emotions of social partners is of fundamental importance in day-to-day life. Humans share their affective states and intentions not only by language, but also by facial expressions, body posture or tone of voice. Nonverbal emotional expressions are specified as being part of an emotional episode, which additionally consists of action tendencies including underlying autonomic changes and subjective feelings. Although the communication of emotions has been studied for decades, our understanding of what exactly is communicated and how it is processed is still far from complete. Emotional expressions are frequently produced to fulfil social requirements calling into question the reliability to infer actual emotional states. As recognizing expressions that accompany underlying emotions would be of relevance for subsequent behavior, the ability to detect emotional deception seems to be essential in emotional communication. Especially vocal expressions seem to be promising for revealing underlying emotions, due to the strong autonomic innervation of the vocal tract.

Moreover, the recognition of emotions has been found not to be invariable but to depend on the speaker-listener relation. Sharing group membership, for example, positively affected emotion recognition, which might be caused by an attention-shift towards people of higher relevance but also by facilitated empathic concern. Successfully understanding others' emotions is closely linked to mirroring or simulating the perceived emotion internally. Research in the field of affective neurosciences could demonstrate a shared neural network during attending and experiencing emotions, which is influenced by the social relevance of the individual showing the expression. The extent to which affect sharing is necessary on the behavioral level to recognize emotional expressions and whether it is positively affected by increased speaker's relevance, is still debated.

In this thesis, I investigated vocal emotion expressions, with the objectives to first understand the relation between spontaneous and play-acted expressions and second to broaden

our knowledge about the importance of affect sharing and speaker's relevance on emotion recognition.

In the first part of this thesis, I compared the recognition of spontaneous and play-acted vocal expressions in a cross-cultural study. In contrast to spontaneous expressions, acted ones were assumed to be influenced by social codes and were therefore less accurately recognized in cultures other than the culture of origin. Alternatively, emotion recognition for both conditions might rest on a universal basis. This cross-cultural comparison was conducted using 80 spontaneous vocal expressions, recorded in emotional situations by a German radio station and the re-enactments by professional actors. Short excerpts of these speech tokens were presented to participants in Germany, Romania and Indonesia with the tasks to indicate the expressed emotion and the authenticity. Generally, participants were poor in distinguishing the encoding condition and German listeners were more accurate in both tasks, independent whether the expression was play-acted or not. Emotion recognition showed a comparable pattern across all cultures, speaking for a universal basis for both encoding conditions. Recognition accuracy for all emotions was low and authenticity affected only anger, which was more frequently recognized when play-acted and sadness, which was more accurately recognized when spontaneous.

In the second part of this thesis, I aimed to understand the source of these differences and to disclose the importance of acting training on the credibility of emotion depiction. I added vocal expressions of acting-inexperienced people to the comparison, and conducted an additional acoustic analysis. Professional actors were predicted to be more suited to produce credible emotion expressions than inexperienced speaker. This was not confirmed, as professionally acted expressions were even more frequently recognized as being play-acted than the ones by inexperienced people. For professional actors I found the same pattern in the emotion recognition as in the cross-cultural study; while expressions by non-experienced speakers only deviated from the spontaneous ones by less accurate sadness recognition. Acoustically, the main difference was that acted expressions had a more vivid speech melody than the spontaneous

ones. Both studies demonstrated a complex, universal interaction between emotion recognition and stimulus authenticity. Acted expressions were only poorly detected and not more stereotyped, and it was shown that acting inexperienced people were more suited to produce vocal expressions that resemble spontaneous ones than were professional actors.

In the third part I focused on investigating the processes of recognizing the emotions of others. To this aim, I experimentally manipulated biographical similarity between fictitious speakers and the listener. I predicted that vocal expressions spoken by the more similar character would be recognized more accurately due to the increased social relevance of the speaker. In order to disclose the impact of affect sharing on emotion recognition, I additionally measured skin conductance responses (SCR) and pupil size, which account for autonomic reactions, while participants judged joyful, angry and neutral vocal expressions. Similarity affected neither the emotion recognition nor the autonomic measurements. Overall, emotional expressions did not trigger arousal related SCR, but emotion-related responses in pupil size. This finding indicates that affective processing does not involve the whole autonomic system and is not an essential component of recognizing emotions, at least when people only attend to vocal expression. Similarity might presumably affect emotion recognition in a more lifelike situation in which an actual tie can be established between both partners, not in a merely artificial setting. Empathic reactions presumably need a more holistic approach to be effective.

My thesis concentrated on the understanding of emotional communication by regarding vocal expressions and I could show that attending to single emotion expressions is not sufficient to reveal the actual affective state of the sender in terms of differentiating acted from spontaneous expressions. Additionally, I demonstrated that vocal expressions do not evoke strong autonomic reactions in the listener. The communication of vocal emotion expression seemingly rests more on cognitive than on affective processing.

II. Zusammenfassung

Ein wichtiger Aspekt des täglichen sozialen Lebens ist das Erkennen von emotionalen Zuständen in unserem Gegenüber. Unsere Emotionen und Intentionen teilen wir nicht nur durch sprachliche Äußerungen mit, sondern auch über die Mimik, Körpersprache und den Tonfall in der Stimme. Diese nichtverbalen, emotionalen Ausdrücke sind Bestandteile einer Emotion, zu denen darüber hinaus das subjektive Empfinden, die Handlungsbereitschaft und die damit zusammenhängenden physiologischen Reaktionen gehören. Obwohl die emotionale Kommunikation schon seit Jahrzehnten im Fokus der Wissenschaft liegt, ist noch unklar, welche Bestandteile einer Emotion genau kommuniziert und wie diese Informationen verarbeitet werden. Zudem spielen emotionale Ausdrücke eine wichtige Rolle in sozialen Interaktionen und werden häufig bewusst verwendet, um sozial-angepasstes Verhalten zu zeigen. Damit ist ihre Reliabilität, die tatsächliche Gefühlswelt des Gegenübers wiederzugeben, fraglich. Das Erkennen von Emotionsausdrücken, die auf empfunden Emotionen basieren ist jedoch von enormer Wichtigkeit für die nachfolgenden Handlungen. Deswegen sollte die Fähigkeit, empfundene von gespielten Emotionen unterscheiden zu können, essentiell sein. Da vokale Ausdrücke durch Einflüsse des autonomen Nervensystems auf den Vokaltrakt gebildet werden, sind diese als besonders vielversprechend anzusehen, um zugrundeliegende emotionale Zustände aufzudecken.

Die Erkennung von Emotionen im Gegenüber ist nicht unveränderlich, sondern hängt unter anderem auch von der Beziehung zwischen dem Sprecher und dem Zuhörer ab. So konnte in einer früheren Studie gezeigt werden, dass bei Personen, die derselben Gruppe angehören, Emotionen besser erkannt werden konnten. Dieser Effekt lässt sich einerseits mit einer Aufmerksamkeitsverschiebung hin zu Personen mit erhöhter sozialer Relevanz deuten. Andererseits gibt es Erklärungsansätze, die auf eine erhöhte Bereitschaft für empathische Reaktionen hinweisen. Erfolgreiches Verstehen von Emotionen wird in der Forschungsliteratur eng mit dem Spiegeln oder dem Simulieren der wahrgenommen Emotion verknüpft. Die

affektiven Neurowissenschaften zeigten bisher ein gemeinsames neuronales Netzwerk, welches aktiv ist, wenn Personen eine Emotion bei anderen wahrnehmen oder selber empfinden. Die neurale Aktivität in diesem Netzwerk wird zudem von der sozialen Relevanz der Person beeinflusst, welche die Emotion zeigt. Welches Ausmaß das Wiederspiegeln einer Emotion auf der Verhaltensebene hat um eine Emotion zu erkennen ist hingegen noch ungeklärt. Auch die Frage nach dem Einfluss des Sprechers auf die empathische Reaktion ist noch nicht abschließend geklärt.

In dieser Arbeit untersuchte ich vokale Emotionsausdrücke und versuchte zunächst das Verhältnis zwischen gespielten und spontanen Ausdrücken zu verstehen. Anschließend konzentrierte ich mich auf die Frage, welche Bedeutung das Teilen einer Emotion und die Relevanz des Sprechers auf die Emotionserkennung haben. Im ersten Teil dieser Arbeit verglich ich die Wahrnehmung von spontanen und gespielten vokalen Ausdrücken in einer interkulturellen Studie. Im Gegensatz zu spontanen Ausdrücken wurde angenommen, dass gespielte Ausdrücke vermehrt auf sozialen Codes basieren und daher von Hörern anderer Kulturen als der Herkunftskultur weniger akkurat erkannt werden. Alternativ könnte die Emotionserkennung beider Bedingungen universell sein. Dieser interkulturelle Vergleich wurde anhand von 80 spontanen Emotionsausdrücken durchgeführt, die von Menschen aufgenommen wurden, welche sich in emotionalen Situationen befanden. Die gespielten Stimuli bestanden aus den nachgespielten Szenen, die von professionellen Schauspielern eingesprochen worden. Kurze Sequenzen dieser Ausdrücke wurden Versuchspersonen in Deutschland, Rumänien und Indonesien vorgespielt. Die Versuchspersonen erhielten die Aufgabe anzugeben, welche Emotion dargestellt wurde und ob der Ausdruck gespielt oder echt war. Im Ganzen konnten die Versuchspersonen nur unzureichend angeben, inwieweit ein Ausdruck gespielt war. Deutsche Hörer waren in beiden Aufgaben besser als die Hörer der anderen Kulturen. Dieser Vorteil war unabhängig von der Authentizität des Stimulus. Die Emotionserkennung zeigte ein vergleichbares Muster in allen Kulturen, was für eine universelle Grundlage der Emotionserkennung spricht. Die

Erkennungsraten im Allgemeinen waren schwach ausgeprägt und ob ein Ausdruck gespielt oder echt war, beeinflusste lediglich die Erkennung von den Emotionen Ärger und Trauer. Ärger wurde besser erkannt wenn er gespielt war und Trauer wenn sie echt war.

Der zweite Teil meiner Arbeit beschäftigte sich mit der Ursache für die oben erwähnten Unterschiede in der Emotionserkennung und untersuchte, welchen Einfluss Schauspieltraining auf die Glaubwürdigkeit der Emotionsdarstellung hat. Zu diesem Zweck erweiterte ich den Stimulus-Korpus um Emotionsausdrücke, die von schauspiel-unerfahrenen Sprechern eingesprochen wurden. Zusätzlich zu der Bewertungsstudie führte ich eine akustische Analyse der Sprachaufnahmen durch. Es wurde vorhergesagt, dass professionelle Schauspieler besser geeignet seien als schauspiel-unerfahrene Sprecher, um glaubwürdig Emotionsausdrücke zu generieren. Diese Vorhersage konnte jedoch nicht bestätigt werden. Die Ausdrücke der professionellen Schauspieler wurden im Gegenteil sogar häufiger als gespielt wahrgenommen als die der unerfahrenen Sprecher. Für die professionellen Sprecher konnte ich das Muster in der Emotionserkennung, welches sich in der interkulturellen Studie zeigte, replizieren. Die Ausdrücke der unerfahrenen Sprecher hingegen wichen nur in den geringeren Erkennungsraten für Trauer von den spontanen Ausdrücken ab. Der Haupteffekt der akustischen Analyse bestand in einer lebhafteren Sprachmelodie der gespielten Ausdrücke.

Im dritten Teil der Arbeit untersuchte ich den Prozess der Emotionserkennung. Zu diesem Zweck manipulierte ich in einem Experiment die biographische Ähnlichkeit zwischen fiktiven Sprechern und dem Hörer. Auf Grund der höheren Relevanz eines ähnlichen Sprechers, sollten emotionale Ausdrücke in der ähnlichen Bedingung besser erkannt werden als in der unähnlichen. Um den Einfluss des gemeinsamen Erlebens einer Emotion auf die Emotionserkennung festzustellen, zeichnete ich außerdem die Hautleitfähigkeit und die Pupillenveränderung auf, welches beides Marker für Reaktionen des autonomen Nervensystems sind. Währenddessen wurden den Versuchspersonen ärgerliche, freudige und neutrale vokale Ausdrücke präsentiert, welche sie zu bewerten hatten. Ähnlichkeit hatte weder einen Einfluss auf die

Emotionserkennung noch auf die peripher-physiologischen Messungen. Die Versuchspersonen zeigten keine Reaktionen der Hautleitfähigkeit auf vokale Ausdrücke. Die Pupille hingegen reagierte emotionsabhängig. Diese Befunde deuten darauf hin, dass die affektive Verarbeitung nicht das gesamte autonome Nervensystem miteinschließt, zumindest nicht, wenn lediglich die Stimme verarbeitet wird. Das Teilen einer Emotion scheint demnach kein notweniger Bestandteil des Verstehens oder der Erkennung zu sein. Die Ähnlichkeit zwischen Sprecher und Hörer könnte die Emotionsverarbeitung in einer lebensnahen Umgebung beeinflussen, in der eine persönliche Verbindung zwischen beiden Interaktionspartnern möglich ist, nicht hingegen in einer mehrheitlich artifiziellen Manipulation. Empathische Reaktionen brauchen um wirksam zu werden einen ganzheitlicheren Ansatz.

Meine Arbeit konzentrierte sich auf das Verständnis von emotionaler Kommunikation in Bezug auf vokale Emotionsausdrücke und konnte zeigen, dass das bewusste Hören einzelner, kontextfreier Emotionsausdrücke nicht ausreichend ist um auf tatsächliche emotionale Zustände rückschließen zu können. Dies wird durch die fehlende Differenzierung von gespielten und spontanen Emotionsausdrücken deutlich. Darüber hinaus konnte ich aufzeigen, dass vokale Emotionsausdrücke im Hörer keine starken Reaktionen des autonomen Nervensystems auslösen. Die Kommunikation mittels vokaler emotionaler Ausdrücke scheint daher vermehrt auf kognitiven als auf affektiven Prozessen zu basieren.

1 General Introduction

Humans share their inner states not only by language, but also by facial expression, body posture as well as by their tone of voice. Hardly any social interaction takes place without using these nonverbal behaviors to communicate emotions (Vrana & Rollock, 1998). Although the use of these expressions is a common process in human life, emotional communication is far from being understood. This thesis concentrates on vocal emotion expressions, with the objectives first to disclose the relation between spontaneous and play-acted expressions in order to investigate the human ability to distinguish between "true" and "deceptive" expressions; and second to reveal the impact of social connectedness between speaker and listener and the importance of affect sharing on the understanding of emotions in others. In this general introduction I give an overview on emotions, describe the mechanisms that underlie vocal expressions, and the processes of understanding others' affective states. I then summarize the knowledge on emotional deceptive behavior, before describing the aims of the thesis.

1.1 Emotions

1.1.1 Definition of emotions

"One of the mysteries of psychology is how it has been possible to define and construe emotion in such apparently incompatible ways […]" (Russell, 2003; p. 167)

Research on emotions started more than 100 years ago (e.g., Darwin, 1872; James, 1884), but a valid definition has yet not been agreed upon (Averill, 1980; Ekman, 1999; Mulligan & Scherer, 2012; Russell, 2003; Scarantino, 2012; K. R. Scherer, 1984). It would go beyond the scope of this thesis to give an extensive overview of the debate on emotion theories. In their encyclopedia entry on emotion definition in *the Oxford Companion to Emotion and the Affective Sciences*, Frijda and Scherer (2009) summarized four aspects that are included in every emotion concept; 1) the event that elicits the emotion is of relevance for the individual's well-being, 2) the

emotional reaction is evolved to prepare the individual for action (motivational aspect), 3) there are reactions of the motor and somatovisceral system to support the action, and 4) the emotion demands priority in behavior. In the following paragraph, I give a summary of the three main directions that dominate emotion research.

Basic emotion theorists (going back to Darwin, 1872) propose the existence of a distinct set of emotions with prototypical characteristics (e.g., physiological reactions and expressive behavior), regulated by a central organizing mechanism (Ekman, 1999; Izard, 1992; Levenson, 2011). These emotions are universal, innate and evolved to deal with fundamental life tasks. The number of basic emotions and their composition is however not consistent across theorists, ranging from about two to eleven (see Ortony & Turner, 1990). On the other hand, representatives of the psychological constructionists approach (Barrett, 2009; Lindquist, Siegel, Quigly, & Barrett, 2013; Russell, 2003, going back to James, 1884) argue that our experience of distinct emotional categories emerge from an intrinsic affective state of arousal and valence (also named core affect) combined with a mental conceptualization of the emotion and is thus psychologically constructed. According to this approach, affective states are based on dimensional scales (the intrinsic physiological states), while their classification is categorical. Emotion categories do not represent specific mental entities, but classify a broad range of different states, meaning that for instance “fear” might be experienced quite differently across situations (see also Pinker, 1997 p. 387: "Fear is probably several emotions"). Lindquist et al. (2013) metaphorically described the dispute between basic emotion theorists and constructionists as the “hundred-year emotion war” that is still not settled. Finally, appraisal theories (going back to Arnold, 1960; Lazarus, Averill, & Opton, 1970) describe emotions as flexible processes rather than distinct mental states and focus on the cognitive evaluation of the situation (Ellsworth & Scherer, 2003; Moors, Ellsworth, Scherer, & Frijda, 2013; K. R. Scherer, 1984; Smith & Ellsworth, 1985). This assumption is in contrast to the constructionist view, in which categorization is done by evaluating the internal state. The situation is appraised on a variety of dimensions such as novelty, intrinsic

pleasantness and goal significance, leading to an infinite number of possible emotional episodes, in contrast to the basic emotion point of view. Classification with an emotional label (such as fear) is done by summarizing the appraisal pattern; e.g. “fear” is elicited when novelty and goal significance are high, while intrinsic pleasantness and coping potential are low (Ellsworth & Scherer, 2003). The emotional episodes consist of a variety of components, flexibly adapting towards reappraisal of the situation (Moors et al., 2013; K. R. Scherer, 2009). Next to the appraisal (that is not necessarily conscious, see Ellsworth & Scherer, 2003; Mortillaro, Mehu, & Scherer, 2013), further components are the motivational component including action tendencies, the somatic component including peripheral physiological reactions, the subjective feeling and the motor behavior including emotional expressions (Moors et al., 2013). Causal relations between the components are unsettled.

1.1.2 Emotional expressions

Emotions can be expressed via the face, the voice and the body. While facial and bodily expressions are strongly based on activation of the somatic nervous system (SNS), vocal expressions are affected to a large extent by the autonomic nervous system (ANS) that is not under voluntary control (Rinn, 1984; K. R. Scherer, 1986). In addition, the various expression channels – face, body, or voice - seem to have a different significance on emotion recognition (Regenbogen, Schneider, Finkelmeyer, et al., 2012). Despite the differences, discussing vocal expression cannot be completely done without mentioning facial and gestural expression, at least for comparative reasons. I will first give a detailed overview about emotional expression in the acoustic domain and will afterwards briefly summarize knowledge on facial and gestural expressions.

1.1.2.1 Vocal expressions

The voice is affected by a variety of physiological changes caused by the ANS and the SNS that influence the structure of our vocal tract (K. R. Scherer, 1986). Emotional expressions normally underlay spoken language; hence emotion-based acoustic changes interact with phoneme based differences. The phonemes that are mostly analyzed for emotional reasons are vowels, as they have stable acoustic characteristics and are produced via phonation. The process of phonation includes the activation of the vocal folds and is strongly influenced by peripheral physiological activity, such as respiration or muscle tone. Vowels are created according to the source filter model of speech production (see Fitch, 2000; Kent & Read, 1992). Air from the lungs is pressed through the glottis (the source). This leads to the vibration of the vocal folds, which causes the air flow to oscillate at a specific frequency (called the fundamental frequency or pitch) and at its multiple integers (the harmonics). These acoustic frequencies then pass the resonance structures of the vocal tract (the filter; including the pharynx, the throat, and the nasal and oral cavities) in which they are filtered or enhanced; this process is called articulation. Vowels are thus characterized by a fundamental frequency (ranging from approximately 50 Hz to 150 Hz in men and from 150 Hz to 250 Hz in women) and frequency regions in the harmonics with high energy densities (called formants) or low energy densities. The distribution of formants in the spectrum characterizes the different vowels (see Kent & Read, 1992 for details). Physiological activation influence speech production in the following ways (for a complete description see K. R. Scherer, 1986). An increase in muscle tone tenses the vocal folds and results in a higher pitched voice. Faster and deeper respiration fortifies the air flow and causes louder, higher pitched vocalizations. At the same time, speech gets faster, as the number of syllables between inhalations is held consistent. Salivation, affected by the ANS, changes the resonance characteristics of the oral cavity, leading to changes in the energy distribution of the spectrum. Lastly, even facial expressions influence acoustic structure; smiling for example shortens the vocal tract and thus affects the resonance structure (Tartter, 1980).

K. R. Scherer (1986) made predictions for emotion effects on acoustic parameters as reactions towards appraisal checks, most of which were later confirmed by Banse and Scherer (1996). Highly aroused anger for example is characterized by a high fundamental frequency, fast speech rate, large amplitude, high variability of fundamental frequency, as well as high energy density in the higher frequency regions, while sadness is characterized oppositely (Banse & Scherer, 1996). Single emotion categories can thus be acoustically distinguished (e.g. Goudbeek & Scherer, 2010; Hammerschmidt & Jürgens, 2007; Juslin & Laukka, 2003; Laukka, Juslin, & Bresin, 2005; Murray & Arnott, 1993), while listeners are able to recognize the intended emotion (Banse & Scherer, 1996; Pell & Kotz, 2011; K. R. Scherer, 2003; K. R. Scherer, Clark-Polner, & Mortillaro, 2011; Van Bezooijen, Otto, & Heenan, 1983). The listener's differentiation whether a voice sounds emotional or not happens quickly, indicating the fast attention shift towards and importance of emotional expressions in social partners. Studies on event-related brain potentials for example indicate that emotions in the voice are detected within 200 ms (Paulmann, Bleichner, & Kotz, 2013; Paulmann & Kotz, 2007; Schirmer, Chen, Ching, Tan, & Hong, 2013).

The communication of emotions via vocalization seems to be deeply biologically and evolutionary rooted (see for an extensive overview Scheiner & Fischer, 2011). Comparative studies revealed similar acoustic structures for aversive vocal expression in humans and squirrel monkeys (Fichtel, Hammerschmidt, & Jürgens, 2001; Hammerschmidt & Jürgens, 2007). Additionally, research on normally hearing and hearing impaired children found a comparable usage of seemingly predetermined emotional vocalization (Scheiner, Hammerschmidt, Jürgens, & Zwirner, 2002, 2004), speaking against vocal learning. Lastly, vocal expressions are universally encoded (Pell, Paulmann, Dara, Alasseri, & Kotz, 2009) and recognized (Pell & Skorup, 2008; K. R. Scherer, Banse, & Wallbott, 2001), although cultural variations (Pell et al., 2009) and an advantage for the same cultural background exists (called in-group effect; Elfenbein & Ambady, 2002; K. R. Scherer et al., 2011).

1.1.2.2 Other channels of emotional expressions

Emotional expressions have been studied most widely in the facial domain, and extensive research started with Izard (1971) and Ekman and colleagues (Ekman & Friesen, 1969b; Ekman, Sorenson, & Friesen, 1969). Facial expressions related to emotions are characterized by activation of different muscle movements (Ekman & Friesen, 1975; Izard, 1971). Whether these expressions are seen as representing distinct emotions or more flexible componential appraisal patterns depends on the theoretical position of the researchers (Ekman, 1993; Levenson, 2011; K. R. Scherer & Ellgring, 2007). Facial expressions are universally recognized (Ekman et al., 1969; Elfenbein & Ambady, 2002), although cultural variations and an in-group effect exist (Elfenbein, Beaupre, Levesque, & Hess, 2007; Jack, Garrod, Yu, Caldara, & Schyns, 2012). Body gestures are also used in emotional communication, but have been studied rarely (but see De Gelder, 2009; De Gelder & Van den Stock, 2011; K. R. Scherer et al., 2011). De Gelder (2009) stated that investigating body expressions is especially informative, as they strongly display action tendencies and might be less easily controlled than the face.

1.2 Understanding emotions in others

"You can only understand people if you feel them in yourself."

(Steinbeck 1952; cited by Preston, 2007)

Emotions possess strong social functions; sadness for example is understood to be a call for support, while happiness signals a lack of threat and an invitation to approach (Fischer & Manstead, 2008; Hendriks & Vingerhoets, 2006; Shariff & Tracy, 2011). A quick recognition and understanding of emotional expressions in others is therefore of importance for effective social interactions. Emotional expressions are quickly classified and distinguished (Paulmann et al., 2013; Pell & Kotz, 2011), but the underlying processes that allow to infer emotions to other people from perceiving their expressions are still debated. Understanding others' emotions is part

of the concept of social cognition, which is defined as "the processing of any information which culminates in the accurate perception of the disposition and intentions of other individuals" (Brothers, 1990, p. 28). Beliefs, intentions and desires are attributed to others via a process called mentalizing (or Theory of Mind (ToM); U. Frith & Frith, 2003; Premack & Woodruff, 1978; Singer, 2006), although it is still debated how mentalizing is achieved. People might use a theory-based strategy to make predictions about others' states (Theory-theory; Carruthers, 1996; Premack & Woodruff, 1978), but they might also put themselves imaginatively into the shoes of others and simulate the perceived mental processes (Simulation theory; Goldman & Sripada, 2005; Singer, 2006) (see also Davies & Stone, 1998; Völlm et al., 2006). Understanding emotions has been closely linked to affect sharing and simulation (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Goldman & Sripada, 2005; Singer, 2006), which will be the focus of the next section.

Sharing the affective state of a social partner is called empathy[1] and is thought not only to be an important aspect to understand inner affective states but also to cause prosocial behavior (Eisenberg & Miller, 1987; Hein, Lamm, Brodbeck, & Singer, 2011; Masten, Morelli, & Eisenberger, 2011; Singer, 2006). Empathic reactions comprises reflections of the complete emotional episode, including subjective feeling, and autonomic responses (see the perception-action model by Preston & de Waal, 2002; Hatfield, Rapson, & Le, 2011) and are followed or linked to the more automatic process of contagion[2] (Coricelli, 2005). Both processes are however difficult to tell apart (Preston & de Waal, 2002). The term "empathy" was introduced by Titchener (1909; cited in Gallese, 2003) as the translation of the German word "Einfühlung", which was established by the philosopher and psychologist Theodor Lipps (1903). According to Lipps, inner imitation of actions is the basis to the understanding of others. In the middle of the 1990s, the discovery of the so-called "mirror neurons" in monkeys emphasized the link between sensory and motor processes (Di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Kilner & Lemon, 2013; Rizzolatti, Fadiga,

[1] Empathy is defined according to Preston and de Waal (2002, p. 4) as "Subject's state results from the attended perception of the object's state."
[2] Contagion is defined according to Preston and de Waal (2002, p. 4) as "Subject's state results from the perception of object's state."

Gallese, & Fogassi, 1996). These neurons were interpreted to be involved in action understanding as they were activated during observing as well as during performing an action (Rizzolatti et al., 1996) (for evidence in humans see Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010). This interpretation was later expanded to the understanding of emotions in others (Carr et al., 2003; Gallese, 2003; Mier et al., 2010; Ramachandra, Depalma, & Lisiewski, 2009). Functional magnetic imaging studies found shared activation of the brain network between imitating emotional expressions and observing them (Carr et al., 2003). Additionally, there has been extensive research on pain perception (Jackson, Meltzoff, & Decety, 2005; Singer et al., 2006; Singer et al., 2004), showing that brain structures activated during experiencing pain, such as the anterior cingulate cortex and the anterior insular cortex, are also involved when perceiving pain in others (Jackson et al., 2005; Singer et al., 2004). Feeling pain is not generally described to be an emotion per se (Ortony & Turner, 1990), but has been extensively investigated in the context of empathy (Jackson et al., 2005; Singer, 2006). These findings highlight the shared representation of one's own and others' emotions (Singer & Lamm, 2009).

At the behavioral level, evidence for affective processing, that is sharing emotions, is also increasing. Congruent facial reactions to pictures of emotional expressions, called facial mimicry (first described by Dimberg, 1982), were reliably found in various studies, even towards vocal expressions (Blairy, Herrera, & Hess, 1999; Dimberg, Andréasson, & Thunberg, 2011; Magnée, Stekelenburg, Kemner, & de Gelder, 2007). For example, looking at angry faces or hearing angry voices induce activation of the *corrugator supercilii* muscle, leading to a frown. This effect is intensified when people's trait empathy is high, that is when they declare to be easily emotionally moved by others (Dimberg et al., 2011; Dimberg & Thunberg, 2012; Sonnby-Borgström, 2002). Facial muscle activity has been proposed to depict a direct emotional response (Dimberg, 1997) or to elicit emotions via proprioceptive feedback according to the facial feedback hypothesis (Niedenthal, 2007; Niedenthal & Maringer, 2009). In fact, attending to emotional faces does

actually influence the subjective feelings of the receiver (Hess & Blairy, 2001; Wild, Erb, & Bartels, 2001).

Sharing the emotion of others is thought to be a quick, and thus efficient road for understanding them (Stel & van Knippenberg, 2008). The embodied simulation theory of emotion proposes explicitly that mirroring emotions is essential for their comprehension and thus predicts that emotion recognition is facilitated when emotion evocation in the beholder is effective (see e.g. Goldman & Sripada, 2005). Recent studies could find the predicted link between facial mimicry, subjective experience and emotion recognition; people showing more intense facial mimicry, indicated stronger subjective feelings and revealed more accurate emotion recognition (Künecke, Hildebrandt, Recio, Sommer, & Wilhelm, 2014; Sato, Fujimura, Kochiyama, & Suzuki, 2013).

The embodied simulation theory is however not unchallenged, as easier roads for recognizing emotions in others are suggested, such as a simple feature-based recognition (Zahavi, 2008) and not all studies on facial mimicry revealed an interaction with emotion recognition (Blairy et al., 1999; Bogart & Matsumoto, 2010; Hess & Blairy, 2001). Autonomic reactions to emotional expressions in others – such as increased respiration, perspiration or cardiovascular activity (Kreibig, 2010) - would complete the shared emotional episode (Moors et al., 2013). These somatic responses have been revealed for interactions in real life settings, such as including context and verbal content (Cwir, Carr, Walton, & Spencer, 2011; Levenson & Rueff, 1992), while responses to context-free presented expressions are ambiguous (Alpers, Adolph, & Pauli, 2011; Aue, Cuny, Sander, & Grandjean, 2011; Wangelin, Bradley, Kastner, & Lang, 2012).

The ability to share emotions by contagion or by empathizing seems a useful tool to react adequately in emotional situations, but people do not empathize equally with everyone. Empathic reactions are influenced by appraisal processes (de Vignemont & Singer, 2006; Preston & de Waal, 2002), and even the more automatic contagion reactions do not happen haphazardly (Norscia & Palagi, 2011). Sharing emotions is supposed to be facilitated when interaction partners are

socially connected and of higher relevance, as when they are familiar (Preston & de Waal, 2002), similar to each other (Cwir et al., 2011; Preston & de Waal, 2002) or share group membership (Brown, Bradley, & Lang, 2006; Mathur, Harada, Lipke, & Chiao, 2010). Additionally, empathic responses are increased when the partner is perceived as more likable or fair, respectively (Singer et al., 2006). Indeed, shared group-membership – a situation in which empathic concern is high- was found to improve emotion recognition (Weisbuch & Ambady, 2008; Young & Hugenberg, 2010). Situations in which empathic concern is facilitated might therefore positively impact the classification of emotional expressions due to an increase in affect sharing. In contrast, however, the effect might also result from higher attention and increased motivation to decode the more relevant emotion by in-group members (Ackerman et al., 2006; Thibault, Bourgeois, & Hess, 2006). In this case the improved performance rests on cognitive processes rather than on affective ones. Out-group members are however not only socially less relevant than in-group members, but might also promote the perception of negative emotions (Bijlstra, Holland, & Wigboldus, 2010; Hugenberg & Bodenhausen, 2003; Weisbuch & Ambady, 2008), as it is of higher importance to detect threat in people from different groups. To investigate the effect of social relevance on emotion processing without the interference of these prejudices, the use of a more neutral social connection might be of advantage. Personal similarity (sharing first name, birth date, or interests) for example, was found to directly create a social link between two individuals (Walton, Cohen, Cwir, & Spencer, 2012), increase behavioral mimicry (Guéguen & Martin, 2009) and has been found to increase empathic reactions (Cwir et al., 2011), and it is therefore promising for studying the impact of social relevance on emotion recognition and affect sharing.

Affective neurosciences developed a detailed picture about the reactions of the central nervous system during empathic responses or affect sharing (Dalgleish, 2006; Mathur et al., 2010; Singer, 2006; Wicker et al., 2003). At the behavioral level there is still a lack of clarity regarding the effect of these processes. One of the open questions is whether the social relevance of interaction partners evokes stronger empathic reactions and more accurate emotion

understanding, due to the fact that people can easier “put themselves into the shoes of others”. Beyond that, the knowledge on actual “bodily sensations” (as termed by Singer, 2006, p. 856, meaning the autonomic reactions of an emotional episode) during the processing of emotional expressions in others is generally scarce. Answers to these questions would enlighten the necessity and the extent of affect sharing during emotion communication.

1.3 Play-acted expressions and reliability

In the previous sections I described the processes how emotions are transmitted by the voice through modulations of the vocal tract as well as the possibility to understand others’ emotions by sharing their affective state. The above mentioned studies were, however, mostly conducted using play-acted expressions that do not mirror the affective state of the individual expressing it. It is an important next step, to disclose the relation between felt and unfelt emotional expressions.

1.3.1 Reliability

Following the concept of emotions (Chapter 1.1.1), an expression is accompanied by physiological reactions and reflects the inner affective state of the person showing the expression. However, the occurrence of emotional expressions is strongly social (Fernández-Dols & Ruiz-Belda, 1995; Fridlund, 1991) and the coherence between expressions and other components of an emotional episode could until now not sufficiently be demonstrated (see Fernández-Dols & Crivelli, 2013; Mauss, Levenson, McCarter, Wilhelm, & Gross, 2005; Reisenzein, Bördgen, Holtbernd, & Matz, 2006; Reisenzein, Studtmann, & Horstman, 2013). Scherer proposed a model accounting for this twofold use of emotional expressions (K. R. Scherer, 2003; K. R. Scherer & Bänziger, 2010; U. Scherer, Helfrich, & Scherer, 1980). He described that expressions are

influenced by internal physiological reactions ("push factors"), and by social requirements or display rules ("pull factors"). Emotional expressions are produced via an interplay of peripheral physiological responses that "push" an expression and deliberate modulations that are used to fulfil social expectations (pull factors). According to this model, whether an emotion is truly felt or deliberately produced is a false dichotomy as it is always both (K. R. Scherer & Bänziger, 2010).

From a biological-evolutionary point of view the differentiation between emotion-based expressions and unfelt, rather socially used ones is however of relevance. Three recent theoretical papers extensively discuss the topic of emotional expressions as signals in a stable communicative system (Dezecache, Mercier, & Scott-Phillips, 2013; Mehu & Scherer, 2012; Schmidt & Cohn, 2001). They highlight that in order to provide any information[3] to the beholder, emotional expressions should, at least on average, be reliable (see also R. A. Johnstone & Grafen, 1993; Maynard Smith, 1991; Maynard Smith & Harper, 1995). I will use an example to clarify the matter. Happiness indicates an invitation to approach and the absence of threat (Fischer & Manstead, 2008). If the expressions of happiness can be used deceptively, everyone would produce it regardless of the actual intention. The expressions would soon lose its value. Coming back to Scherer's model, it is the question whether expressions purely produced by push factors are different from expressions that are only based on pull factors, and whether listeners or beholders can distinguish both from each other.

Research on deceptive behavior in facial communication has been conducted by Ekman and colleagues (e.g. Ekman, Davidson, & Friesen, 1990; Ekman & Friesen, 1969a; Ekman & O'Sullivan, 2006). They proposed that facial expressions can be masked or faked but that the felt emotion will shimmer through by subtle, rapid muscle movements that are difficult to be influenced voluntarily, a process called "leakage" (Ekman & Friesen, 1969a; Ekman & O'Sullivan, 2006, see also Porter & ten Brinke, 2009; Porter, ten Brinke, & Wallace, 2012). Specific muscles are assumed to be activated only under effort, these are called "reliable muscles" and include the

[3] Defined as the "reduction of uncertainty" (Wheeler et al., 2011, p. 188)

prominent *orbicularis oculi* (cheek raiser), which in combination with the *zygomaticus major* (lip corner pull) accounts for the Duchenne smile (Ekman et al., 1990; Ekman & Friesen, 1982). This smile, created by activation of the *zygomaticus* and the *o. oculi*, has been proposed to be a read-out of felt positive emotions, in contrast to a smile generated by *zygomaticus* activity only (non-Duchenne smile) (Ekman et al., 1990; Ekman & O'Sullivan, 2006). Mehu, Mortillaro, Bänziger, and Scherer (2012) supported the assumption of reliable muscles experimentally. In their study, facial expressions including muscle activity, which was rated as being difficult to activate deliberately beforehand, were actually perceived as more authentic than expressions that were produced without activation of these muscles (see also Warren, Schertler, & Bull, 2008). Recent research, however, failed to confirm the dichotomic differentiation between Duchenne and non-Duchenne smiles, indicating rather a relation with smile intensity and thus challenging the assumption of reliable muscles (Gunnary & Hall, 2014; Krumhuber & Manstead, 2009; Riediger, Studtmann, Westphal, Rauers, & Weber, 2014). Timing pattern seems to distinguish between felt and deliberately produced facial expression (Ekman & O'Sullivan, 2006; Hess & Kleck, 1990), but further empirical support is needed here. From the receiver's side, recognition of emotional deception was only poorly achieved, mostly at or barely above chance level (Ekman & O'Sullivan, 1991; Porter & ten Brinke, 2009; Porter et al., 2012; Warren et al., 2008).

This short overview indicates the dominance of facial expressions in the line of research on reliability of emotional expressions. Vocal expressions rests more strongly upon responses of the autonomic nervous system (see chapter 1.1.2; K. R. Scherer, 1986) and might thus be more prone to reveal underlying emotional episodes. No markers have been suggested to expose deceptive expressions - with the exception of fluctuations in the fundamental frequency (jitter, see Juslin & Laukka, 2001) -, but deliberately expressed emotions were assumed to be more intense and stereotypical, as the speaker might not be able to control the subtle adaptations of the vocal tract (Laukka, Audibert, & Aubergé, 2012; Wilting, Krahmer, & Swerts, 2006).

Considering that the knowledge on differences between felt and socially used expression is scarce, it is of interest that research on emotional expressions was mostly conducted using actors' portrayals. While this procedure avoids the difficulties of recording spontaneous expressions, such as ethical constraints or technical problems of high quality recording conditions (Bänziger & Scherer, 2007; K. R. Scherer, 2003), it certainly restricts ecological validity, especially as stimuli were preselected in order to ensure the correct emotional content, leading to highly stereotypical and intense expressions (i.e., Banse & Scherer, 1996). With regard to the scarce knowledge on actual expression pattern, Barrett (2011) stated that researchers study merely symbols of emotional expressions instead of emotional expressions as they occur in daily life. According to her, researchers might consider the wrong expression patterns as long as it is not clarified how emotions are actually expressed - for example when looking at coherence between expressions and subjective feelings (Barrett, 2011; Carroll & Russell, 1997). The knowledge on spontaneous vocal expressions so far comes from depressed or anxious patients (Laukka, Linnman, et al., 2008) or from recordings of talk shows (Grimm, Kroschel, & Narayanan, 2008), telephone services (Laukka, Elenius, Fredrikson, Furmark, & Neiberg, 2008) or during emergencies (Williams & Stevens, 1972). Emotion induction methods were also used to gain stimulus material (T. Johnstone, van Reekum, Hird, Kirsner, & Scherer, 2005). The results so far are corresponded with the findings on acted expressions, but with smaller effect sizes (Laukka, Elenius, et al., 2008). Yet most of these studies had a restricted sample of different emotion categories, or dealt with low intense expressions.

A set of studies concerning the comparison of spontaneous and play-acted expressions was conducted in our research group. We collected radio sequences, in which people were interviewed in emotional situations and compared these with re-enactments of the same situation by professional actors. An acoustic analysis (Jürgens, Hammerschmidt, & Fischer, 2011) revealed that the production in both conditions differed, as articulation differences and a more variable speech melody distinguished acted from spontaneous expressions. The acoustic structure

differed only slightly across the emotion categories (in contrast to Banse & Scherer, 1996; Hammerschmidt & Jürgens, 2007) and no interaction between emotion and recording condition (whether the recording was acted or spontaneous) was found. Drolet, Schubotz, and Fischer (2012) showed in their imaging study using the same stimulus set that BOLD responses (blood oxygenation level dependent responses) during listening to spontaneous expressions differed compared to listening to play-acted ones. Participants were poor in distinguishing the play-acted expressions, however. Most interestingly, play-acted expressions were not more easily recognized than spontaneous ones, which reflects the acoustic analysis and speaks against the assumption that acted expressions are more stereotypical than spontaneous ones. Play-acting had nevertheless an influence on emotion recognition, namely a more accurate recognition of angry and a less accurate recognition of sad speech compared to spontaneous expressions. This result indicated that play-acting might have a more complex, emotion-specific influence on recognition, but that people are poor in recognizing whether the expressions was acted or not. The play-acted stimuli of this study were produced by professional actors, who represent a special case of encoders. Actors are trained to act emotional expression and might therefore be especially suited to produce credible expressions (K. R. Scherer & Bänziger, 2010). A comparison of acting-inexperienced people would disclose the issue whether people are generally capable to play-act expressions convincingly.

It is important to clarify that when referring to “spontaneous” expressions in the context of this stimulus set, I do not state that these rely solely on push factors. This is often misconceived. The spontaneous recordings are done in social situations, and expressions are certainly influenced by social requirements (pull factors), although they are not staged (see also K. R. Scherer & Bänziger, 2010). The play-acted expressions on the other hand can be regarded as incoherent with the underlying emotional episode. Although specific acting techniques use the recollection of emotional episode to create actual emotions and some actors do feel into their role, this procedure requires preparation and training (Goldstein & Winner, 2010; Stanislavskij,

1989) and the actors did not prepare themselves in that way. There might be the possibility that the spontaneous expressions are play-acted, however unlikely this explanation seems to be - regarding the situations in which they were recorded (people speaking about the death of their children or winning in a lottery). In this case the whole concept of emotion communication in daily life would be challenged.

1.3.2 Acted expressions in cross-cultural emotion research

The use of preselected, highly intense, play-acted stimuli might bias research findings on emotional expressions as stated by Barrett (2011). The problem of unrealistic, acted emotion portrayals has been discussed for example in the context of universality of emotional expressions. As mentioned in Chapter 1.1.2 emotional expressions are recognized accurately across different cultures and language families, but people from the same cultural background were found to possess an advantage in recognizing the expressions (in-group effect, Elfenbein & Ambady, 2002; K. R. Scherer et al., 2001). On the one hand, this in-group-effect might be caused by the higher motivation, the increased attention or facilitated empathic concern towards people of the same group (see Chapter 1.2, Thibault et al., 2006), or, in the case of vocal expressions, by familiarity with the language. On the other hand Matsumoto, Olide, and Willingham (2009) proposed that the in-group-effect across cultures might be created artificially by using play-acted expressions. They argued that acting rely more strongly on social codes than spontaneous expression (see also Hunt, 1941) and that while spontaneous emotional expressions are universally equivalent, acted expressions differ across cultures. In their study, Matsumoto et al. (2009) demonstrated a lack of intragroup advantage when looking at spontaneous facial expressions of joy. Another unintentional effect of play-acted expressions was introduced by Elfenbein, Mandal, Ambady, Harizuka, and Kumar (2002), who stated that using preselected, highly aroused and intense expressions might cover possible culture specific decoding rules (see also Wagner, 1993). Although emotional expressions are universal, their appraisal and the evaluation in which

situation the expression is perceived as appropriate does vary across cultures (Matsumoto & Hwang, 2011). In collectivistic societies attributing negative emotions to people has been assumed to endanger group stability, which might lead to a bias against using negative emotion categories during recognition (Matsumoto, 1989, 1992). This effect might be covered by the unambiguousness of preselected play-acted emotion stimuli (Elfenbein et al., 2002; Wagner, 1993) and lead to an overestimation of cross-cultural similarity. These examples, which again referred to facial expressions, highlight the importance to look at daily life expressions and to reveal the effect of acting in order to fully understand the impact of human emotional communication.

1.4 Aims

In the previous sections, I disclosed open questions concerning vocal emotion expressions. I summarized that the relation between spontaneous and deliberately produced expressions and thus their reliability is still not disclosed. Additionally, I pointed to the fact that the processes which lead to an understanding of others' expressions and especially the importance of affect sharing are also far from being understood. Generally, the literature review emphasized the lack of knowledge concerning vocal expression.

In my thesis I investigated the relation between spontaneous and play-acted speech tokens to reveal the human ability to produce and to detect deliberately expressed vocal emotions portrayals in Chapter 2 and Chapter 3. These two chapters belong to a set of studies that aimed to reveal how play-acted and spontaneous expressions are differently processed and perceived (see Drolet et al., 2012; Jürgens et al., 2011). **Chapter 2** concentrates on the emotion recognition and the perceived authenticity of spontaneous expressions and their re-enactments in a cross-cultural comparison including German, Romanian and Indonesian listeners. This study enlightens the universality of play-acted and spontaneous vocal expressions and gives evidence whether the

emotion recognition patterns, found by Drolet et al. (2012), have a universal basis or are a result of listeners' culture. In case that acting rests on social codes, I predicted that emotion recognition would be less accurate for acted expressions in cultures other than German. If the relation of spontaneous and play-acted expressions in contrast is based on a universal basis, similar recognition rates across the three cultures would be found. In accordance to Elfenbein et al. (2002), I predicted additionally that the cultural-dependent decoding biases against negative emotions in collectivistic cultures (Romania and Indonesia) have stronger effects on the more ambiguous expressions, which would include the authentic anger and the fear stimuli.

In **Chapter 3**, I clarify whether acting training affects the production of emotional expressions and their recognition, by including emotional portrayals by acting in-experienced people. This aims to disclose whether acting emotions has to be trained to be convincing. Furthermore, including non-trained people clarifies the source of the emotion recognition pattern mentioned above, namely whether it is caused by acting in general, or by especially trained actors' speech. This study consists of an acoustic analysis and a rating experiment to focus both on the production and on the recognition of vocal expressions. Under the hypothesis that professional actors are better suited to produce credible vocal expressions, as they are trained for this task, I predicted that expressions by non-trained people deviated more strongly in their emotion recognition from the spontaneous expressions than the ones by actors; namely possessing even higher recognition rates for anger and lower for sadness. In their production, actors' portrayals would resemble the spontaneous expressions more strongly than the non-trained people. It might however be the case, that acting- and speech training interferes with the production of emotional expressions, in this case I predicted that the expressions by professional actors are most extreme both in the acoustic analysis and in the rating data.

Chapter 4 expands the topic of recognizing and processing vocal expressions. In chapter 2, I implicitly dealt with the effect of social connectedness on emotion recognition by regarding cultural group membership. In Chapter 4, I focused more explicitly on the question whether

increased social relevance, manipulated by biographical similarity, improves vocal emotion recognition. This study investigated whether attending to vocal expression alone elicits emotional engagement including autonomic reaction (skin conductance and pupil size) and whether increased relevance of the speaker interacts with emotion recognition and emotion sharing. I hypothesized that sharing biographical characteristics increased emotion recognition either by shifting attention, or by stronger empathic concern. In the first case I would predict an increase in pupil dilation, which is a marker for attention (Laeng, Sirois, & Gredeback, 2012), for expressions spoken by a similar character, and in the latter a general increase in emotional engagement measured by autonomic skin conductance response.

In **Chapter 5**, I discuss the results in a broader context and give future perspectives.

2 Encoding conditions affect recognition of vocally expressed emotions across cultures

Rebecca Jürgens[1], Matthis Drolet[1], Ralph Pirow, Elisabeth Scheiner, Julia Fischer

Cognitive Ethology Laboratory, German Primate Center

[1] these authors contributed equally to the work.

Frontiers in Psychology (2013), 4:111

doi 10.3389/fpsyg.2013.00111

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 in terms of 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.

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 for example Mason & Capitanio, 2012; Mulligan & Scherer, 2012; Scarantino, 2012), there is a growing consensus among theorists that emotion needs to be viewed as a multi-component phenomenon (Frijda, 1986; Lazarus, 1991; K. R. 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; K. R. 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 & Capitanio, 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 et al., 1969; Ekman and Friesen, 1971; Ekman and Oster, 1979) 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 & Hwang, 2011), in small dialectic-like differences (Elfenbein et al., 2007) and in the culture-specific display rules that alter facial expressions (Ekman & Friesen, 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 & Scherer, 1996; Hammerschmidt & Jürgens, 2007; Juslin & Laukka, 2003; Sauter et al., 2010). Expression of emotion in the voice occurs via modifications of voice quality (Gobl & Ni Chasaide, 2003) and prosody in general (K. R. 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 & Ambady, 2002). Ratings of vocalizations by listeners showed that they were able to infer vocally expressed emotions at rates higher than chance (Banse & Scherer, 1996; Juslin & 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 & 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 & 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, Masumoto (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 & Bloom, 2011), acted emotional expressions may still be more stereotyped and more intense than spontaneous expressions (Wilting, Krahmer, & Swerts., 2006; Laukka et al., 2012, but see Jürgens et al., 2011; K. R. 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 re-enacted 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 had a significant effect on 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).

Based on 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 in terms of the distance to 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, Lin, & Clark, 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).

Material and Methods

Recordings

We focused on four emotions that differ in terms of valence, dominance and intensity: anger, fear, joy and sadness (Bryant & Barrett, 2008; de Vignemont & Singer, 2006; Ethofer, de Ville, Scherer, & Vuilleumer, 2009). These are the most commonly used emotions in this field of research (Juslin & Laukka, 2003; K. R. Scherer et al., 2001; Sobin & Alpert, 1999) 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 represents a way of speaking more related to acting than to natural speech. We compared emotional expressions that were obtained during radio interviews to re-enacted 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. Prior to 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 s (*SD*: 1.2 s). 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 (incl. 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 re-enactements 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).

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.

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 five authentic and five 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 three 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) via a program called Emosurvey (developed by Martin Schmeisser). Participants heard the stimuli via earphones (Sennheiser HD 497). They could activate the playback of the stimuli themselves and each stimulus could be activated a maximum of three times. The ratings were made via mouse clicks on the screen. When all questions were answered, the next stimulus could be activated. 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).

Statistical Analysis

All models were implemented in the R statistical computing environment (R Developmental 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 ~ 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 ~ 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, Bretz, & Westfall, 2008), adjusting the p-values for multiple testing via 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 have a strong preference for 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, Wilson, & 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 (α), 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, Thomas, Best, & Speigelhalter, 2000) using the R2WinBUGS interface package (Sturtz, Ligges, & Gelman, 2005) to account for the four response options ("anger," "fear," "sadness," and "joy").

Results

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 ($z = 18.39$, $p < .001$; Figure 2.1). 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 $z = 2.99$, $p = .028$; German–Indonesian $z = 2.95$, $p = .031$).

The analysis of ratings using choice theory revealed that participants had a strong bias towards choosing the response 'authentic' in the authenticity ratings (Figure 2.2), 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 ($z = 2.64$, $p = .045$; Figure 2.2).

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 < .001).

Figure 2.1 *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.*

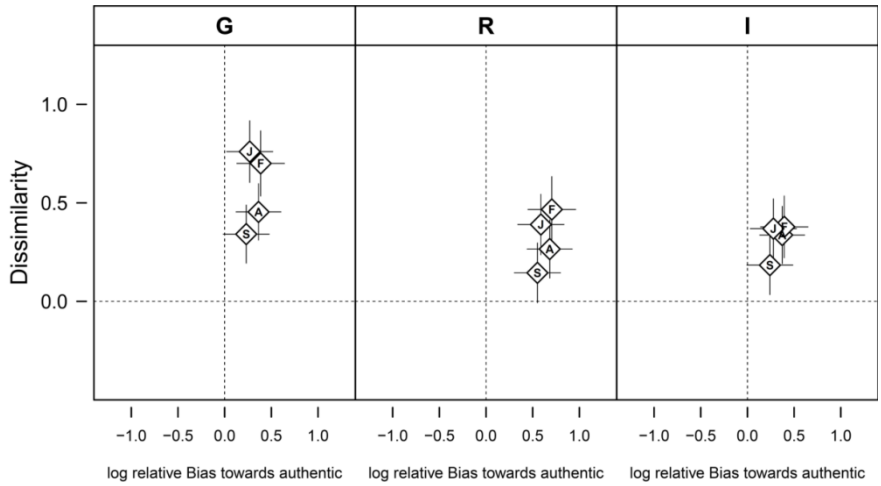

Figure 2.2 *Discrimination of authentic and play-acted vocal expressions of emotions as assessed by choice theory. The discriminatory ability is described by the dissimilarity between authentic and play-acted stimuli (depicting how well the stimuli could be discriminated) and by the participants' relative bias towards choosing authentic as a response, 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 ability to distinguish the stimuli. Data are given are as means ± 95% confidence intervals.*

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 in general showed similar patterns in the three countries (Figure 2.3). The GLMM analysis revealed that the rate of correct emotion recognition was influenced by Intended emotion, Stimulus authenticity and Country (see Table 2.1 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 2.1).

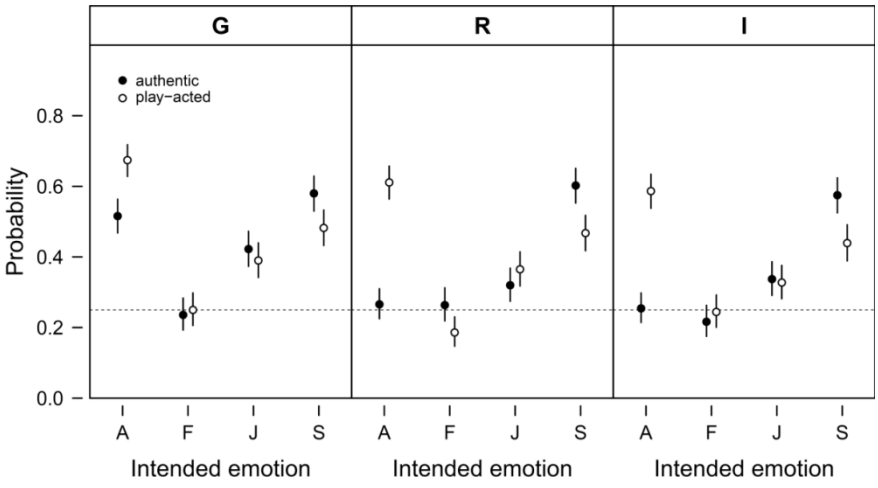

Figure 2.3 *Probability of correct emotion recognition. Given is the probability of correct emotion recognition with respect to the intended emotion (A - anger, F – fear, J – joy, S – sadness) and stimulus authenticity (authentic or play-acted). The data are split by cultural affiliation (G - Germany, 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 2.1
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	<.001	***
Germany - Indonesia == 0	0.351577	0.059665	-5.893	<.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 (Figure 2.4). 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 with respect to the responses ‘joy’ and ‘fear’. Indonesian participants, whose bias against ‘joy’ was less distinct than for Romanian or German participants, were the only exception.

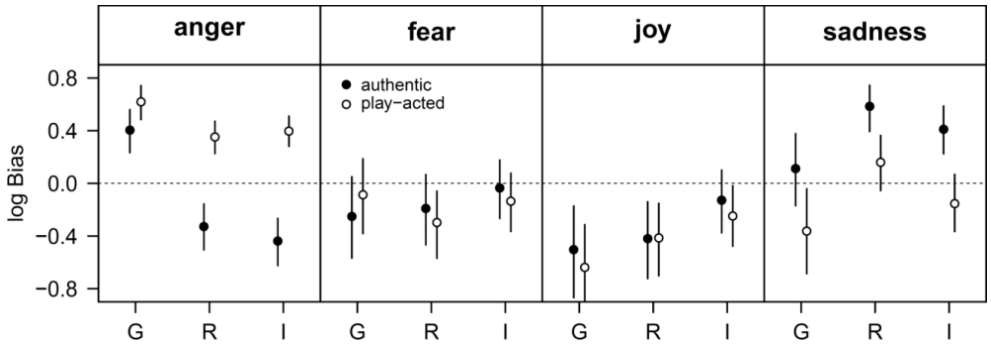

Figure 2.4 *Analysis of emotion recognition data by choice theory. Given is the log-transformed response bias for each of the four possible choices (anger, fear, joy, sadness) with respect to cultural affiliation (G - Germany, R - Romania, I - Indonesia). The filled and open symbols indicate the response 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.*

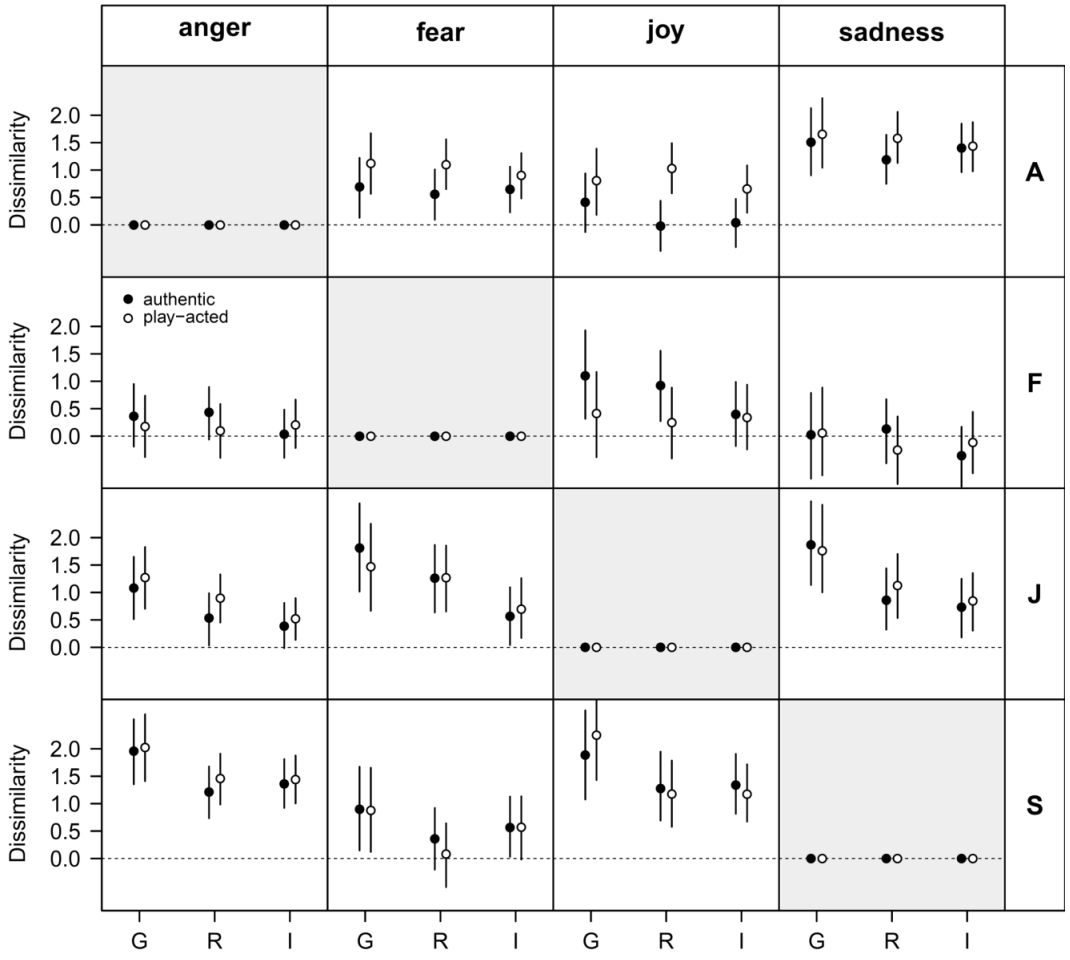

Figure 2.5 *Analysis of emotion recognition data using choice theory. Given is the dissimilarity for different pairs of emotion stimuli with respect to cultural affiliation (G - Germany, R - Romania, I - Indonesia). The rows and columns of this matrix plot indicate the four emotion stimuli (A - Anger, 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).*

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 2.5. 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 Figure 2.5) indicate high confusion with the other emotion categories and reflect the low recognition rates for fear.

Discussion

Participants in all three cultures had difficulties 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 vs. play-acted) had a clear impact 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.

With regard to 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

, Krahmer, & Swets, 2007; Laukka, Audibert, & Aubergé, 2012). 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 & Scherer, 2012; Schmidt & Cohn, 2001). The inability to distinguish between play-acted and spontaneous expressions is, therefore, counter-intuitive, but has also been found in previous studies (see for corresponding results Ekman & O'Sullivan, 1991; Audibert, Aubergé, & Rilliard, 2008). People tend to believe in the truthfulness of a statement rather than mistrust it (Zuckerman et al., 1984; Levine et al., 1999). 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).

In addition to the well documented in-group effect for German participants (Elfenbein & Ambady, 2002; K. R. 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, McKeown, McRorie, & Vukicevic (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 over-simplification, 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 (e.g. Pell & Kotz, 2011; K. R. Scherer et al., 2001; Van Bezooijen et al., 1983;). 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 (Pell et al., 2009; K. R. Scherer et al., 2001; Van Bezooijen et al., 1983). 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 laymen and inexperienced actors before (Pell et al., 2009; Van Bezooijen et al., 1983;). In particular, the low recognition rates for joy and fear at or close to chance levels might reveal interesting facts about emotional expressions in general. 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. It is clear that 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.

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.

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.

3 Effect of acting experience on emotion expression and recognition in voice: Non-actors provide better stimuli than expected

Rebecca Jürgens [a,b], Annika Grass [a], Matthis Drolet [a], and Julia Fischer [a,b]

[a]Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany

[b]Courant Research Centre "Evolution of Social Behaviour", University of Göttingen, Göttingen, Germany

Journal of Nonverbal Behavior (2015)

doi 10.1007/s10919-015-0209-5

Abstract

Both in the performative arts and in emotion research, professional actors are assumed to be capable of delivering emotions comparable to spontaneous emotional expressions. This study examines the effects of acting training on vocal emotion depiction and recognition. We predicted that professional actors express emotions in a more realistic fashion than non-professional actors. However, professional acting training may lead to a particular speech pattern; this might account for vocal expressions by actors that are less comparable to authentic samples than the ones by non-professional actors. We compared 80 emotional speech tokens from radio interviews with 80 re-enactments by professional and inexperienced actors, respectively. We analyzed recognition accuracies for emotion and authenticity ratings and compared the acoustic structure of the speech tokens. Both play-acted conditions yielded similar recognition accuracies and possessed more variable pitch contours than the spontaneous recordings. However, professional actors exhibited signs of different articulation patterns compared to non-trained speakers. Our results indicate that for emotion research, emotional expressions by professional actors are not better suited than those from non-actors.

Introduction

Acting is not only an essential part of human performative culture, but also of everyday social life, since emotion expressions in natural settings are frequently play-acted due to social requirements (Goffman, 1959; Gross, 1998; Hochschild, 1979; Kappas, 2013). At the same time, actors' portrayals can be strongly influenced by subjective feelings, especially when produced via techniques based on emotional imagination or memory (Gosselin, Kirouac, & Dore, 1995; K. R. Scherer & Bänziger, 2010). Therefore, it has been argued that genuine expressions of emotions and play-acted ones are difficult, if not impossible, to distinguish (K. R. Scherer & Bänziger, 2010). Other authors have criticized clearly staged expressions as being stereotypical, exaggerated and more intense than spontaneously occurring expressions (Barrett, 2011; Batliner, Fischer, Huber, Spilker, & Nöth, 2000; Douglas-Cowie, Campbell, Cowie, & Roach, 2003). However, only a handful of studies have directly compared authentic expressions and actors' portrayals (Aubergé, Audibert, & Rilliard, 2004; Drolet et al., 2012; Greasley, Sherrard, & Waterman, 2000; Laukka et al., 2012; Williams & Stevens, 1972). In most of these studies, play-acted expressions were found to be more intense or more stereotypical (Laukka et al., 2012; Wilting et al., 2006). Yet, some more recent studies failed to detect such a pattern (Drolet et al., 2012; Jürgens et al., 2013; K. R. Scherer, 2013).

In a set of earlier studies, we compared vocal expressions of emotions taken from natural, non-staged situations recorded by a local radio station with their re-enactments by professional actors. In the course of the paper we will use the terms "authentic," "play-acted," and "realistic" according to the following definitions. "Authentic" is used for stimuli that are recorded in spontaneous non-staged, daily life situations, reflecting the expressions we use in our day-to-day emotion communication. The term does not reflect the physiological (affective) state or inner feelings of the encoder. "Play-acted" stimuli are recorded under the instruction, to transmit specific emotional information using a given wording, without an intrinsic motivation of the encoder. "Realistic" is used for play-acted stimuli that are perceived as authentic, that is believed

to be spontaneous. Our results showed that listeners were poor at identifying the encoding condition (that is whether the stimuli were authentic or play-acted). Furthermore, in contrast to the prediction that they are more stereotypical, play-acted expressions were not generally recognized more accurately (Drolet et al., 2012; Jürgens et al., 2013). Instead, we found a significant interaction between emotion category and encoding condition with anger being recognized more frequently when play-acted, while sadness was recognized more frequently when authentic. This effect has been replicated across different cultures (Jürgens et al., 2013). An imaging study comparing brain activation via BOLD response (blood oxygenation level dependent, measured by functional magnetic resonance imaging) of the authentic and play-acted stimuli showed that listening to the authentic but not to the play-acted stimuli activates the Theory of Mind network (ToM) (Drolet et al., 2012). The encoding condition of emotional stimuli thus interacts with neural processing, indicating its importance on human response behavior.

A comparison of the acoustic structure revealed differences in articulation and a more variable pitch contour for play-acted stimuli, showing that measurable differences in the stimulus material between play-acting and authentic encoding condition exist (Jürgens et al., 2011). As we compared not only acting to non-acting but professional actors' voices to normal people's voices and speaking style, our results raised the question whether the effects referred to acting in general, or to the elaborated articulation of professional actors.

Professional actors may produce emotional expressions that are more realistic than expressions by lay people (hereafter "non-actors"), due to their acting training (hypothesized by Krahmer & Swerts, 2008; K. R. Scherer & Bänziger, 2010). Specific acting styles include own feelings as part of the actors' performance; these methods require extensive training and are supposed to increase realism, precisely because they rely on inner affective states, thereby emphasizing the advantage of using actors for creating emotional stimuli that resemble expressions in spontaneous situations (Enos & Hirschberg, 2006; Gosselin et al., 1995; K. R. Scherer & Bänziger, 2010).

Actors, however, need to transmit their emotional expression to the back row of the theater, which might lead to overexpression (Kracauer, 2005) and their speech training may influence phonation and articulation in order to produce loud, intelligible, and persisting speech (Master, Debiase, Chiari, & Laukkanen, 2008; Nawka, Anders, Cebulla, & Zurakowski, 1997; Roy, Ryker, & Bless, 2000). Thus, professional actors may not necessarily produce more realistic emotional expressions compared to non-actors (see Krahmer & Swerts, 2008). Spackman, Brown, & Otto (2009) tested the influence of acting training on emotional expressions in voice comparing eight drama students with an inexperienced control group. An acoustic comparison revealed interaction effects between encoding conditions and the acoustic structure for single emotions. From the perspective of the listener, anger and fear stimuli produced by actors were recognized more accurately than those by laymen, but the reverse was true for happiness and sadness. Krahmer and Swerts (2008) induced positive and negative affective states in their participants via the Velten induction method (Velten, 1968) and compared the resulting facial expressions with portrayals by experienced theatre actors and non-actors. Contrary to their prediction, facial expressions by actors were perceived as the most intense. These studies indicate that professional actors may not be more suited to producing emotional expression than non-actors, at least when resemblance to spontaneous expressions is the goal.

Our study aims to deepen the understanding of training effects on vocal expressions of emotions and to put them in relation to expressions produced in spontaneous situations. With this approach, we aim to advance the discussion about what differences concerning the effect of authenticity (Drolet et al., 2012; Jürgens et al., 2011; Jürgens et al., 2013) are due to acting per se and which might be due to the actors' way of speaking.

We formulated two opposing hypotheses: (1) If professional actors are more suited to producing realistic emotional expression through their acting training, we would predict that the acoustic structure of non-actors' speech tokens deviate more from the authentic expressions than the actors' portrayals. In this case, we would predict that portrayals by non-actors were more

stereotypical and exaggerated and, thus, were more easily recognized as being play-acted. Recognition accuracies for the emotion categories would be the highest for the non-actors' expressions. (2) If however, professional acting and speech training leads to different speech patterns, we would predict that expressions by professional actors differ from the other conditions, both in their acoustic structure and in their perception, while the differences between non-actors and authentic emotion expressions were negligible. Recognition accuracies both for authenticity recognition and emotion recognition would in this case be the highest for actors. Based on earlier research (Jürgens et al., 2011), we made clear predictions for the acoustic parameters. In the past decades acoustic parameters have been described that differentiate the expressions of different emotion categories (Hammerschmidt & Jürgens, 2007; Juslin & Laukka, 2001; K. R. Scherer, 1986). These parameters mirror the phonation (sound production) and the articulation process (modulation of sound via nasal and oral cavities) respectively. Highly aroused emotions such as anger, are spoken faster, less monotonously, more loudly, with more energy in the higher frequencies, more noise in the signal and in a higher fundamental frequency (pitch); while low aroused emotions such as sadness are spoken slower, monotonously, quietly, with more energy in the lower frequencies, with less noise, and in a lower fundamental frequency. Speed of speech, speech melody, fundamental frequency, harmonic-to-noise-ratio, and peak frequency are thus parameters that distinguish emotional speech and that are related to arousal differences in general (Juslin & Laukka, 2001; K. R. Scherer, 1986). In the previous study on the extensive acoustic analysis of the authentic and professionally acted expressions (Jürgens et al., 2011), none of these parameters differed systematically between authentic and actors' speech tokens, with the exception of peak frequency, which was found to be slightly lower in play-acted expressions and the more variable speech melody in acted portrayals. The most pronounced differences between actors and authentic speech tokens were the broader bandwidths of the first formants, the more dominant fundamental frequencies, both of which are not affected by emotion, and the more variable pitch contour (speech melody) in actors' speech (Jürgens et al.,

2011). If non-actors' portrayals were more exaggerated (and thus more aroused) than actors' expressions, we would predict higher values for the arousal related parameters (fundamental frequency, speed of speech, peak frequency, harmonic-to-noise ratio, energy distribution, and pitch contour) in the non-actors condition. Additionally, the bandwidths of the first formant and the more dominant fundamental frequency should be even more pronounced. However, if the articulation and modulation differences are something related to the actors' voice, we predict negligible differences between non-actors and authentic speech in these acoustic structures.

Method

Stimuli

Authentic. The authentic speech recordings were selected from the database of a German radio station and were taken from interviews made while the individuals were talking about an emotionally charged on-going situation or describing their emotional state while recollecting a past event. 80 recordings were selected that had a good recording quality and a low amount of background noise. The selected recordings contained interviews in which the individuals expressed anger, fear, sadness, or happiness (specified via situation context and verbal content of the recordings). The radio recordings were then converted into wav files (sampling rate of 44.1 kHz). From these interviews, short segments up to 5.5 s in length were cut and consisted of neutral verbal content that does not indicate any specific emotion. Neutral content was rated prior to the study by an additional set of 64 naïve participants. Only these short segments, which ranged from three words to half-sentences, were used for the study, such as " -up to the crossbar" [German original: "bis zum Fensterkreuz"], "twice in a row and such" ["zweimal hintereinander und so"], and "read it again" ["lesen Sie es noch mal vor"]. The 80 speech segments were spoken by 78 speakers and consisted of 22 anger, 18 fear, 20 joy, and 20 sadness stimuli (half spoken by female speakers). The emotional content of the recordings (whether we classified the interview to the anger, joy, fear, or sadness condition) was determined via context

analysis by a post-doctoral member of our research group. Recordings in which speakers were talking about a loss were categorized as sadness, while situations regarding winning and celebration were categorized as happiness. Recordings in which people reported or lived through a threatening event were grouped as fear and the ones in which people verbally attacked someone were grouped as anger. The selected recordings represented a broad variety of emotion situations and emotion intensities. We could neither exclude the possibility of mixed emotions by the sender nor could we control their actual physiological affective state. However, our focus was on the natural communication of emotion that is seldom clearly distinct and controlled. The recording instructions of the actors and non-actors were adjusted to allow for comparable mixed expressions. Examples of the stimuli and of the context situations are found in the Appendix.

Play-acted expressions. Professionally play-acted stimuli (hereafter stimuli produced by "actors") were produced by 21 male and 21 female actors (*M* age = 31 years, SD = 7.9, age range = 21 - 48 years), 30 of them were professional actors who mainly worked on stage, 11 were acting students at the end of their education and one was a professional singer with acting experience. All of the actors had taken part in professional acting training. They were asked to enact one to three of the authentic recordings. Most of them (33 of 41) reproduced two original recordings with the same intended emotion that is two times anger, or two times joy, respectively. The actors used an information sheet (indicating the gender of the speaker, a situational description, and a transcription of the spoken text, including the respective text segment later used for the study) and were told to express the text in their own way. The respective emotion was mentioned in the situational description (e.g., "…She said full of joy…, …an inhabitant reports her fears…, she reports her pain and her sadness…, he got terribly agitated…"). This allows mixed emotions and different intensities expressed by the actors to mirror the recording condition of the authentic expressions. The actors were instructed not to speak in their stage voice, to imagine the situation and to feel into it. The short segment that was later used for the study was not known to the actors. Actors were allowed to express the text as often as they wanted and could select the

expression they considered to be most successful. The recordings were made with a Marantz Professional Portable Solid State Recorder (Marantz, Kenagawa, Japan) with a sample rate of 44.1 kHz, a sampling depth of 16 bit and a Sennheiser directional microphone (Sennheiser, Wedemark, Germany, K6 power module and ME64 recording head).

Non-professionally play-acted recordings ("non-actors") were recorded similarly. Twenty women and 19 men (*M* age = 45 years, SD = 14.8, age range = 21 – 67 years) were recruited via postings at a university notice board and by recruitment in the second authors' circle of acquaintances. The sample of non-actors consists of students, teachers, and normal employees. The non-professional actors were thus older on average than the professional actors; however, age classes could not be determined for the authentic speakers. Ten of the speakers indicated experiences in amateur theatre groups (like school theatres), but none of them received professional acting training. Recordings were made using the same procedure and the same transcriptions as for the actors and were made with a Field Memory Recorder (Fostex, Tokyo, Japan, FR-2LE) and a Sennheiser directional microphone (Sennheiser, Wedemark, Germany, K6 power module and ME64 recording head) with a sampling frequency of 44.1 kHz and a 16 bit sampling depth. To reduce category effects between authentic and play-acted stimuli, both the professional and the non-professional re-enactments were partly recorded outside with varying background noise, as the radio recordings also varied in their background noise.

For all three conditions, the recordings were edited with AvisoftSASLab Pro Version 5.1 (AvisoftBioacustics, Berlin, Germany) to cut the short segments used for the study out of the longer interviews. The final stimulus set consisted of 240 short text segments (80 authentic, 80 professional play-acted, and 80 non-professional play-acted) with non-emotional text content (e.g., "up to the crossbar") flanked by 0.5 s silences. The mean duration of all stimuli was 1.87 sec (*SD* = 1.29, range: 0.327- 8.03 sec). Duration did not vary between the encoding conditions (*M* authentic = 1.89 s, *M* actors = 1.95 s, *M* non-actors = 1.79 s, linear mixed model comparison χ^2 = .114, *df* = 2, *p* = .946).

Rating Experiment

Design. The 240 stimuli were divided into four sets of 60 stimuli made up of 20 "authentic" stimuli, 20 "actors," and 20 "non-actors" expressions so that subjects were not confronted with the same sentence spoken by the authentic speaker, the actor, and the non-actor, in order to avoid a direct comparison of sentences. The 60 stimuli of one set were selected in such a way that there was neither a repetition of one specific stimulus (e.g., an authentic stimulus and the same stimulus re-enacted by an actor) nor a repetition of one speaker in one set. Each participant listened thus to only one fourth of the whole stimulus material.

The rating experiment was performed using the program NBS Presentation (Neurobehavioral Systems, Inc., Albany, California). Participants had to evaluate each of the 60 stimuli in regards to the specific vocal expression of emotion and to authenticity. During the experiment they either had to rate whether the stimulus represents "anger" [German original: Wut], "fear" [Angst], "sadness" [Trauer], or "joy" [Freude] (emotion recognition), and whether it is "play-acted" [gespielt], or "authentic" [echt] (authenticity recognition). The sets were pseudo-randomized to avoid serial repetition of trial order (order of the emotion and authenticity judgment task) more often than two times, of encoding condition ("authentic," "non-actors," or "actors") more often than two times, and of intended emotion more often than three times. This was done to reduce any systematic or pattern-related effects. In addition, the order of the four possible emotion-responses ("joy," "anger," "fear," and "sadness") and the two possible authenticity-responses ("authentic" and "play-acted") was counterbalanced per participants to avoid enhancement of a specific response by preferential effects for a specific response button.

Participants and experimental procedure. Participants for the rating experiment were recruited in the Cafeteria of the Georg-August-University of Göttingen and at the German Primate Center, Göttingen, Germany. They were all native German speakers. Two-hundred and twenty-eight subjects participated (69 female and 59 male) in the rating experiment. The subjects were students *(N* = 99) or scientific assistants (*N* = 29). Every single stimulus was thus rated by 32

subjects. Eighty of the subjects were between 18 and 24 years of age, 36 between 25 and 29 years, seven between 30 and 34 years and five subjects 35 years or older.

The stimuli were played back with a laptop (Toshiba Satellite M70-187 with a Realtek AC97 Soundcard) via NBS Presentation. Subjects heard the stimuli via earphones (Sennheiser HD 448 and HD 280 pro). Before the experiment started, subjects read a description about their task and the experimental procedure. All remaining questions were answered before the experiment started, after which there was no further interaction between participant and experimenter and the trials were played back automatically by Presentation as defined in the script.

Ethics. The study was approved by the ethics committee of the Georg-Elias-Müller-Institute of Psychology (University Göttingen). Professional and non-professional actors consented to the use of their shortened recordings in our rating experiment and to the anonymous acoustic analysis. Professional actors were paid 20 Euros for their participation. The non-professional actors and the participants of the rating experiment received candy bars for their participation. For the rating study, we did not obtain informed consent as data was collected and analyzed anonymously.

Acoustic Analysis

The acoustic analysis was conducted on two levels – on single vowels (Level 1) and on the short speech sequence (Level 2). Level 1) Vowels (a, e, i) were cut out of the speech tokens to obtain comparable units. For these vowels, we calculated the mean fundamental frequency (F0), the harmonic-to-noise-ratio, the frequency with the highest amplitude (peak frequency), the bandwidth of the first formant (hereafter “first formant”) and the amplitude ratio between third frequency band and F0 (hereafter “amplitude ratio”). All measurements were conducted using the spectrogram analysis software *LMA* (Lautmusteranalyse, developed by K. Hammerschmidt), except the first formant that was analyzed using *Praat* (Boersma & Weenink, 2009). We used this set of parameters, as they mirror the phonation process (F0, harmonic-to-noise ratio) and the

articulation (peak frequency, first formant, amplitude ratio). Furthermore, they are independent of each other, known to be affected by emotions (F0, harmonic-to-noise ratio, and peak frequency) and were already used in the comparison between the professional actors and the authentic recordings (see Jürgens et al., 2011).

Level 2) We measured the speech tempo and the variability of the F0-contour for the entire speech tokens that were used in the rating experiment. To determine the speech tempo, we calculated the speech rate (syllables/sec including pauses) and the articulation rate (syllables/sec excluding pauses). The variability of the F0-contour was measured via the standard deviation of the F0 for each speech token. All measurements were done manually using AvisoftSASLab. For a detailed description of the acoustic analysis see Jürgens and colleagues (2011).

Statistical Analysis

Recognition of encoding condition. The statistical analysis was done using R (R Developmental Core Team, 2012). Pure recognition rates reflect the behavior of the listener, but do not mirror the listeners' actual ability to distinguish the categories. High recognition rates in one condition might simply be caused by the participant's bias to only or preferentially choose the respective response category. Therefore, we calculated unbiased hit rates according to Wagner (1993). Unbiased hit rates reflect the probability of one participant that a stimulus is correctly recognized and that a response is correctly given, thus incorporating individual biases in response behavior. We tested the effect of *emotion* ("anger," "fear," "sadness," or "joy"), *encoding condition* ("authentic," "actors," or "non-actors") and their interaction on the recognition of encoding condition establishing a Linear mixed model (lmer function of the lme4 R package Bates, Maechler, & Bolker., 2011). As "actors" and "non-actors" provided two to three stimuli to the dataset, we had to deal with the dependency among our data. For the rating experiment, we divided the data set into four sets, so that participants rated only one speech token of every

actor. We then included *participant-ID* and *stimulus block* (1 to 4, representing the set, in which the stimulus was presented) as random effects into the statistical model to account for the influence of these variables. *Participant gender* was added as a fixed factor, while speakers' gender was not, as the calculation of individually unbiased hit rates incorporated every stimulus presented to one participant. Unbiased hit rates are proportions and were thus arcsine transformed prior to the analysis. The full model was compared to the null model (only including intercept and random factors) using a likelihood ratio test (function anova with the test argument "Chisq"), to establish the significance of the full model. The interaction effect of both categorical predictors was also tested using a likelihood ratio test. Afterwards, we conducted twelve post–hoc comparisons for all emotions between the three encoding conditions using the function glht (from the package multcomp Hothorn et al., 2008). P-values were adjusted using a Bonferroni correction.

Following the suggestion by Wagner, we calculated for every participant the chance probability that a stimulus is recognized correctly and compared the unbiased hit rates to the chance levels. The statistical model was built with hit rates as the response variable, *type of hit rate* (unbiased or chance), *encoding condition* and *emotion* as fixed factors, as well as *participant-ID* as a random effect. The full model was compared to the null model; post-hoc tests for *type of hit rate* (chance or unbiased hit rate) regarding every condition were done using the glht function, p-values were Bonferroni corrected.

Emotion recognition. Similar to the recognition of encoding condition, we calculated unbiased hit rates as well as the respective chance probabilities for the emotion recognition. The analysis followed the procedure mentioned above for the recognition of encoding condition.

Acoustic structure of vowels. Altogether we included 1176 vowels into the analysis, divided into 446 by authentic speakers, 346 by actors, and 384 by non-actors. We analyzed the acoustic parameters separately for vowels a, e, and i as a previous study revealed interaction effects between encoding condition and vowels (Jürgens et al., 2011). The effects on the acoustic

parameters *F0, harmonic-to-noise ratio, peak frequency, amplitude ratio,* and *bandwidth of the first formant (*hence *first formant)* were tested using linear mixed models (LMMs). *Emotion, encoding condition, speaker's gender,* and the interaction between *emotion* and *encoding condition* were entered as fixed factors. As random effects, we included *speaker-ID* and *stimulus-text.* Normal distribution and homogeneity of residuals were tested by inspecting Quartile-Quartile-Plots (QQ-plots) and residual plots. For the following acoustic parameters (a, e, and i) a deviation of assumptions was found and they were thus log transformed: the amplitude ratio, the peak frequency, and the first formant. We then compared the full models to the null models (function anova test argument "Chisq") to establish significance of the models. We tested for interaction effects by comparing the model including the interaction with the model excluding the interaction and used the reduced model when appropriate (function anova test argument "Chisq"). Main effects for fixed factors were also tested by model comparisons. We treated the acoustic parameters separately and adjusted the p-values with a Bonferroni correction for multiple testing within the three vowels. Finally, we conducted a post-hoc analysis for the acoustic parameters that were affected by encoding condition (function glht with Bonferroni adjustment for all possible comparisons).

Speed of speech. We tested the influence of *emotion*, *encoding condition* and their interaction on both *speech rate* and *articulation rate* by using linear mixed models (lmer from the package lme4). *Stimulus text and speaker-ID* was entered as a random effect. The assumptions of normal distribution and homogeneous residuals were tested by inspecting the QQ-plots and the residual plots. We then compared the full model to the null model using the likelihood ratio test of the function anova (test function "Chiq").

Pitch variability. We tested the influence of *encoding condition* on the *F0-standard deviation (F0-SD)* by using linear mixed models (lmer from the package lme4). We included *encoding condition* as predictor and *stimulus text and speaker-ID* as random effect into the model. To obtain a normal distribution of the data, *F0-SD* was log transformed. The assumptions of

normal distribution and homogeneous residuals were tested by inspecting the QQ-plots and the residual plots. The variability of pitch contour depends strongly on the type of sentence (that is exclamatory or interrogative sentence, beginning, or end of a sentence) or on stimuli length. Therefore, different sentences cannot be compared without restrictions. For the encoding condition, we could compare the same sentence across different modalities (as every sentence was present in every condition), which could not be done for the emotion effects. Hence, emotion was not regarded as a predictor for this analysis. We compared the full model to the null model (test anova, test function "Chisq") to look for an effect of *encoding condition* on *F0-SD* and used the function glht (with Bonferroni adjustment) for post hoc testing.

Results

Rating Experiment

Recognition of encoding condition. Encoding condition was correctly classified in 59.8% of all cases. The authentic recordings were recognized correctly as being authentic in 72.7% of all cases, while the actors' recordings were recognized as being play-acted in 57.5%, and the non-actors' recordings in 49.1% of all cases. The comparison to the null model established an overall effect of the predictors on unbiased hit rates (χ^2 = 314.31, *df* = 12, $p < .001$). Additionally, an interaction effect between emotion and encoding condition was found (interaction: χ^2 = 46.684, *df* = 6, $p < .001$, Figure 1). Post-hoc tests revealed that authentic speech tokens were recognized as such with the highest accuracy, regardless of the emotion category. Actors' expressions were recognized more accurately than the non-actors' expression, but only for anger, fear, and joy stimuli (Table 3.1, Figure 3.1). Participant gender (χ^2 = 0.2156, *df* = 1, p = .642) had no effect on the recognition of the encoding condition. The recognition rates were generally quite low, indicating a poor ability to judge encoding condition. In fact, recordings by non-actors of anger, fear, and joy were not recognized above chance level (post-hoc comparison for anger: z = 2.071, p = .460, fear: z = 2.861, p = .0505, and joy: z = 1.778, p = .904). All other conditions were recognized

above chance (all $z > 3.660$, $p < .003$). In sum, the *encoding condition* was only poorly recognized by the participants, but nevertheless differed between encoding conditions. Although the authentic stimuli were recognized at a higher rate than the acted stimuli, both acting conditions were often misjudged as being authentic as well. The portrayals by non-actors were significantly more frequently misjudged as being authentic than the expressions by professional actors.

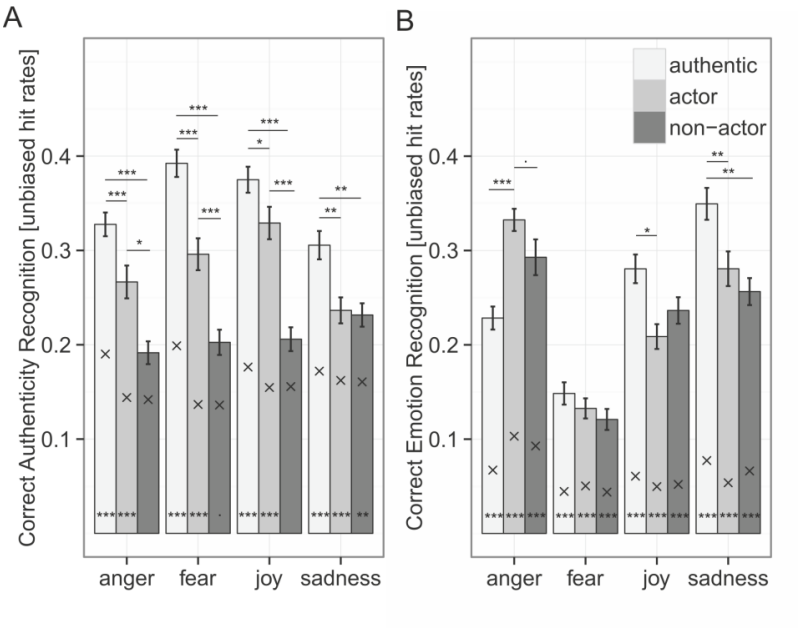

Figure 3.1 *Recognition of encoding condition (A) and emotion recognition (B) across all conditions. A) Unbiased hit rates for recognition of encoding condition (responding "authentic" when the stimulus is authentic, "play-acted" when stimulus is either from actors or non-actors), given as mean values ± SEM. B) Unbiased hit rates for the emotion classification (mean values ± SEM). X marks the mean chance levels per condition. Asterisks refer to the significance level . < .1; * p < .05; ** p < .01, *** p < .001. Asterisks at the bottom of the bars indicate differences between unbiased hit rates and individual chance level.*

Recognition of emotion. The different emotions were correctly recognized in 44.7% of cases, with a recognition accuracy of 46% for authentic stimuli, 45% for actors' recordings and 43% for non-actors' recordings. Regarding the recognition accuracy (unbiased hit rates), the full model, including *emotion*, *encoding condition* and their interaction, was significantly different from the null model ($\chi^2 = 251.55$, $df = 12$, $p < .001$). Furthermore, we found a significant interaction between *emotion* and *encoding condition* ($\chi^2 = 30.565$, $df = 6$, $p < .001$) (Figure 3.1). Post-hoc-tests between "authentic," "actors," and "non-actors" for all four emotions revealed that professionally acted anger was recognized better than authentic anger, while the reverse was true for joy and sadness. Non-actors' expressions showed a less accurate sadness recognition than the authentic speech tokens (Table 3.1). Overall, unbiased hit rates for expressions by professional actors were similar to the ones by non-actors, while both differed from the

expressions in the "authentic" condition. No effect of participant gender on the emotion recognition was found (χ^2 = 0.2601, *df* = 1, *p* =.610). Unbiased hit rates were very low, but differed in every emotion condition from the individual chance level, even in the fear condition (post-hoc comparison for every condition, unbiased hit rates – chance level: *z* > 4.94, *p* <.001). Fear was seldom recognized, but when participants gave the response "fear," they were mostly correct.

Table 3.1
Post-hoc comparisons for recognition of encoding condition and emotion (unbiased hit rates).

Type	Emotion	Encoding conditions		Estimate[a]	*SE*[a]	*z*-value	*p*[b]
Encoding condition	anger	authentic	actors	0.00848	0.002049	4.138	<.001***
		authentic	non-actors	0.01520	0.002049	7.418	<.001***
		actors	non-actors	0.00672	0.002049	3.279	.013*
	fear	authentic	actors	0.01057	0.002049	5.159	<.001***
		authentic	non-actors	0.02029	0.002049	9.901	<.001***
		actors	non-actors	0.00972	0.002049	4.743	<.001***
	joy	authentic	actors	0.00545	0.002049	2.661	.094
		authentic	non-actors	0.01795	0.002049	8.76	<.001***
		actors	non-actors	0.01250	0.002049	6.099	<.001***
	sadness	authentic	actors	0.00800	0.002049	3.902	.001**
		authentic	non-actors	0.00772	0.002049	3.769	.002**
		actors	non-actors	-0.00027	0.002049	-0.133	1
Emotion	anger	authentic	actors	-0.01111	0.002314	-4.8	<.001***
		authentic	non-actors	-0.00465	0.002314	-2.009	.535
		actors	non-actors	0.00646	0.002314	2.792	.063.
	fear	authentic	actors	0.00116	0.002314	0.501	1
		authentic	non-actors	0.00405	0.002314	1.749	.964
		actors	non-actors	0.00289	0.002314	1.248	1
	joy	authentic	actors	0.76010	0.002314	3.285	.012*
		authentic	non-actors	0.00464	0.002314	2.005	.539
		actors	non-actors	-0.00296	0.002314	-1.28	1
	sadness	authentic	actors	0.00776	0.002314	3.353	.0096**
		authentic	non-actors	0.00886	0.002314	3.829	.0016**
		actors	non-actors	0.00110	0.002314	0.47	1

Asterisks mark the significance level .<.1; *p < .05; **p < .01, ***p < .001
[a]Based on arscine transformed data
[b]Adjusted *p*-values (Bonferroni correction)

Acoustic Analysis

Acoustic structure of vowels. The variables in the analysis were affected by encoding condition, emotion, and gender (comparison of full models with null models; all $\chi^2 > 29.381$, $df = 12$, $p < .01$), except for the first formant (vowels "e" and "i") (Chi-statistics lower than $\chi^2 = 12.713$, $df = 12$, $p = 1$). We found no interaction between *emotion* and *encoding condition* in any of the acoustic parameters (Chi-statistics lower than $\chi^2 = 7.65$, $df = 6$, $p = .7$). Figure 3.2 shows the pattern of the acoustic variables in the different conditions. The acoustic profiles of all three conditions varied, with the biggest variation being between "non-actors" and "actors" speech tokens.

Table 3.2
Results of the linear mixed models on the acoustic structure of vowels

		Emotion[a]		Encoding condition[a]		Gender[a]		
Parameter	Vowel	$\chi^2_{(3)}$	p[b]	$\chi^2_{(2)}$	p[b]	$\chi^2_{(1)}$	p[b]	Estimates ± SE[c]
F0	a	15.34	.005**	0.42	1	43.56	<.001***	70.49 ± 9.72
	e	15.02	.005**	0.86	1	45.07	<.001***	69.72 ± 9.30
	i	14.48	.007**	3.41	.540	32.12	<.001***	70.18 ± 11.46
HNR	a	1.35	1	1.28	1	21.15	<.001***	0.046 ± 0.01
	e	3.32	1	0.19	1	22.99	<.001***	0.045 ± 0.009
	i	11.91	.023*	2.59	.822	41.78	<.001***	0.076 ± 0.010
Amplitude ratio	a	5.61	.397	52.07	<.001***	5.09	.072	-0.221 ± 0.099[d]
	e	1.74	1	26.15	<.001***	27.61	<.001***	-0.651 ± 0.117[d]
	i	11.65	.026*	3.79	.450	32.44	<.001***	-1.028 ± 0.164[d]
Peak Frequency	a	21.42	<.001***	24.22	<.001***	0.76	1	0.067 ± 0.0774[d]
	e	8.13	.13	17.53	<.001***	6.99	.024*	0.169 ± 0.064[d]
	i	0.55	1	24.58	<.001***	6.50	.032*	0.121 ± 0.048[d]
First formant	a	4.66	.198	19.08	<.001***	5.00	.025*	0.190 ± 0.086[d]

Asterisks mark the significance level .<.1; *p < .05; **p < .01, ***p < .001
[a]Statistical values are obtained from the model comparison (full model to reduced model excluding the respective predictor).
[b]*p*-value adjustments (Bonferroni correction) were done for the different vowels within one acoustic parameter and one predictor.
[c]Estimates for the predictor gender were gained from the *LMM*, with male speakers included in the intercept. Estimates refer to the female speakers in comparison to the male speakers.
[d]Values base on log transformed data

Notably, acoustic differences were putatively related to articulation. Specifically, amplitude ratio, peak frequency, and the bandwidth of the first formant (Table 3.2 and Figure 3.2) varied between "authentic," "actors," and "non-actors" recordings. The professional actors' speech tokens differed most strongly from the other two encoding conditions. The authentic and the non-actors' recordings also varied between each other, although they deviated in a similar way from the actors' speech tokens. For instance, professional actors had a lower amplitude ratio (referring to a pronounced F0 and a less intense third frequency band) and wider bandwidths of the first formant compared to the "authentic" and the "non-actor" expressions; "authentic" and "non-actors" did not differ in their formant bandwidths, but "non-actors" vowels possessed higher amplitude ratios than the "authentic" ones (see Table 3.3 for the results of the *post-hoc* analysis). To sum up, the acoustic structure varied between "authentic," "actors," or "non-actors" recordings, but encoding condition affected other variables than those affected by emotion (see below).

The factor *emotion* influenced the parameters F0 and peak frequency, as well as to a lesser degree the harmonic-to-noise (for vowel "i") and the amplitude ratios (for vowel "i"). Anger stimuli deviated most strongly from the other emotions by possessing higher F0 and peak frequencies (Figure 3.2). Speaker gender influenced the acoustic structure of vowels most strongly (Table 3.2). Women spoke with a higher F0 and peak frequency, increased bandwidths of the first formant, higher harmonic-to-noise ratio and lower amplitude ratio.

Table 3.3

Results of the post-hoc analyses on the influence of encoding condition on the acoustic structure of vowels

Parameter	Vowel	Encoding condition		Estimate	*SE*	*z*-value	*p*[a]
Amplitude Ratio	a	authentic	actor	0.552	0.115	4.949	<.001***
		authentic	non-actor	-0.423	0.109	-3.895	<.001***
		actors	non-actors	-0.975	0.123	-7.941	<.001***
	e	authentic	actor	0.385	0.121	3.177	.004**
		authentic	non-actor	-0.323	0.125	-2.591	.029*
		actors	non-actors	-0.709	0.133	-5.323	<.001***
Peak frequency	a	authentic	actor	0.196[b]	0.077[b]	2.550	.0323
		authentic	non-actor	-0.231[b]	0.075[b]	-3.093	.006**
		actors	non-actors	-0.428[b]	0.084[b]	-5.077	<.001***
	e	authentic	actor	0.103[b]	0.567[b]	1.814	.209
		authentic	non-actor	-0.163[b]	0.058[b]	-2.798	.015*
		actors	non-actors	-0.266[b]	0.063[b]	-4.243	<.001***
	i	authentic	actor	0.022[b]	0.050[b]	0.450	1
		authentic	non-actor	-0.206[b]	0.045[b]	-4.555	<.001***
		actors	non-actors	-0.229[b]	0.052[b]	-4.381	<.001***
First formant	a	authentic	actor	-0.444[b]	0.105[b]	-4.299	<.001***
		authentic	non-actor	-0.052[b]	0.102[b]	-0.511	1
		actors	non-actors	0.392[b]	0.114[b]	3.442	.002**

Asterisks mark the significance level .<.1; *p < .05; **p < .01, ***p < .001

[a]Adjusted *p*-values (Bonferroni correction)

[b]Values base on log transformed data

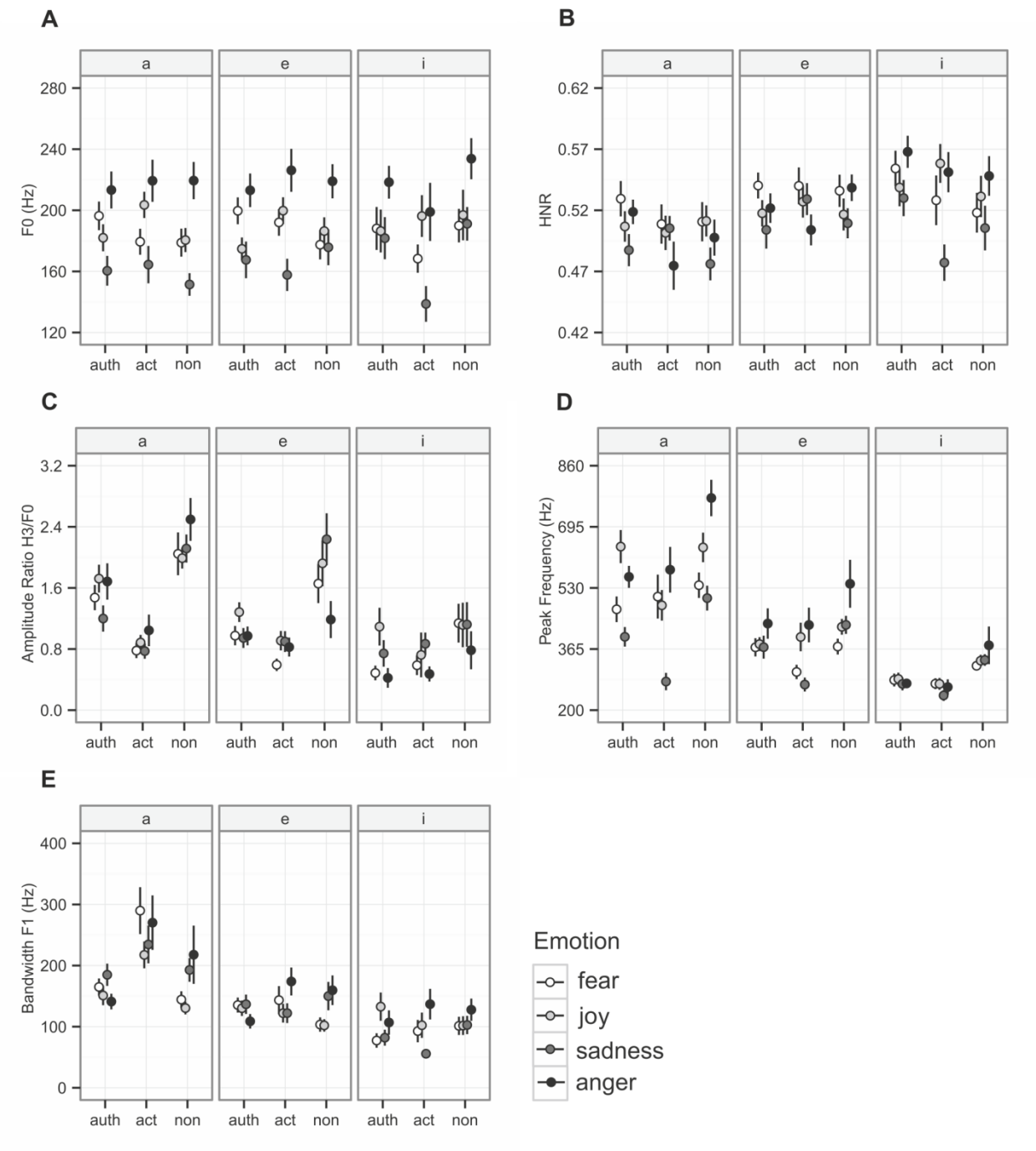

Figure 3.2 *Selected acoustic parameters separated for emotion and encoding condition. Mean values are given for vowels a, e, i ± SEM.*

Speed of speech. Speech rate and articulation rate were not affected by the three encoding conditions, the four emotion categories, or their interaction (comparison of full with null model: speech rate χ^2 = 9.2631, *df* = 11, *p* = .598; articulation rate χ^2 = 5.797, *df* = 11, *p* = .887, Figure 3.3).

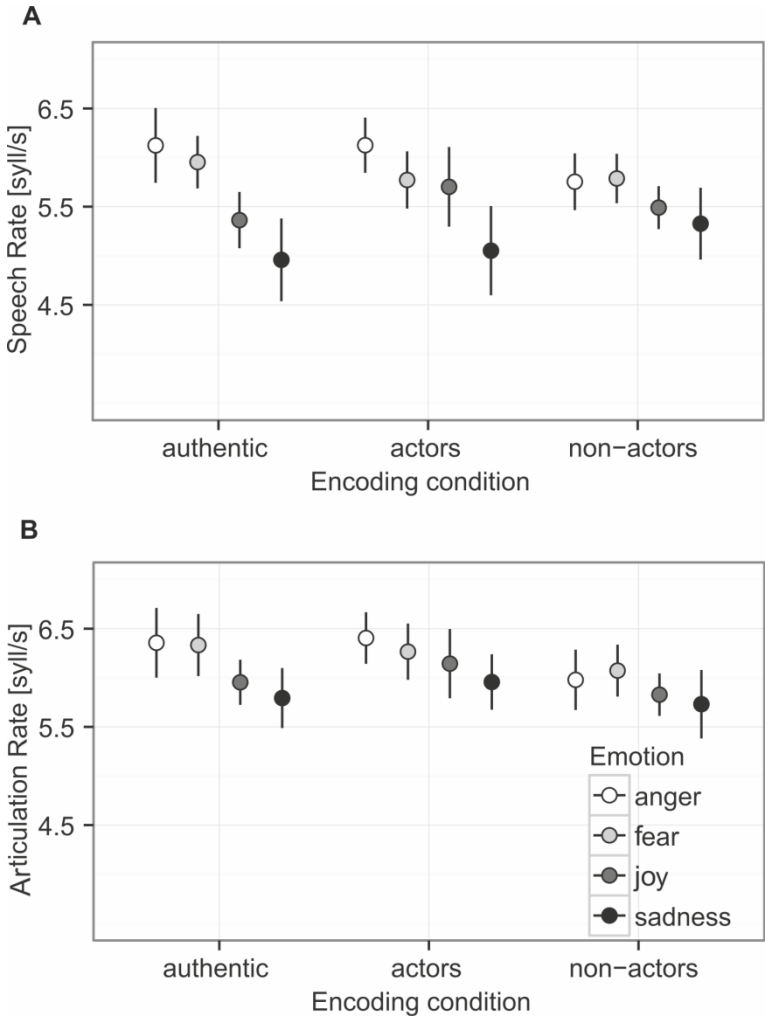

Figure 3.3 *Speech tempo for the different encoding conditions and emotion categories. Mean ± SEM is given for A) speech rate and B) articulation rate.*

Pitch variability. Pitch variability *(F0-SD)* was affected by *encoding condition* (χ^2 = 13.48, *df* = 2, *p* = .001). The post-hoc comparison revealed a flatter prosody in authentic speech compared to both acting conditions (authentic – actors: estimates (log transformed data) ± *SE* = -0.3114 ± 0.097, *z* = -2.879, *p* = .012; authentic – non-actors: -0.375 ± 0.109, *z* = -3.45, *p* = .002, non-actors – actors: 0.063 ± 0.116, *z* = 0.547, *p* = 1). Pitch variability for authentic speech was 23.24 Hz (± 1.94 Hz SEM), while actors speech was characterized by a variability of 31.00 Hz (± 2.41 Hz SEM), and non-actors speech by 33.07 Hz (± 2.36 Hz SEM).

Discussion

In terms of their acoustic characteristics, vocal expressions of emotions delivered by professional actors were not more similar to authentic expressions than the ones by non-actors. Moreover, vocal expressions by professional actors and non-actors evoked similar recognition patterns. Thus, our findings do not support the hypothesis that compared to non-actors, professional actors have a superior ability to produce emotional portrayals that resemble spontaneous expressions (hypothesis 1). Our results furthermore do not support the view that play-acted expressions in general, and the ones by non-professional actors in particular, are necessarily stereotyped caricatures of authentic expressions. The lack of an interaction between encoding condition and emotion in the acoustic variables we analyzed indicates that emotions were expressed in a similar fashion in all of the three recording types (but see Spackman et al., 2009 for contrasting results).

Nevertheless, we found acoustic differences between the encoding conditions. Acting in general was distinguished from the authentic recordings by a more variable speech melody (see also Audibert et al., 2010; Williams & Stevens. 1972). Interestingly, high variability in pitch contour has been related to more aroused emotions, such as anger, while low variability characterize less aroused expressions such as sadness (Juslin & Laukka, 2003; K. R. Scherer, 1986). The different intonation may thus interact with emotion perception, which might explain the differences in emotion recognition for authentic speakers, actors and non-actors. The variable speech melody in play-acted expressions might be confounded with anger perception, affecting the high recognition rates for play-acted anger. Low variability on the other hand might be misinterpreted as sadness, and facilitates the sadness recognition for authentic speech tokens.

As in previous studies, the recognition of encoding condition was rather low (Drolet et al., 2012; Jürgens et al,. 2013; Porter & ten Brinke, 2009). Listeners were thus unable to reliably recognize whether an expression was acted or not. Expressions by experienced actors were rated as “play-acted” more often than the ones by non-actors, which were not recognized as “play-

acted" above chance level. The differences in articulation supports this notion, as acoustic profiles of non-actors' speech resembled the structure of authentic speech tokens, while the acoustic profiles of professional actors differed from the other two categories. The acoustic differences of vowels between the encoding conditions might be caused by variation in articulation, as a result of speech training, rather than on differences in emotion encoding between acting and spontaneous expression. These results indicate that acting training interacts with the perception of authenticity and supports hypothesis 2 (see also Krahmer & Swerts, 2008).

Although listeners appear to process some of the acoustic variation between encoding conditions implicitly (Drolet et al., 2012, 2014), they appeared to be unable to use these cues for recognizing the encoding condition, as evidenced by the poor hit rates. One important open question is the source of variation in speech melody, specifically whether it is due to acting, or to variation between speaking styles, such as reading (Batliner, Kompe, Kießling, Nöth & Nieman, 1995; Eskenazi, 1992; Laan, 1997). Considering that during the acting conditions the actors and non-actors were not asked to learn the sentences by heart, the differences may also reflect the reading and not the acting process. Future studies will be needed to clarify this issue.

The effect of emotion on the acoustic structure is consistent with the literature (Hammerschmidt & Jürgens, 2007; K. R. Scherer, 2003), although effect sizes (as well as the corresponding recognition rates) for both the acted and spontaneous expressions were weak compared to previous studies on vocal expressions (Hammerschmidt & Jürgens, 2007; Laukka et al., 2005; K. R. Scherer, 2003). Our speech tokens were not preselected and were taken from long text sequences in which not all of the words were equally emotionally accentuated (in contrast to studies in which only one word was expressed emotionally). This procedure seems suitable for avoiding exaggerated portrayals and thus increasing realism of play-acted expressions. The low emotional content may reflect the actual emotionality transmitted via speech (in contrast to emotional outbursts) and emphasized the importance of studying realistic expressions to understand daily communication.

The fact that vocal expressions of emotions are so easily play-acted, without being detected as such, contributes to the discussion about the reliability of emotional expressions. The question is whether emotion expressions need to be tied to specific affective states including subjective feeling or physiological changes. (Dezecache et al., 2013; Fernández-Dols & Crivelli, 2013; Mehu & Scherer, 2012). For facial expressions, the Duchenne smile (smiling including the zygomaticus major and the orbicularis oculi muscle) was suggested to signal felt happiness only, while smiles without activation of the orbicularis oculi were classified as faked (Ekman et al., 1990). Recent studies, however, demonstrated the common use of Duchenne smiles in acted expressions (Krumhuber & Manstead, 2009, for discussion Riediger, Studtmann, Westphal, Rauers, & Weber, 2014). As noted above, we did not measure physiological state, or subjective feelings of our spontaneous speakers and do not claim to test coherence between feelings and expressions. However, the general similarity in expression patterns in the three encoding conditions is in keeping with the view that in humans, expressions of emotions can be successfully decoupled from subjective feelings (Fernández-Dols & Crivelli, 2013).

As in every study using daily life data, our study does suffer from some limitations, mostly due to the nature of our authentic expressions being recorded by radio reporters. For this reason, we were unable to create fully comparable stimuli. The recording quality of real-life situations may be substantially worse compared to the play-acted ones. We did, however, try to obtain play-acted recordings under a variety of acoustic conditions. If the higher recognition rates for authentic stimuli were simply explained by the recording quality, play-acted expressions would have had much higher recognition rates than they actually did (for other studies finding a bias to preferentially choose authentic see Gosselin et al., 1995; Jürgens et al., 2013; Levine et al., 1999). In any case, our current study aimed to compare play-acted expressions by trained professional actors and non-actors and these recording conditions were equivalent. Another limiting aspect is the fact that our sample of decoders consisted mainly of students, thus being rather homogeneous and limited for a generalization of results. However, in a preceding study, the

authentic and professionally acted stimuli were rated by students from three different cultures (Germany, Romania, and Indonesia) (Jürgens et al., 2013). All three study populations showed highly similar rating patterns, suggesting that some of our findings may be generalizable. On the other hand, our sample of speakers was rather heterogeneous, with non-actors being older than the professional actors. Previous studies showed that decoders are generally more accurate in judging expressions from their own age category, while older encoders are supposed to express emotions less distinctly (Borod et al., 2004; Riediger, Voelkle, Ebner, & Lindenberger, 2011), both suggesting an advantage for the younger encoder group ("actors"). This effect does not appear to be strong in our sample, as both acting conditions were rated similarly, despite the age difference of the speakers. One advantage of our stimulus set is the high number of speakers, which minimizes the probability that our results are based on individual differences between speakers rather than on the group differences. Future studies should focus on individual differences in emotion encoding to disentangle the effects of age, gender, non-professional acting experience, voice use, and even current mood of speakers on the play-acting of emotion, factors that we could not consider in our analysis.

In summary, our study centered on methodological issues that may have strong effects on the interpretation of previous results, and are relevant for the planning of future studies. We showed that compared to professional actors, non-actors are equally capable of transmitting emotional information via the voice when asked to portray an emotion; additionally non-actors' expressions were perceived as more realistic. For future studies, recordings of daily life emotion expressions should clearly be preferred, but recording spontaneous emotion behavior is unfortunately rarely possible. As an alternative, our findings on vocal expressions speak for the use of non-professional actors when realistic stimulus databases are required.

Appendix

3.1 Examples of transcripts and stimulus texts used for the play-acted recordings (taken from Jürgens et al., 2011, Appendix 1).

Only the words in quotation marks were used for the rating study and the acoustic analysis.

- Male Spoken Anger

Context

Two fighting dogs attacked 6-year-old V. in the schoolyard. He was bitten to death. Fighting dogs are a big problem in the area and people do not feel protected by the police. They are furious and are looking for a culprit. The anger is directed to the police. The people are shouting at a police officer, blaming him for being too late.

Man:

Original (German): Der Kiosk ruft vor Viertelstund an, "*nach Viertelstund*" kommt ihr erst, oder was?"

Translation: The kiosk called 15 minutes ago, you only come "*after 15 minutes*" or what?

- Female Spoken Sadness

Context

The 73-year-old W. was attacked in his shop by two 16-year-old boys. He was robbed and stabbed to death. It is the date of the funeral. A weeping woman reports.

Woman:

Original (German): "*Ich kenn den 43 Jahr*." Und er war für uns alle ein Freund. Und ich finde das furchtbar, was da passiert ist.

Translation: "*I have known him for 43 years*". And he was a friend, for all of us. And I think what happened is dreadful.

- Male Spoken Joy

Context

The Fall of the Berlin Wall. A citizen of the German Democratic Republic reports excitedly and happily about the border crossing.

Man:

Original (German): Vorhin haben sie noch einzeln durchgelassen. Dann haben sie das Tor aufgemacht, "*und jetzt konnten wir alle*" so, wie wir waren, ohne vorzeigen, ohne alles, konnten wir gehen.

Translation: Previously they let the people pass individually. Then they opened the gate and "*now we could all*", as we were, without showing anything, without everything, we could go.

- Female Spoken Fear

Context

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.

Woman:

Original (German): 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 dolle Angst

Translation: 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.

4 Biographical similarity does not affect vocal emotion processing

Rebecca Jürgens[a,b], Julia Fischer[a], Annekathrin Schacht[b]

[a]Cognitive Ethology Laboratory, German Primate Center, Göttingen, Germany

[b]CRC Text Structures, University of Göttingen, Göttingen, Germany

Prepared for submission

Abstract

Similarity between two individuals may immediately create a social connection that positively affects their interaction. Sharing this connection putatively leads to a more attentive processing of social signals and might increase empathic concern. This study addresses the question to which degree such social links have an effect on the recognition of emotions in human speech, to test the hypothesis that facilitated empathic concern induce more accurate emotion recognition. We aimed to investigate whether manipulated similarity in terms of biographical data between a fictive speaker and the participant, increases the recognition of vocal emotion expressions and intensifies emotional engagement. Experiment 1 concentrated on vocal emotion recognition, while Experiment 2 integrated autonomic measures (pupil size and skin conductance) to investigate emotional engagement. As a control, we additionally investigated the processing of affective sounds like dentist drills or baby cries. In experiment 3, we investigated the effect of cognitive load on emotion recognition. Surprisingly, we found no effect of similarity on emotion recognition. Autonomic reactions to angry and joyful vocal expressions were in general low compared to the processing of affective sounds. Pupil dilation however differed in response to emotion categories, for both vocal expressions and sounds. Our findings revealed that biographic similarity may not necessarily affect emotion processing and that brief excerpts of vocal expressions alone may not trigger perceptible autonomic reactions. Future studies should make use of a more holistic approach when investigating emotional engagement.

Introduction

Sharing attitudes, interests, and personal characteristics with another person may immediately create a social link (Jones, Pelham, Carvallo, & Mirenberg, 2004; Miller, Downs, & Prentice, 1998; Vandenbergh, 1972; Walton et al., 2012). Similarity, even with regard to such irrelevant pieces of information as the date of birth, increases the willingness to help (Burger, Messian, Patel, del Prado, & Anderson, 2004), to cooperate (Miller et al., 1998), to adopt the goal and motivation of others (Walton et al., 2012) as well as to share their emotions (Cwir et al., 2011). Even in narrative settings, our perception and evaluation of fictive characters is strongly influenced by perceived similarity (Maccoby & Wilson, 1957; Raney, 2004). As cognitive resources are limited, humans seem to selectively allocate their attention in interpersonal exchanges towards individuals of relevance (Ackerman et al., 2006). We here address the question whether this positive effect of social connectedness facilitates the recognition of emotion expressions in others and whether facilitation is caused by an attention-shift or by increased empathic concern.

The recognition of emotions in others is an integral aspect of interpersonal exchange. We interact differently with people, in which we recognize anger or happiness, respectively. Emotions may be expressed by the face (Ekman & Friesen, 1978; K. R. Scherer & Ellgring, 2007), body postures (De Gelder & Van den Stock, 2011), or the voice (Banse & Scherer, 1996; Hammerschmidt & Jürgens, 2007). Importantly, emotions are not only recognized in others, but are probably also shared between both interaction partners – humans “catch the emotion” of others (Hatfield et al., 2011). Attending to emotion expressions may induce congruent facial muscle movements (called facial mimicry, Dimberg, 1982; Dimberg & Thunberg, 2012), or influence subjective experience (Wild et al. 2001). This mirroring of someone else’s emotional state may be mediated via automatic processes of emotional contagion or via more deliberate, conscious processes of empathy (Preston & de Waal, 2002). Emotional contagion and empathy are presumed to be of importance for understanding the inner affective state of others (Goldman & Sripada, 2005; Niedenthal & Maringer, 2009; Sato et al., 2013). Recent studies demonstrated

the link between facial mimicry, subjective experience, and the recognition of emotional expressions (Künecke et al., 2014; Sato et al., 2013, but see Blairy et al., 1999). While empathy might occur in a wide range of situations, it has been revealed that the strength of an empathic response is influenced by appraisal processes and is not just an automatic reaction (de Vignemont & Singer, 2006; Preston & de Waal, 2002). People do seemingly not empathize equally with everyone, but show increased reactions towards people that are more relevant for them, such as people that are more likable, more familiar, share group-member ship with them, or are similar (Cwir et al., 2011; de Vignemont & Singer, 2006; Mathur et al., 2010; Preis & Kroener-Herwig, 2012; Singer et al., 2006). Cwir et al. (2011) for example recorded cardiovascular activity of people witnessing others preparing a public speech. Cardiovascular activity in the beholder, that indicates a shared stress reaction, was higher when both people had common interests than when both were not connected at all. Following this line of thought, social connectedness or relevance between speaker and listener might thus affect emotion recognition, either by an attention-shift, or by increased emotional engagement.

Jürgens et al. (2013) demonstrated that vocal expressions of emotions in daily life situations are only poorly recognized by (socially unconnected) listeners, in contrast to the highly intense and stereotypical expressions commonly used in research (cf. K. R. Scherer et al., 2001). Assuming that vocal emotion transmission works in daily life, emotion communication might be based on the social link between interaction partners and is facilitated when signaler and the signal are more relevant. For group membership - a second factor that increases social relevance (Ackerman et al., 2006; Brown et al. 2006) - a positive effect on simple emotion recognition for facial expression was found. Emotions expressed by other in-group members are classified faster and more accurately in comparison to those of out-group members (Elfenbein & Ambady, 2002; Thibault et al. 2006; Weisbuch & Ambady, 2008), even when the groups were only randomly created during the experiment (Young & Hugenberg, 2010). Group membership may also, however, elicit negative connotation like prejudices (Bijlstra et al., 2010; Hugenberg &

Bodenhausen, 2003) that could cause additional effects concerning the recognition of emotional expressions. European Americans for example preferentially perceived anger in emotional faces by African Americans, while they did not have this bias in regard to other European Americans (Hugenberg & Bodenhausen, 2003). Anger might be more relevant to out-group members, while happiness for example might be perceived as more important in in-group members (Bijlstra et al., 2010; Weisbuch & Ambady, 2008). Similarity does presumably not trigger these negative connotations in contrast to group-membership and might hence be more adequate to investigate whether the perception of emotional expressions is directly influenced by social relevance between sender and speaker.

The main aim of the present study is to investigate the impact of speaker-listener similarity on the recognition of vocal expressions of emotions. Vocal expressions are only scarcely studied in the context of empathic reactions or other context effects. Similarity is expected to increase the attention towards emotional stimuli or to facilitate empathic concern, which should both positively affect emotion recognition, although underlying autonomous reactions would be different. To create social connectedness, we manipulated the similarity of speakers by providing the listeners with (fictive) biographical information of the speaker, such as place of birth, education, age, or leisure activities. In the case that the processing of emotional expressions base on emotional engagement of the perceiver, we expected increased activation of the autonomic nervous system (ANS) to vocal expressions of emotion, which should be even enlarged under conditions of speaker-listener similarity (Brown et al., 2006; Cwir et al., 2011).

In Experiment 1, we investigated the processing of prosodic information at the behavioral level and tested whether emotion recognition is facilitated when the protagonist of a short story is similar (that is more relevant) to the listener. Experiment 2 focused on the question whether emotional prosody elicits enhanced physiological responses in the listener, and whether these responses are modulated by similarity. In addition to emotional expressions, participants had to categorize affective sounds, such as bombs or baby cries (Bradley & Lang 2000). With this

manipulation we aimed at controlling differences in physiological sensitivity to affective stimuli in the auditory modality. Experiment 3 serves as control experiment, in which we detailed the analysis of recognition performance for affective sounds and vocal expressions, in order to reveal the cognitive difficulties during emotion recognition.

Experiment 1

In this experiment, we investigated the effect of simulated similarity between listener and fictive speaker on the recognition and perception of vocal expressions of emotion. Vocal expressions of five different emotion categories (anger, fear, joy, sadness, and neutral) were integrated in short acoustically-presented stories to allow the participant "to get into" the character. Every story ended with a vocal emotion expression. We hypothesized that emotion expressions by fictive characters resembling the perceivers in their biographical data were recognized more accurately and perceived as more intense.

Methods

Ethics. All experiments of the present study were approved by the local Ethics committee of the Georg-Elias-Müller Institute of Psychology at the Georg-August-University of Goettingen. All participants were fully informed about the procedure and gave written informed consent prior to the experiment.

Participants. Thirty-eight German native speaker (20 women, aged 19 – 30 year, *M* = 22.9) participated in this study. All of them were undergraduate or graduate students of the University of Goettingen. Participation was reimbursed by course credits or 8 Euro/hr.

Stimuli. Stimulus material consisted of ten different short stories of neutral content which were followed by target sentences of different emotional prosody. The stories had a word count of about 150 words each and dealt with daily life situations, such as sitting in a cafeteria, going shopping or returning home (see Appendix 4.1 for an example) in third-person versions. Every

story had one protagonist performing the respective actions and ended with the beginning of a direct speech, such as “then he/she says:” [German Original: ”dann sagt er/sie:”].

Each story was recorded from two different speakers (one male (28 years) and one female (25 years)) who read the texts with neutral prosody. The spoken versions of the texts did not differ in their mean duration, *M* = 48.4 sec (± 3.0 sec) for the female speaker and *M* = 47.8 sec (± 1.9 sec) for the male speaker. We constructed ten different target sentences that were controlled for length (six syllables each) and, importantly, of neutral content semantically related to one of the stories, such as “What have you said?” [“Was hast du da gesagt?”] or “I have expected this one.“ [“Den hab’ ich erwartet.”]. The prosodic sentences were recorded from unprofessional speakers (students and members of the University of Goettingen; 14 female and 17 male, mean age 27.4 ± 4.9 years). During recording, the speakers were instructed to express the given sentence in a specific emotion category (anger, joy, fear, sadness, or neutral, respectively), and were allowed to repeat the sentence until they felt satisfied with their performance. To facilitate the emotion expression, we provided the speakers with short vignettes that described an emotion scenario (overlapping with the short stories used in the experiment).

To select appropriate target sentences, we conducted a pre-experimental rating, in which the emotional content of each sentence was rated by N = 20 - 22 participants (balanced between women and men). Participants listened to a subset of sentences and specified whether the speech token was “anger,” “neutral,” “joy,” “fear,” or “sadness”. Stimuli were selected that revealed emotion recognition rates between 57 – 81% (mean 70%), to ensure emotion recognition above chance while avoiding ceiling effects. The final set consisted of 40 prosodic sentences (by 13 women (*M* age 26.3 ± 3.9 years) and 14 men (*M* age 28.1 ± 6.3 years); four stimuli per emotion category per gender), with no repetition of a given speaker per emotion category or sentence. The sentences were connected to the stories in the following way: For each gender, we constructed two different stimulus sets. Each stimulus set included all of the ten stories, combined with one prosodic sentence, so that every emotion category was included twice in a

given set. The recognition rates for the two stimuli per emotion category in every stimulus set were comparable to each other according to the pre-experimental ratings (mean differences in %).

On the basis of participants' demographic data obtained prior the main experiment, such as first name, date and place of birth, field of study, place of domicile, living situation and hobbies, we constructed personal profiles of the fictive protagonists. They either resembled or differed from the participants' profiles. Similarity was created by using the same gender, first name (or similar equivalents, e.g., *Anna* and *Anne*), same or similar dates and places of birth, same or similar study program, and same hobbies. Dissimilar characters were characterized by not being a student, being around 10 years older, not sharing the birth month and date, living in a different federal state of Germany, having a dissimilar first name, and being interested in different hobbies. Manipulations for every participant were done using the same scheme.

Procedure. After arriving at the laboratory, participants filled out the informed consent and the demographic questionnaire that was introduced as being necessary to control for individual characteristics. Afterwards, participants had to fill out a set of questionnaires (that were not regarded further in the analysis) to give the experimenter time to conduct the similarity manipulation. Participants were then seated in front of a computer screen and listened to the ten stories in a randomized order. Prior to a given story, they were presented with one of the personal profiles on screen (for 7 seconds) of the story's main character (similarity manipulation). Every emotion category of the target sentences was once connected to a similar profile and once to a dissimilar profile. Participants were instructed to read the profiles carefully and to vividly imagine the character and the situation described by the short story and finalized by an utterance spoken by the fictive character in specific emotional prosody. After the target sentence, participants had to specify the expressed emotion (forced choice options "anger," "fear," "sadness," "joy," "neutral") via mouse selection and to rate the intensity of the expression on a 1-5-likert scale. The experiment lasted about 20 minutes.

Statistical analysis. Statistical analyses were done in R (R Developmental Core Team, 2012). Correct emotion recognition was analyzed using a Generalized linear mixed model with binomial error structure (GLMM, lmer function, R package lme4 Bates et al., 2011). We included emotion category (anger, joy, fear, sadness, neutral), similarity (similar, dissimilar) and their interaction as fixed factors into the model, and added participant-ID as random effect. Intensity ratings were analyzed using a cumulative link mixed model for ordinal data (package ordinal, Christensen, 2012) that also included emotion category, similarity and the interaction as fixed effect, and participant-ID as random effect. Models were compared to the respective null models only including the random effects via likelihood ratio test (function anova). In addition, model comparisons were conducted to test for interaction or for an effect of similarity on the model. Post-hoc tests were conducted using the glht function of the multcomp package (Hothorn et al., 2008) with Bonferroni correction.

Results

We established the full model for emotion recognition (χ^2 = 26.47, *df* = 9, *p* = .002) and intensity ratings (LR.stat = 21.89, *df* = 9, *p* = .009). However, neither the interaction (emotion recognition: χ^2 = 6.57, *df* = 4, *p* = .160; intensity rating: LR.stat = 5.25, *df* = 4, *p* = .262) nor the similarity (emotion recognition: χ^2 = 0.9, *df* = 1, *p* = .343; intensity rating: LR.stat = 0.87, *df* = 1, *p* = .351) had a significant effect on emotion recognition and intensity rating. Both emotion recognition (χ^2 = 19.04, df = 4, *p* < .001) and intensity ratings (LR.stat = 15.911, *df* = 4, *p* = .003) only differed with regard to emotion category (Figure 4.1). Specifically, neutral prosody was recognized less accurately and perceived as less intense in comparison to other emotional expressions (Table 4.1).

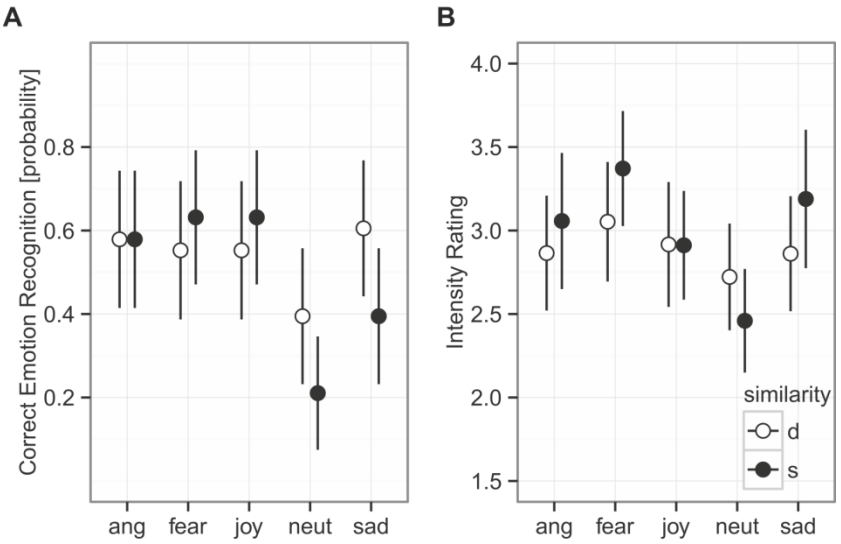

Figure 4.1 *Correct emotion recognition (A) and intensity ratings (B) for Study 1. Given are the mean values ± 95 % CI. d = dissimilar, s = similar.*

Table 4.1
Results of the post-hoc tests for emotion recognition and intensity perception

Rating	Emotion		Estimates	*z*-value	p^a
Emotion recognition	fear	anger	0.054 ± 0.33	0.165	1
	joy	anger	0.054 ± 0.33	0.165	1
	neutral	anger	-1.156 ± 0.34	-3.385	.007**
	sadness	anger	-0.319 ± 0.33	-0.977	1
	joy	fear	0.000 ± 0.33	0.000	1
	neutral	fear	-1.210 ± 0.34	-3.537	.004**
	sadness	fear	-0.374 ± 0.33	-1.140	1
	neutral	joy	-1.210 ± 0.34	-3.537	.004*
	sadness	joy	-0.373 ± 0.33	-1.140	1
	sadness	neutral	0.837 ± 0.34	2.465	.137
Intensity rating	fear	anger	0.434 ± 0.30	1.441	1
	joy	anger	-0.113 ± 0.30	-0.375	1
	neutral	anger	-0.735 ± 0.30	-2.408	.160
	sadness	anger	0.061 ± 0.30	0.198	1
	joy	fear	-0.547 ± 0.29	-1.881	.599
	neutral	fear	-1.169 ± 0.34	-3.441	.006**
	sadness	fear	-0.372 ± 0.26	-1.460	1
	neutral	joy	-0.621 ± 0.37	1.674	.942
	sadness	joy	0.175 ± 0.304	0.574	1
	sadness	neutral	0.796 ± 0.35	2.273	.230

Note: [a]Adjusted p-values (Bonferroni correction)

Discussion

Similarity did not affect the recognition of emotion or intensity ratings. The lack of similarity effects on both recognition performance and intensity ratings in our data might be caused by several (methodological) reasons and does not necessarily imply that personal similarity is irrelevant for processing emotional information in the voice. The context stories might have interacted not only with the emotion recognition, but also with the similarity manipulation. A participant, listening to a story of a character drinking coffee, while he himself detests coffee, is most likely to be disrupted in his similarity illusion. Further, the stories in combination with the high number of emotion categories might have complicated the study, restricted the interchangeability of prosodic stimuli and limited possible repetition within participants.

Overall, both emotion recognition and perceived intensity of emotional expressions were reduced here in comparison to our pre-experimental ratings in which the same target sentences were presented but in isolation. Particularly the recognition of utterances spoken with neutral prosody suffered from the context condition, even though all stories were of both neutral content and prosody. Presumably, the stories led participants to expect emotional endings, which seems plausible considering the typical emotionality in human interactions (Vrana & Rollock, 1998). These expectations might strongly depend on individual differences, such as moods or knowledge (Halberstadt, Dennis, & Hess, 2011; Schmid & Schmid Mast, 2010).

Experiment 2

The main aims of this experiment were twofold: First, we wanted to optimize the experimental design of Experiment 1. Therefore, we reduced the number of emotion categories to three, namely anger, joy, and neutral. Furthermore, we replaced the context stories with short context sentences to allow an increase in repetitions and to diminish the uncontrollable effects of the stories. In order to elicit strong emotional engagement in the listeners and to increase the quality as well as the comparability of the prosodic stimuli, vocal expressions of emotions were

taken from a database of professional and intensely acted stimuli (Burkhardt, Paeschke, Rolfes, Sendlmeier, & Weiss, 2005). Additionally, we restricted the sample to female participants only. That was done for the practical reason to only need one stimulus set by female speakers. Beyond that, men and women are found to vary in their emotional reactivity (Bradley, Codispoti, Sabatinelli, & Lang, 2001; Kret & De Gelder, 2012). Second, we included peripheral measures (pupil dilations and electrodermal responses), which are known to reflect even very subtle emotion-related physiological chances and to be mainly robust against conscious evaluative appraisals.

Reactions of the autonomic nervous system, related to emotional episodes, include changes in the cardiovascular system, respiration and perspiration (Kreibig, 2010; Moors et al., 2013). Attending emotions in others is thought to elicit reflections of the emotional states in the listeners, including the corresponding autonomic reactions (Hatfield et al., 2011; Preston & de Waal, 2002). To investigate this emotional engagement and to reveal possible modulations by similarity, we recorded two peripheral physiological measures. Skin conductance response (SCR) is one of the most often used peripheral physiological markers; it cannot be induced voluntarily and is exclusively activated by the sympathetic nervous system, making it an ideal measurement for arousal (Dawson, Schell, & Filion, 2007). Another promising maker for detecting unconscious or subtle effects during stimulus processing is pupil size (Laeng et al., 2012). Pupil size is triggered both by the sympathetic and the parasympathetic nervous system and is voluntarily affected only under effort (Sirois & Brisson, 2014). Changes of pupil diameter can be caused by variation on luminance, but also by emotional arousal (Bradley, Miccoli, Escrig, & Lang, 2008; Hess & Polt, 1960) and cognitive load (Kuchinke, Schneider, Kotz, & Jacobs, 2011; Laeng et al., 2012; Stanners, Coulter, Sweet, & Murphy, 1979). Stronger attention towards stimuli, as predicted in this experiment when participants were confronted with similar (fictive) speakers, leads for example to pupil dilation (see Sirois & Brisson, 2014). The interplay of the listeners' emotion recognition, SCRs and pupil size provides insight into the cognitive and affective reaction during processing of

the prosodic information. We hypothesized that listeners showed stronger physiological responses towards vocal expression by similar characters compared to dissimilar characters.

Responses of autonomic reactions (SCRs and pupil size) to arousing stimuli have been established for affective pictures and sounds rather than for emotional expressions (Bradley, Codispoti, Cuthbert, & Lang, 2001; Bradley & Lang, 2000). Affective pictures or sounds of high arousal, mainly representing violence and erotica, have been shown to induce subjective feelings, emotion-congruent facial expressions, increased SCRs, and pupil dilation in the beholder (Bradley, Codispoti, Cuthbert, et al., 2001; Bradley et al., 2008; Lithari et al., 2010; Partala & Surakka, 2003). The evidence for reactions to emotional expressions, is however less clear (Alpers et al., 2011; Aue et al., 2011; Wangelin et al., 2012).

To relate our results on the processing of emotional expressions with the processing of acoustic affective stimuli, we added a second task, in which positive, negative and neutral affective sounds were presented and had to be categorized regarding their emotional meaning by participants. We predicted that listeners reacted with increased autonomic responses towards arousing sounds.

Methods

Participants. Twenty-eight female German native speakers, ranging in age between 18 and 29 years (*M* = 22.8), participated in the study. The majority of participants (23 out of 28) were undergraduates at the University of Goettingen, three just finished their studies and two worked in a non-academic profession. Due to technical problems during recordings, two participants had to be excluded from the analysis of pupil data.

Stimuli.

Spoken utterances with emotional prosody. The emotional voice samples were selected from the Berlin Database of Emotional Speech (EmoDB, Burkhardt et al., 2005). The database

consists of 500 acted emotional speech tokens of ten different sentences. These sentences were of neutral meaning, such as “The cloth is lying on the fridge” [German original “Der Lappen liegt auf dem Eisschrank”], or “Tonight I could tell him” [“Heute abend könnte ich es ihm sagen”]. From this database we selected 30 anger, 30 joyful, and 30 neutral utterances, spoken by five female actors. Each speaker provided 18 stimuli to the final set (6 per emotion category). The stimuli had a mean duration of 2.48 ± 0.71 s (anger= 2.61 ± 0.7, joy= 2.51 ± 0.71, and neutral = 2.32 ± 0.71), with no differences between the emotion categories (Kruskal-Wallis chi-squared = 2.893, *df* = 2, *p* = .24). Information about the recognition of indented emotion and perceived naturalness were provided by Burkhard and colleagues (2005). We only chose stimuli that were recognized well above chance and perceived as convincing and natural (Burkhardt et al., 2005). Recognition rates did not differ between emotion categories (See Table 4.2 for descriptive statistics, Kruskal-Wallis chi-squared = 5.0771, *df* = 2, p = .079). Anger stimuli were however perceived as more convincing than joyful stimuli (Kruskal-Wallis chi-squared = 11.1963, *df* = 2, *p* = .004; post-hoc test with Bonferroni adjustment for anger – joy *p* = .003). During the experiment, we presented prosodic stimuli preceded by short context sentences that were presented in written form on the computer screen. With this manipulation we aimed at providing context information in order to increase the plausibility of the speech tokens. These context sentences were semantically related to the prosodic target sentence and neutral in their wording, such as “She points into the kitchen and says” [German original: “Sie deutet in die Küche und sagt”] followed by the speech token “The cloth is laying on the fridge” [“Der Lappen liegt auf dem Eisschrank”] or “She looks at her watch and says” [German original “Sie blickt auf die Uhr und sagt”] followed by the speech “It will happen in seven hours” [“In sieben Stunden wird es soweit sein”].

Table 4.2
Descriptive statistics of the stimulus material.

Prosody[a]			*Sound*[b]		
	Recognition	Naturalness		*Pleasantness*	*Arousal*
Anger	96.17 ± 7.39	84.55 ± 10.61	*Negative arousing*	*2.8 ± 1.76*	*6.9 ± 1.86*
Neutral	93.52 ± 6.46	80.08 ± 11.16	*Neutral*	*4.91 ± 1.75*	*4.46 ± 2.04*
Joy	93.19 ± 8.85	72.51 ± 15.98	*Positive arousing*	*7.23 ± 1.78*	*6.75 ± 1.81*

Note: [a] Burkhardt et al., 2005;[b] Bradley & Lang, 2007

Similarity manipulation was done equivalently to Study 1, resulting in four personal profiles of (fictive) characters that resembled the respective participant in her data, and four profiles that differed from the participant's profile. To detract participants from the study aim, we included trait memory tasks between acquiring the biographical information and the main experiment. Additionally, we instructed the participants to carefully read every profile that was presented during the experiment, as they later should respond to questions regarding bibliographic information.

Sounds of emotional content. Forty-five affective sounds (15 arousing positive, 15 arousing negative, 15 neutral[4]) were selected from the IADS database (International Affective Digital Sounds, Bradley & Lang, 1999). All of them had a duration of 6 s. Erotica were not used in our study as they have been shown to be processed differently compared to other positive arousing stimuli (Partala & Surakka, 2003; van Lankveld & Smulders, 2008). The selected positive and negative stimuli did not differ in terms of arousal (see Table 4.2 for descriptive statistic; $t(27) = -0.743$, $p = .463$) and were significantly more arousing than the neutral stimuli ($t(25) = 12.84$, $p < .001$). In terms of emotional valence, positive and negative stimuli differed both from each other ($t(24) = 21.08$, $p < .001$) and from the neutral condition (positive-neutral $t(19) = 11.99$, $p < .001$, negative-neutral $t(25) = 15.15$, $p < .001$), according to the ratings provided in the IADS database. Positive and negative sounds were controlled for their absolute valence value

[4] Stimulus selection (taken from Bradley & Lang, 1999):
Positive sounds: 110, 311, 352, 353, 360, 363, 365, 367, 378, 415, 704, 717, 813, 815, 817
Negative sounds: 134, 261, 281, 282, 289, 380, 423, 501, 626, 699, 709, 711, 712, 719, 910
Neutral sounds: 130, 170, 246, 262, 322, 358, 376, 698, 700, 701, 720, 722, 723, 724, 728

from the neutral condition ($t(24) = 0.159$, $p = .875$). Note that this stimulus selection was based on ratings by female participants' ratings only, provided by Bradley and Lang (2007).

As the affective sounds were relative diverse in their content, we controlled for differences in specific acoustic parameters that might trigger startle reactions or aversion and thus influence the physiological indicators used in the present study in an unintended way. These parameters included intensity, intensity onset (comprising only the first 200 ms), intensity variability (intensity standard deviation), noisiness, harmonic-to-noise ratio (HNR), and energy distribution (frequency at which 50 % of energy distribution in the spectrum was reached). Intensity parameters were calculated using Praat (Boersma & Weenink, 2009), while noisiness, energy distribution and HNR were obtained by using LMA (Lautmusteranalyse developed by K. Hammerschmidt; Hammerschmidt & Jürgens, 2007; Schrader & Hammerschmidt, 1997). We calculated linear models in R to compare the parameter values across the three emotion category (Table 4.3). We conducted post-hoc analysis even when the general analysis was only significant at trend level. We found differences at trend level for intensity and intensity variability, and significant effects for energy distribution across the emotions. Differences were marginal and unsystematically spread across the categories, meaning that no emotion category accumulates all aversion related characteristics. Differences depict the normal variation when looking at complex sounds. The probability that acoustic structure confounds the physiological measure is thus low.

Table 4.3
Acoustic parameter values for the emotional sounds grouped by emotion.

Parameter	negative	neutral	positive
Intensity [db]	65.66 ± 13.07	64.57 ± 9.33 *a*	71.50 ± 6.62 *a*
Intensity onset [db]	61.28 ± 18.64	62.44 ± 9.04	68.36 ± 7.20
Intensity variability [db]	12.66 ± 7.66 *a*	8.48 ± 4.07	7.83 ± 7.64 *a*
% noise	68.07 ± 34.13	80.33 ± 24.94	58.73 ± 33.03
harmonic-to-noise ratio	390.50 ± 469.72	369.80 ± 235.05	665.09 ± 638.37
50% Energy distribution [Hz]	1802.40 ± 961.37 *b*	1046.67 ± 868.05 *b*	1231.67 ± 389.62

Note: Differences in one parameter across emotion are indicated by uncapitalized letters (*a*: $p < .1$, *b*: $p < .05$)

Procedure. First, participants filled out questionnaires regarding their demography and their handedness (Oldfield, 1971). After completing the questionnaires, participants were asked to wash their hands and to remove eye make-up. Participants were then seated in a chin rest 72 cm in front of a computer screen. Peripheral physiological measures were recorded from their non-dominant hand, while their dominant hand was free to use a button box for responding. Stimuli were presented via headphones (Sennheiser, HD 449) at a volume of around 55 db. During and shortly after auditory presentation, participants were instructed to fixate a green circle displayed at the center of a screen in order to prevent excessive eye movements. The circle spanned a visual angle of 2.4° x 2.7° and was displayed on an equi-luminant grey background. Additionally, participants were asked not to move and to avoid blinks during the presentation of target sentences.

The experiment consisted of two parts. Figure 4.2 gives an overview about the procedure of the stimulus presentation. Within the first part, prosodic stimuli were presented. Stimuli were presented twice (once in the similar / once in the dissimilar condition), resulting in a total number of 180 stimuli. The stimulus set was divided into 20 blocks of 9 stimuli (three stimuli per emotion category that is anger, neutral, joy). All Stimuli within one block were spoken by the same speaker and were presented in random order within a given block. Prior to every prosodic stimulus, a context sentence was presented for three seconds. The personal profile, which manipulated the similarity, was shown prior to each block for six seconds. Every second block was followed by a break. Rating was done six seconds after stimulus onset. Participants had to indicate the valence of each stimulus (positive, negative, neutral) by pressing one of three buttons. In order to avoid early moving and thus assuring reliable SCR measures, the rating options appeared not until six seconds after stimulus onset and valence-button assignment changed randomly for every trial. Participants were instructed to carefully read the personal profiles and to feel into the speaker and the situation respectively. This part lasted for about 40 min. At the end of this part, participants answered seven questions regarding bibliographic information of the fictive speakers.

After a short break, the second part started, in which the 45 affective sounds were presented. Every trial started with a fixation cross in the middle of the screen for 1 s. The sound was then replayed for 6 s each, while a circle was displayed on screen. When the sound finished, the response options (neutral, positive, negative) were provided on the screen. The order of the response options was random for every trial; thus, button order was not predictable. The 45 affective sounds were presented twice in two independent cycles, each time in randomized order. In analogy to the prosodic part, participants were instructed to listen carefully and to indicate the valence they associate most with the sounds without thinking about the sound's meaning. Short breaks were included after every 15th trial. This part of the experiment lasted for about 20 minutes. The experiment took approximately 60 min in total.

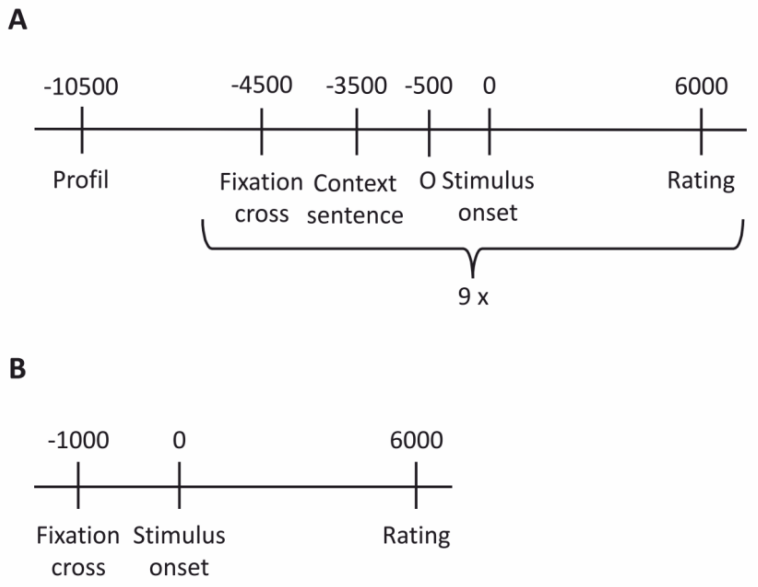

Figure 4.2 *Overview of stimulus presentation procedure. A) One of the 20 presentation blocks created for the prosodic stimuli. All nine stimuli of one block were spoken by the same speaker, and included in randomized order three neutral, three anger and three joy sentences; B) Stimulus presentation of sounds.*

Psychophysiological data recording, pre-processing, and analysis.

Pupil diameter was recorded from the dominant eye using the EyeLink 1000 (SR Research Ltd.), at a sampling rate of 250 Hz. The head position was stabilized via a chin and forehead rest that was secured on the table. Prior to the experiment, the eyetracker was calibrated with a 5-point calibration, ensuring correct tracking of the participant's pupil. Offline, blinks and artefacts

were corrected using spline interpolation. Data was then segmented around stimulus onset (time window: -1000 ms to 7000 ms) and referred to a baseline 500 ms prior to stimulus onset. Data was analyzed in consecutive time segments of 1 s duration each. We started the analysis 500 ms after stimulus onset, to allow a short orientation phase, and ended 5500 ms afterwards.

Skin conductance was recorded at a sampling rate of 128 Hz using ActivView and the BioSemi AD-Box Two (BioSemi B.V.). The two Ag/AgCl electrodes were filled with skin conductance electrode paste (TD-246 MedCaT supplies) and were placed on the palm of the non-dominant hand approximately 2 cm apart, while two additional electrodes on the back of the hand served as reference. Offline, data was analyzed using the matlab based software LedaLab V3.4.5 (Benedek & Kaernbach, 2010a). Data was down-sampled to 16 Hz and analyzed via Continuous Decomposition Analysis (Benedek & Kaernbach, 2010a). Skin conductance (SC) is a slow reacting measure based on the alterations of electrical properties of skin after sweat secretion. SC has long recovery times leading to overlapping peaks in the SC signal when skin conductance responses (SCR) are elicited in quick succession. Conducting standard peak amplitude measures is thus problematic, as peaks are difficult to differentiate and subsequent peaks are often underestimated. Benedek and Kaernbach (2010a) developed a method that separates the underlying driver information, reflecting the sudomotor nerve activity (and thus the actual sympathetic activity) from the curve of physical response behavior (sweat secretion causing slow changes in skin conductivity) via standard deconvolution. Additionally, tonic and phasic SC components are separated, to allow a focus on the phasic, event-related activity only. The phasic driver subtracted by the tonic driver is characterized by a baseline of zero. Event-related activation was exported for a response window of 1 to 6 sec after stimulus onset, taking into account the slow signal (Benedek & Kaernbach, 2010b). Only activation stronger than 0.01 µS was regarded as an event-related response (Bach, Flandin, Friston, & Dolan, 2009; Benedek & Kaernbach, 2010a). We used averaged phasic driver within the respective time window as measure for skin conductance response (SCR).

Statistical analysis. To test the effects of emotion category and similarity on recognition accuracy we built a generalized linear mixed model with binomial error structure (GLMM, lmer function, R package lme4, Bates et al., 2011). Effects on SCRs and pupil size were analyzed using linear mixed models (LMM, lmer function). Models included emotion category, similarity, and the interaction between these two as fixed factors and participant-ID as random factor. All models were compared to the respective null model including the random effects only by likelihood ratio tests (function anova). Additionally, we tested the interaction between emotion category and similarity by comparing the full model including the interaction with the reduced model excluding the interaction. We used the model without interaction when appropriate. Models for the affective sounds included only emotion category as fixed factor and participant-ID as random effect. The models were compared to the respective null models by likelihood ratio tests. Normal distribution and homogeneity of variance for all models were tested by inspecting Quartile-Quartile-Plots (QQ-plots) and residual plots. SCR data deviated from normal distribution and were log transformed. Pairwise post-hoc tests were conducted using the glht function of the multcomp package (Hothorn et al., 2008) with Bonferroni correction.

Results

Emotion recognition.

Spoken utterances with emotional prosody. Overall, emotional prosody was recognized relatively well, at around 92% (Figure 4.3). The comparison to the null model established an overall effect of the predictors on emotion recognition (χ^2 = 138.44, *df* = 5, $p < .001$), while the interaction between similarity and emotion category was not significant (χ^2 = 4.53, *df* = 2, $p = .104$). Similarity influenced the emotion recognition at trend level (χ^2 = 3.21, *df* = 1, $p = .073$, Figure 4.3), while emotion category had a strong influence on recognition (χ^2 = 130.81, *df* = 2, $p < .001$). Post-hoc tests revealed differences in every pairwise comparison (anger – joy: $z = 6.117$,

$p < .001$; anger – neutral: $z = 10.176$, $p < .001$; joy = neutral $z = 5.088$, $p < .001$). Anger was recognized most frequently, followed by joyful and neutral prosody (Figure 4.3).

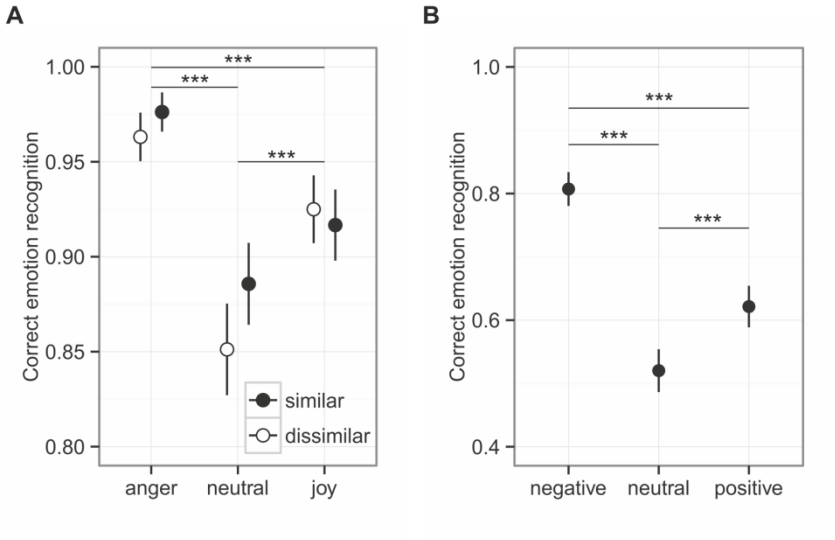

Figure 4.3 *Emotion recognition for prosody (A) and sounds (B). Given are the mean values ± 95% CI. Asterisks mark the significance level: * p < .05, ** p < .01, *** p < .001.*

Sounds of emotional content. The emotional content of sounds was recognized to a lesser degree than the emotionality of prosody, with a recognition accuracy of around 65% (Figure 4.3). Emotion had a clear influence on the recognition (comparison of full model and null model: $\chi^2 = 167.52$, $df = 2$, $p < .001$). With a recognition accuracy of about 52%, neutral sounds were recognized least accurately (negative – neutral: $z = 12.972$, $p < .001$; negative – positive: $z = 8.397$, $p < .001$; positive - neutral: 4.575, $p < .001$).

Skin conductance.

Spoken utterances with emotional prosody. Skin conductance response (Figure 4.4) represented by the phasic driver activity was not affected by any of the predictors (comparison to null model $\chi^2 = 1.605$, $df = 5$, $p = .9$).

Sounds of emotional content. We did, however, find an effect of the emotional sounds on SCR (Figure 4.4; comparison to null model $\chi^2 = 15.828$, $df = 2$, $p < .001$). Consistent with the prediction, more arousing sounds elicited stronger SCRs than neutral sounds (Figure 4.4, post-hoc tests: negative – neutral: estimates on log transformed data = 0.336 ± 0.081, $z = 4.129$, $p < .001$;

positive – neutral: estimates on log data = 0.222 ± 0.081, z = 2.723, p = .019). Negative and positive sound elicited SCRs of similar size (negative – positive: estimates on log data = 0.114 ± 0.0814, z = 1.406, p = .479).

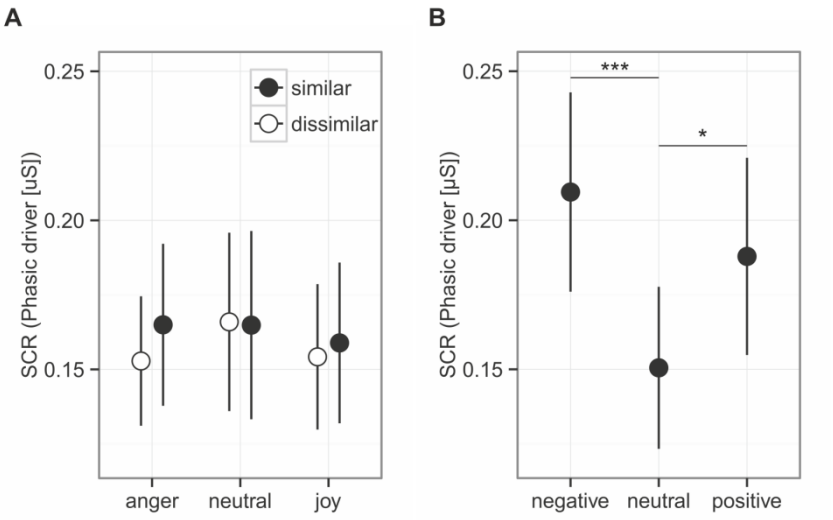

Figure 4.4 *Skin conductance response for the prosodic stimuli (A) and the sounds (B). Given is the mean ± SEM phasic driver activity within the response window of 1-6 sec after stimulus onset. Asterisks mark the significance levels of the post- hoc tests * < .05, ** < .01, *** < .001.*

Pupil dilation.

Spoken utterances with emotional prosody. We found an effect of the predictors on pupil size for the time windows 2.5 – 3.5 and 3.5 - 4.5 seconds after stimulus onset (comparisons to null models, see Table 4.4). There was no interaction between emotion category and similarity on pupil size (Table 4.4). Pupil size was affected by emotion category of speech samples (Figure 4.5, Table 4.4). Interestingly, increases of pupil size dynamically differed between prosodic conditions: Pupil size increased fast in response to angry stimuli, while responses to joyful stimuli were delayed by about one second (see Figure 4.5 and Table 4.5). Neutral stimuli triggered the weakest pupil response in comparison to anger and joy (Figure 4.5, Table 4.5). The similarity condition had no effect on pupil size for the respective time windows (model comparisons $\chi^2 < 1.16$, $df = 1$, $p > .28$).

Sounds of emotional content. The pupil size was affected by emotional content of sounds in three time windows (Table 4.4, Figure 4.5). Post-hoc tests revealed that negative sounds elicited a stronger pupil size response compared to positive sounds (Table 4.5). Differences

between negative and neutral sounds almost reached significance. Our results indicate that pupil dilation does not purely reflect arousal differences.

Table 4.4

Effects on pupil sizes for prosody and sounds. Presented are the results of the model comparison for each time segments.

		null model comparison[a]			interaction			emotion		
Stimulus	time steps	χ^2	*df*	*p*	χ^2	*df*	*p*	χ^2	*df*	*p*
Prosody	0.5-1.5	9.86	5	.079.		2				
	1.5-2.5	5.56	5	.351						
	2.5-3.5	11.56	5	.041*	0.786	2	.675	10.66	2	.005**
	3.5-4.5	11.82	5	.037*	1.63	2	.444	9.108	2	.011*
	4.5-5.5	7.722	5	.172						
Sounds	0.5-1.5	1.74	2	.419						
	1.5-2.5	4.29	2	.117						
	2.5-3.5	6.95	2	.031*						
	3.5-4.5	12.81	2	.002**						
	4.5-5.5	6.07	2	.048*						

Note: [a] For the sounds, the emotion effect is reflected by the null model comparison

Table 4.5

Emotion effects on pupil size for prosody and sounds. Presented are post-hoc tests for the respective time segments

Stimulus	time step	Emotion		Estimates	*z*-value	p^a
Prosody	2.5-3.5	joy	neutral	0.025 ± 0.0143	1.72	.258
		anger	joy	0.023 ± 0.0143	1.58	.345
		anger	neutral	0.047 ± 0.143	3.29	.003**
	3.5-4.5	joy	neutral	0.041 ± 0.017	2.58	.03*
		anger	joy	0.001 ± 0.017	0.1	1
		anger	neutral	0.047 ± 0.017	2.67	.023*
Sounds	2.5-3.5	positive	neutral	0.036 ± 0.025	1.43	.458
		negative	positive	0.067 ± 0.025	2.67	.023*
		negative	neutral	0.031 ± 0.025	1.24	.646
	3.5-4.5	positive	neutral	0.032 ± 0.025	1.29	.592
		negative	positive	0.092 ± 0.025	3.68	<.001***
		negative	neutral	0.06 ± 0.25	2.39	.051.
	4.5-5.5	positive	neutral	0.001 ± 0.024	0.06	1
		negative	positive	0.053 ± 0.024	2.18	.087.
		negative	neutral	0.052 ± 0.024	2.12	.102

Note: [a] *p-values base on Bonferroni correction within one time window*

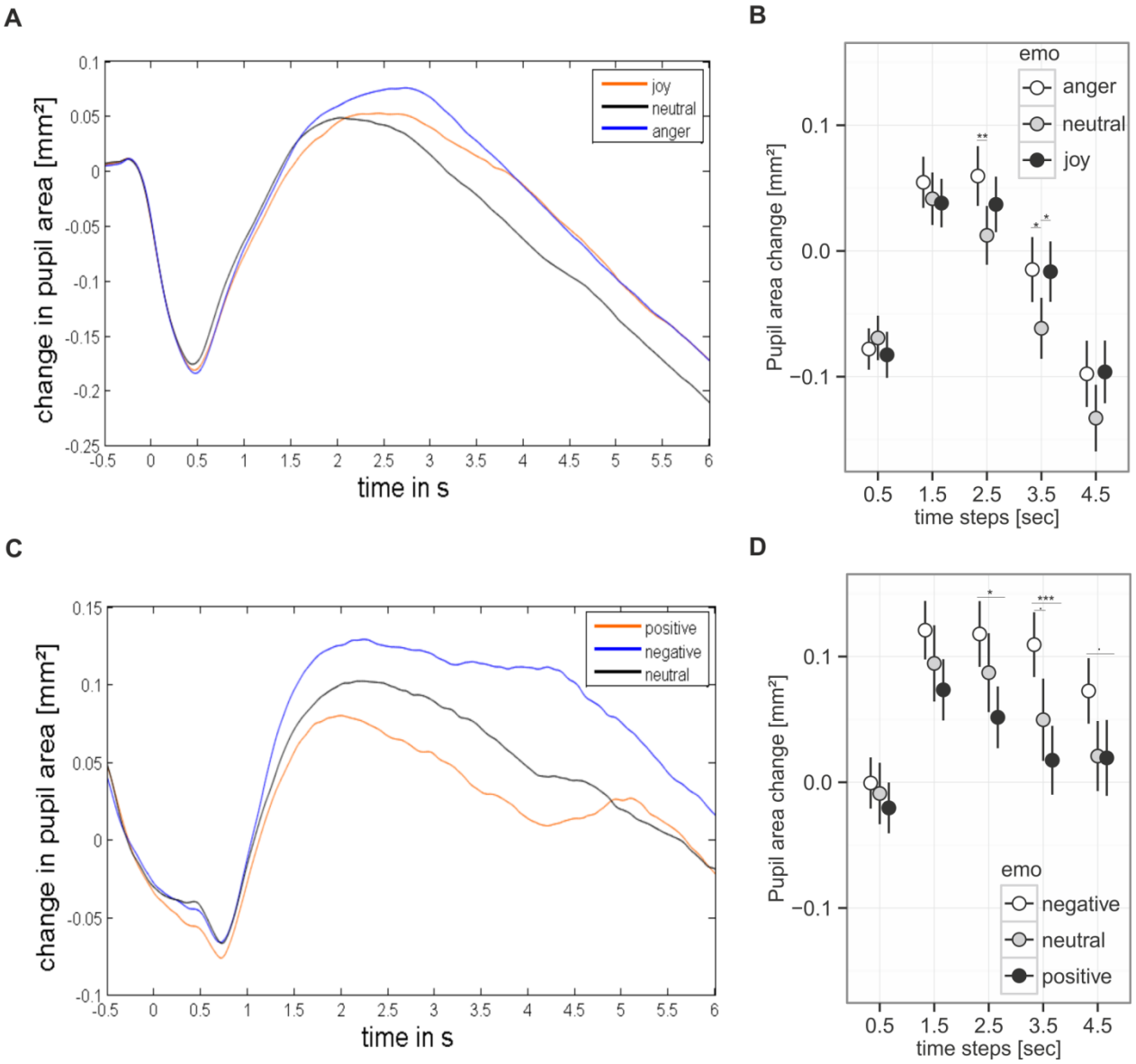

Figure 4.5 *Pupil dilation during presentation of prosodic stimuli (A, B) and sounds (C, D). Figure parts B, D base on mean values ± SEM for the analyzed time steps. Stimulus onset happened at time point 0. Asterisks mark the significance levels of the post-hoc tests: . p < .1, * p < .05, ** p < .01, *** p < .001.*

Discussion

In accordance with Experiment 1, manipulated similarity between (fictive) speaker and listener did not influence the affective processing of vocal prosody. Similarity failed to draw more attention to the vocal expressions – as would have been indicated by modulations of the pupil size and improved recognition performance – or to boost physiological responses in terms of increased arousal. Considering previous findings, this result seems surprising (Burger et al., 2004; Cwir et al., 2011; Walton et al., 2012). Nevertheless, our study does imply that similarity has no effect on vocal emotion processing, but that our manipulation using biographical similarity in a laboratory setting does not seem to be sufficiently potent to induce effects (see General discussion).

Interestingly, the peripheral physiological data indicate that listening to vocal expressions of emotions did not trigger a marked activation of the sympathetic nervous system, as indicated by the null-findings on SCRs in the present experiment. These results are in accordance with previous studies on facial expressions of emotions (Alpers et al., 2011; Wangelin et al., 2012) that also failed to show increased SCRs to emotional expressions. However, in contrast to the spoken utterances, affective sounds elicited the hypothesized arousal effects in SCRs, confirming previous results (Bradley & Lang, 2000). Together, the SCR findings imply that emotional prosody is of generally lower arousal in comparison to affective sounds, which is supported by overall decreased SCR magnitudes to prosodic stimuli compared to sounds. However, emotional prosody differentially affected pupil size with larger dilations for utterances spoken with angry or joyful prosody, which is in line with a recent study reported by Kuchinke and colleagues (2011). Since pupil responses have been demonstrated to reflect the dynamic interplay of emotion and cognition and are thus not simply related to arousal (cf. Bayer, Sommer, & Schacht, 2011), our finding is not surprising. Instead, it provides additional evidence that SCRs and pupil responses reflect functionally different emotion-related ANS activity.

Another finding supports the idea that pupil dilation reflects cognitive effects on emotional processing. In a study by Partala and Surakka (2003), affective sounds, taken from the same data base as in our study, triggered stronger pupil dilations for both negative and positive compared to neutral sounds. In our study, however, emotionally negative sounds elicited larger dilations compared to positive sounds, with neutral in between. While participants in our study had to overtly judge the emotional content of sounds, no explicit task had to be performed in the study by Partala and Surakka (2003). These processing differences due to task demands might explain the inconsistencies between studies. Similar arguments have been made by Stanners and colleagues (1979) suggesting that changes in pupil size only reflect arousal differences under conditions in which no cognitive effort is required. For both domains – affective sounds and prosodic utterances – participants had to explicitly categorize the emotional content or prosody of each stimulus. Since accuracy rates provide rather unspecific estimates of cognitive effort, we conducted another experiment in which we collected reaction times and confidence ratings in addition to accuracy rates. This allowed us to further investigate the cognitive load during recognition of both prosodic stimuli and affective sounds. To be able to apply the results with the measures of Study 2, we used a subset of the participants form Study 2.

Experiment 3

In Experiment 3 we explored the cognitive difficulties for recognizing vocal expressions and affective sounds in more detail. In light of the previous recognition rates in Experiment 2 and the respective pupil data, we predicted that neutral prosody and especially neutral sounds would be most difficult to recognize.

Methods

Participants. The sample consisted of 20 (female, aged 21-30 years, *M* = 24.45) participants from the 28 participants of study 2.

Stimuli. Stimuli were identical to Experiment 2 but both the context sentences and the similarity manipulation were omitted.

Design and procedure. The experiment took place 6 month after Experiment 2. Participants sat in front of a computer screen, and listened to the acoustic stimuli via headphones. They were first confronted with the affective sounds (first part) in a randomized order and were instructed to stop the stimulus directly as fast as they had recognized the emotion within a critical time window of 6 seconds. The time window was in accordance to the one in Experiment 2 and corresponded to the duration of the sounds. After participants pushed a button, reflecting the time needed for successful emotion recognition, they had to indicate which emotion they perceived (positive, negative, neutral) and how confident they were in their recognition (likert-scale 1-10), both by paper-pencil. The next trial started after a button press. In the second part, they listened to the prosodic stimuli that had to be classified as expressing joy, anger, or neutral, respectively, within the same procedure as in the first part. The critical time window was again 6 s after stimulus onset.

Statistical analysis. We did not compare prosody and sounds statistically, knowing about the differences in stimulus length, quantity of stimuli and, regarding the broader perspective, the stimulus structure overall (Bayer & Schacht, 2014). Reaction time data was not normally distributed and was thus log transformed prior to the analysis. Recognition accuracy and reaction time data were only calculated for those stimuli that were responded to within the time window of 6 s, whereas certainty ratings were analysed for all stimuli in order to not overestimate the ratings. We tested the effect of emotion category on recognition accuracy (using GLMM), reaction time (using a LMM), and certainty ratings (using a cumulative link mixed model for ordinal data, package ordinal, Christensen, 2012) for both prosodic stimuli and affective sounds. The models included emotion category as fixed factor and participant-ID as random effect. The models were compared to the respective null models by likelihood ratio tests. Pairwise post-hoc tests were conducted using the glht function with Bonferroni correction for recognition accuracy and

reaction time. As cumulative link models cannot be used in the glht post-hoc tests, we used the single comparisons of the model summary, and did the Bonferroni correction separately.

Results

Spoken utterances with emotional prosody. Participants responded within the specified time window in 90% of all cases (anger: 91%, neutral: 90%, joy: 89%). We calculated recognition accuracy and reaction time only for these trials. Emotion category had a clear influence on emotion recognition accuracy (comparison to null model $\chi^2 = 26.39$, $df = 2$, $p < .001$), reaction time ($\chi^2 = 42.29$, $df = 2$, $p < .001$), and the certainty ratings (*LR.stat* = 21.50, $df = 2$, $p < .001$). With a recognition rate of 91 %, joy was recognized significantly less accurately, and more slowly (Reaction time: M = 2022 ms, Figure 4.6 and Table 4.6). In the certainty ratings, however, joy resembled anger expressions.

Sounds of emotional content. Participants responded within the specified time window in 81% of all cases (negative: 85%, neutral: 74%, positive: 83%). The recognition varied between emotion categories for recognition accuracy (comparison to null model $\chi^2 = 41.41$, $df = 2$, $p < .001$), reaction time ($\chi^2 = 19.97$, $df = 2$, $p < .001$), and certainty ratings (*LR.stat* = 26.39, $df = 2$, $p < .001$). These results indicate difficulties in the categorization of neutral sounds, as reflected in lower accuracy (57% correct), prolonged reaction times (2988ms), and lower certainty ratings (Figure 4.6 and Table 4.6).

Table 4.6

Post-hoc comparisons for emotion recognition, reaction time and confidence ratings for emotional prosody and affective sounds

Rating	Stimulus	Emotion		Estimates	*z*-value	p^a
Recognition accuracy	Prosody	neutral	anger	0.345 ± 0.379	0.909	1
		joy	anger	-1.110 ± 0.304	-3.646	<.001***
		joy	neutral	-1.455 ± 0.337	4.321	<.001***
	Sounds	neutral	negative	-1.327 ± 0.217	-6.109	<.001***
		positive	negative	-0.454 ± 0.224	-2.026	.128
		positive	neutral	0.873 ± 0.202	4.315	<.001***
Reaction time	Prosody	neutral	anger	-0.586 ± 0.023	-2.543	.033*
		joy	anger	0.092 ± 0.023	3.976	<.001***
		joy	neutral	0.151 ± 0.023	6.495	<.001***
	Sounds	neutral	negative	0.181 ± 0.403	4.494	<.001***
		positive	negative	0.083 ± 0.039	2.138	.098
		positive	neutral	-0.097 ± 0.040	-2.417	.047*
Confident ratings	Prosody	neutral	anger	0.332 ± 0.118	2.81	.015*
		joy	anger	-0.208 ± 0.116	-1.785	.223
		joy	neutral	-0.540 ± 0.117	-4.608	<.001***
	Sounds	neutral	negative	-0.829 ± 0.148	-5.596	<.001***
		positive	negative	-0.161 ± 0.152	-1.057	0.6
		positive	neutral	0.668 ± 0.151	5.413	<.001***

Note: [a]Adjusted p-values (Bonferroni correction)

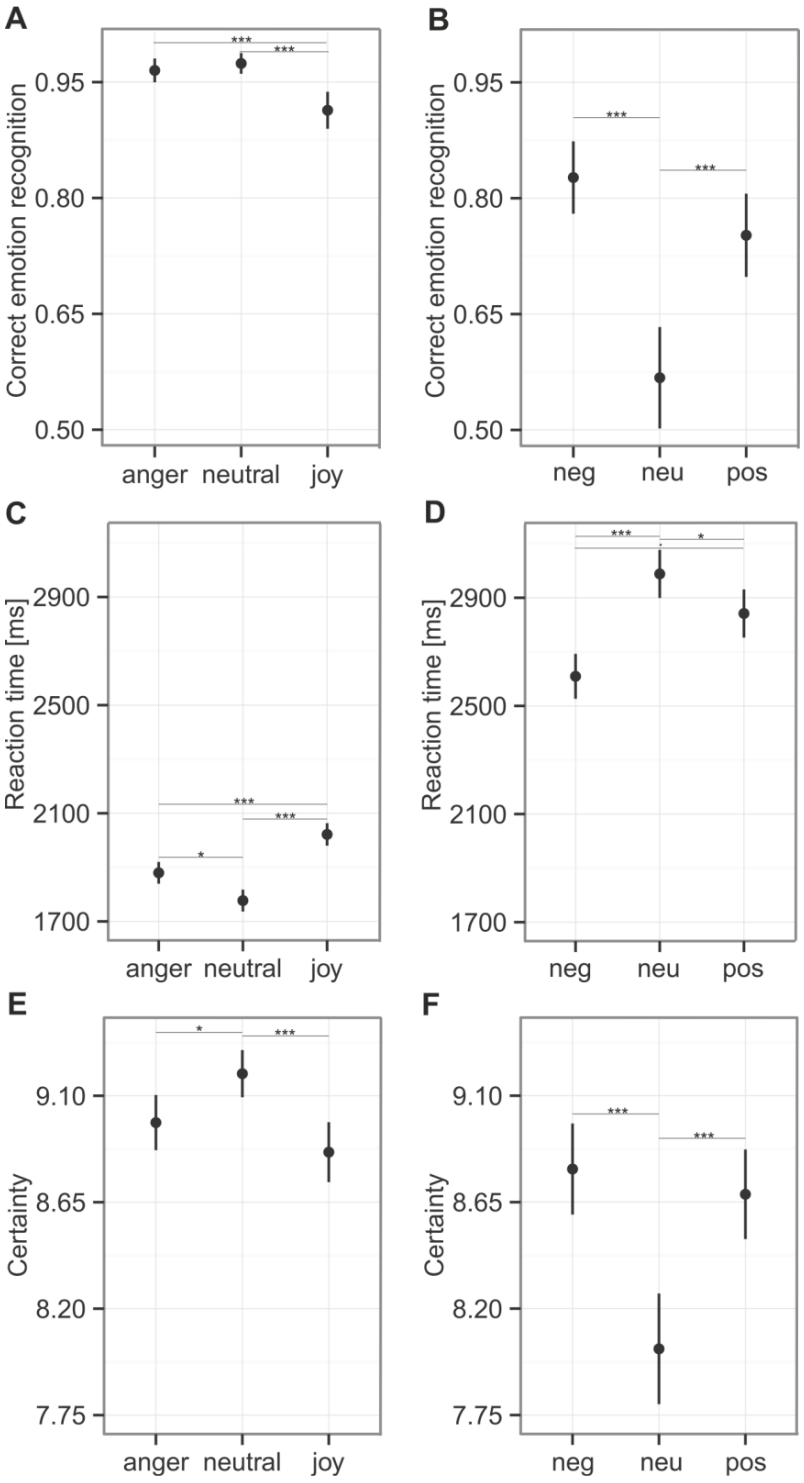

Figure 4.6 *Emotion recognition. The first column (A, C, E) depicts the results of the emotional prosody with the emotion categories "anger", "neutral" and "joy", the second column (B, D, F) represents the sounds with the categories "negative", "neutral" and "positive". A), B) Correct emotion recognition (mean ± 95% CI) was calculated using stimuli that were responded to within the time window of 6 seconds. C), D) Reaction time measures (mean ± SEM) on stimuli that were responded to with the critical time window. E, F) Certainty ratings (mean ± 95% CI) obtained from the 10 point likert-scale (ranging from 1 uncertain to 10 completely certain) were calculated for every stimulus. Asterisks mark the significance level: . p < 0.1, * p < .05, ** p < .01, *** p < .001.*

Discussion

Recognition of emotional prosody was quick and accurate, and participants were relatively confident about their decisions when the utterances were presented without any context information. Overall, the recognition of emotional content of affective sounds appeared to be more difficult as predicted. Importantly, we replicated the ambiguity participants experienced when being presented with neutral sounds that had to be categorized in terms of emotional meaning. Neutral prosody, however, was classified most accurately, which contrasts the behavioral data in Experiment 2 (in which the same stimuli were used) and Experiment 1. These contrasting findings indicate that embedding prosodic stimuli into context situations diminishes the subjective probability that they were perceived as neutral (see also Vrana & Gross, 2004).

The detailed analysis of participants' recognition ability in this experiment suggests that the recognition of vocally expressed emotions does not require large cognitive resources in general. In this case, the impact of emotion on pupil size might be basically caused by arousal (Stanners et al., 1979), even though the arousal level might not have reached a sufficient level to elicit SCRs. The temporal recognition pattern, with neutral and anger classified quickly and joy classified after a longer delay, fit to the timing data by Pell and Kotz (2011) and might also explain the delay in pupil dilation to joyful prosody in Experiment 2. The putative higher cognitive load caused by the emotion judgment task for sounds - and neutral sounds in particular - might explain the surprising pattern of pupil responses to sounds, obtained in Experiment 2, where neutral sounds elicited similar pupil dilations compared to negative sounds.

General Discussion

The present study aimed to investigate the influence of biographical similarity on the recognition of vocal expressions as well on emotional engagement during their processing. We did not find any effect of similarity manipulation. Autonomic reactions in response to vocal

expressions were generally weak, as only pupil size differentiated emotional and neutral stimuli, speaking against the notion that emotional engagement is necessary for the understanding of emotions in others.

Personal similarity between speaker and listener, which manipulates social relevance, did not improve the processing of prosody. Neither emotion recognition and perception nor the peripheral physiological measures indicated any influence. This might be explained by several, not mutually exclusive, reasons. First, vocal emotion expressions might be perceived in a general way, in which case emotions are recognized without differentiating between speakers' relevance. Whether an empathic reaction or prosocial behavior is triggered afterwards, may thus underlie further appraisal processes (see Cwir et al., 2011; de Vignemont & Singer, 2006). Considering emotion recognition for Experiment 2, a ceiling effect, however, might have simply prohibited further improvement of recognition accuracy. The missing effect of similarity on the autonomic measures indicates on the other hand that the processing of an expression is not affected by social connectedness overall. Second, the lack of effects might be explained by the fact that similarity unfolds its beneficial effect only in more realistic settings, in which an actual link between both interaction partners can be developed (see Burger et al., 2004; Cwir et al., 2011; Walton et al., 2012). Our manipulation might remain artificial and hence not affect the listeners in the intended way. Superficial manipulation of sharing group-membership, however, has been demonstrated to increase the ability to recognize emotions (Young & Hugenberg, 2010). Group membership is probably more relevant (Weisbuch & Ambady, 2008). Actually, recent findings indicated that group membership and similarity have different effects on human behavior (Mussweiler & Ockenfels, 2013). There might be a third explanation why personal similarity does not improve emotion processing in general. Todd, Hanko, Galinsky, and Mussweiler (2011) described that similarity leads to an egocentric perspective. People thus strongly include their own knowledge, norms, and beliefs into their situational appraisal (see also Mussweiler & Ockenfels, 2013). Following this line of thought, participants in the similar condition might

incorporate personal tendencies, moods, or perspectives into their emotion perception, leading to a more complex influence of similarity than simply boosting attention or emotional engagement. Further studies, using a more realistic approach and allowing for analysis of individual differences are therefore needed.

Regarding the assumed relevance of mirroring others emotions (Hatfield et al., 2011; Preston & de Waal, 2002), it is rather striking that we found only weak responses in the autonomic nervous system towards prosodic stimuli (for similar results concerning facial expressions see Alpers et al., 2011; Wangelin et al., 2012). Presumably, the emotional responses elicited by prosodic information were too subtle to be reflected in changes of electrodermal activity (cf. Levenson 2014). Emotional expressions are highly relevant social stimuli and might not function as emotion elicitors in socially irrelevant situation such as lab experiments (see Hess & Fischer, 2013). In our study, we aimed at improving the social relevance of speech tokens by embedding them into context and by providing biographical information about the fictive speaker. As mentioned above, this setting remained artificial and future research is needed to investigate whether more relevant social stimuli, such as direct second-person messages, avatars looking directly at the participants while speech tokens are presented, or utterances spoken by personally familiar people would increase physiological responsiveness to emotional prosody. However, it might also be possible that vocal expressions alone do not elicit sufficiently strong affective or empathic responses - in this context it is not surprising that emotion processing is not affected by our similarity manipulation. Vocal expressions in daily life are seldom expressed without the appropriate verbal content. Regenbogen and colleagues (2012) for example demonstrated that empathic concern is reduced when speech content is neutralized. Similarly, there is evidence that although emotional prosody can be recognized irrespectively of the actual semantic information of the utterance (Pell, Jaywant, Monetta, & Kotz, 2011), semantics seem to outweigh emotional prosodic information when presented simultaneously (Kotz & Paulmann, 2007; Wambacq & Jerger, 2004). Prosody is an important channel for emotional communication,

but semantics and context might be more important than the expression alone. Future studies on multimodal stimuli in more realistic situations are promising to enlighten the discussion about how and when we share the emotions of our social partner.

It is noteworthy that processing affective sounds presumably triggers much stronger affective and thus autonomous reactions, while the recognition of their emotional content is poor compared to prosodic stimuli. Autonomic reactions to acoustic stimuli can thus be evoked in laboratory settings (Bradley & Lang, 2000). The variation in affective processing of sounds and vocal expressions might be explained by overall differences between the two stimulus domains. Bayer and Schacht (2014) described two levels of differences between the domains which render a comparison almost impossible, namely physical and emotion-specific features. Firstly, at the physical level, affective sounds are more variable in their acoustic content than the vocal expressions, and hence more diverse, while emotional expressions vary only in subtle acoustic differences (Hammerschmidt & Jürgens, 2007). Secondly, there are strong differences in their overall emotion-specific features. While pictures and sounds have a direct emotional meaning (thus depicting the complete emotional situation), an emotional expression primarily depicts the expresser's emotional appraisal of a situation, rather than the situation itself. In contrast to stimuli with a direct emotional meaning, Walla and Panksepp (2013) described emotional expressions as possessing indirect meaning. Emotional sounds used in our study, such as baby cries, explosions, applause, and baby babble, require quick responses for which an explicit knowledge of the situation might be less important. For emotional expressions, in contrast, explicit knowledge may be vital. To behave adaptively, the context has to be analyzed as well, which might prevent immediate emotion responses (Barrett, Mesquita, & Gendron, 2011). Attending to emotional stimuli such as pictures or sounds seemingly evokes emotional responses in the beholder while attending to pure emotion expressions rather elicits recognition efforts (see Britton et al., 2006 for similar conclusion) than emotional engagement. As stated above,

emotional engagement or empathic reactions in form of autonomic responses, might follow emotion recognition after further evaluation, but are not involved in emotion recognition per se.

To sum up, our study revealed that listening to vocal expressions of emotions did not evoke strong emotion-congruent autonomic responses in the listener, and that neither emotion recognition nor emotional engagement was affected by biographical similarity between sender and listener. Our findings indicate that the recognition of vocal expressions does not depend on social connectedness or emotional engagement. The formation of empathic concern requires a more holistic situational appraisal than simply processing emotion expressions.

Acknowledgment

The authors thank Mareike Bayer for helpful comments on psychophysiological data recording and pre-processing, and Lena Riese, Sibylla Brouer, Katrin Riese, Anna Grimm, and Ramona Kopp for assisting in data collection.

Appendix

4.1 Example of a short story (Experiment 1)

Today is Monday and he wants to work calmly from home. He planned to start working directly after breakfast, but instead he first finishes his housework that he left undone during the week. He takes care of his dirty cloths, vacuums and tidies up. It is half past ten, when he finally turns on his PC and fetches his documents. He spends the next two hours in front of his notes. He works focused and only leaves the desk to get fresh coffee from out of the kitchen. Suddenly, the doorbell rings. As nobody else is at home, he gets up, goes to the door and opens it. There is the postman, who hands him a certified letter and asks for a signature. He examines the addressor of the letter and the date of the postmark. He looks up and says:

“I have expected this one” [Target sentence]

German original:

Heute am Montag will er in Ruhe von zu Hause aus arbeiten. Doch anstatt sich sofort nach dem Frühstück an den Schreibtisch zu setzen, wie er sich das vorgenommen hatte, erledigt er zunächst den Haushalt, der die Woche über liegen geblieben ist, Er kümmert sich um seine dreckige Wäsche, saugt Staub und räumt ein wenig auf. Es ist halb elf als er schließlich seinen Rechner hochfährt und seine Unterlagen hervorholt. Die nächsten zwei Stunden verbringt er vor seinen Aufzeichnungen. Er arbeitet konzentriert und verlässt den Schreibtisch nur um sich frischen Kaffee aus der Küche zu holen. Plötzlich klingelt es an der Tür. Da niemand sonst zu Hause ist, steht er auf geht zur Tür und öffnet sie. Es ist der Postbote, der ihm ein Einschreiben in die Hand drückt und nach einer Unterschrift verlangt. Er betrachtet den Absender des Briefes und das Datum des Poststempels. Dann schaut er auf und sagt:

„Den hab ich erwartet.“

5 General Discussion

In this thesis I investigated the impact of vocal expressions on understanding emotional states in other people. I found that vocal expressions of emotions could be convincingly play-acted, both by professional actors (Chapter 2 and 3), and by acting inexperienced people (Chapter 3). However, there was a notable interaction between emotion category and encoding condition for sadness and anger, which was consistent across cultures. The cross-cultural comparison (Chapter 2) revealed a comparable emotion recognition pattern for German, Romanian, and Indonesian participants with a slight in-group-effect for German listeners, which was independent of the encoding condition. The recognition of acted expressions was not more strongly affected by culture than the recognition of spontaneous expressions. Cultural differences became mainly apparent in decoding biases. Recognition rates for acting-inexperienced people closely resembled the rates for spontaneous expressions and only showed lower values for sadness (Chapter 3). Notably, emotion portrayals by professional actors were more consistently recognized as being play-acted than the ones by inexperienced people, which contradicted the hypothesis that actors were especially suited to produce credible emotional expressions. In Chapter 4, I demonstrated that biographical similarity that was experimentally manipulated between a fictitious speaker and the listener in a rating study did not improve emotion recognition, in contrast to the prediction. Additionally, listening attentively to the vocal emotion expressions did not trigger skin conductance responses (SCR), but affected pupil size. None of the autonomic reactions was affected by similarity manipulation. Listening to sounds with emotional content, in contrast, evoked arousal-related SCRs in the participants, showing that the methodological approach chosen was suitable to detect such effects when occurring. In the following chapters I will discuss these findings in a broader context.

5.1 Relation between play-acted and spontaneous expressions

"All the world's a stage"

(Shakespeare, from *As you like it*, Act II Scene VII)

5.1.1 Authenticity - The complete picture

Can we infer the inner affective states of our social partners by listening to their voices? It is assumed that this is the case (i.e., Mier et al., 2010), considering that emotional expressions rely on peripheral physiological reactions during emotional episodes (i.e., Moors et al., 2013; Mulligan & Scherer, 2012; K. R. Scherer, 1986). Although emotional expressions can be masked, feigned, suppressed, or intensified due to cultural display rules or social requirements (U. Scherer et al., 1980), considering their predictive value, "true" expressions should be distinguishable from deliberately produced ones. My study revealed that listeners could not consistently distinguish whether the emotional expressions were spontaneous or play-acted. Chapter 2 and Chapter 3 of this thesis were part of a larger set of studies on the comparison of spontaneous and play-acted expressions. In the following section I summarize the complete findings so far.

Play-acted emotion expressions differed acoustically from spontaneous ones, namely by their less monotonous speech and their voice quality that mark articulation differences, but these acoustic variations did not affect emotion encoding specifically (Jürgens et al., 2011). In Chapter 3, I showed that speech melody consistently depicts acted expressions, but that voice quality does not, leaving merely one acoustic parameter for differentiation (cf. Williams & Stevens, 1972). From the receivers' point of view, whether speech tokens were authentic or play-acted was recognized infrequently, with authentic expressions recognized correctly more often (an effect called veracity effect or truthful bias, see Ekman, 1996; Levine et al., 1999). Response behavior during emotion recognition, however, varied depending on acting and this effect was stable across different cultures. Anger was more accurately recognized when professionally acted, while sadness was recognized less accurately when acted either by actors or by inexperienced speakers (see Chapter 2 and 3, Drolet et al., 2012). This effect on emotion recognition seems to be based

on stimulus-inherent features, as participants primed with “acted” and “authentic” cues prior to stimulus presentation, did not show any behavioral differences (Drolet et al., 2013). German participants were in general more accurately in their recognition and cross-cultural difference mainly became apparent by attributional biases, such as a disposition against judging authentic stimuli as anger in Romanian and Indonesian listeners. While the explicit question whether a recording was play-acted or not could only poorly be answered by listeners, Drolet et al. (2012) found that listening to authentic stimuli increased neural activity of the Theory of Mind network (U. Frith & Frith, 2003). Thus both stimuli groups were perceived differently, and this difference was linked to mentalizing. Further analyses showed that the activation of the medial prefrontal cortex (mPFC, belonging to the ToM network) during an explicit authenticity discrimination task was related to the variability of fundamental frequency, the acoustic parameter discriminating play-acted and spontaneous emotional speech (Drolet, Schubotz, & Fischer, 2014). Summing up, our studies indicate that vocal emotional expressions can be play-acted convincingly, even more from people that did not received acting training. Yet, specific effects on emotion recognition across encoding conditions were found. One parameter, the more variable speech melody, characterizes acted expressions, which might affect the mentalizing process.

5.1.2 Emotion-specific recognition of vocal expressions

The presented data on emotion recognition demonstrated that anger, fear, joy, and sadness were not equally well recognized and that acting influenced the recognition of emotions specifically. Emotion-specific differences in expression and recognition have been studied for disgust, an expression that is clearly depicted in the face but not in the voice (T. Johnstone & Scherer, 2000; K. R. Scherer, 2003). This observation was explained by the function of disgust, which is to warn people in proximity not to eat rotten food. The ability to transmit disgust over longer distances was not necessary and did therefore not evolve in the acoustic domain (K. R. Scherer, 2003). In Chapter 2 and Chapter 3, I found that correct emotion classification was

generally low (see Chapter 3, and Jürgens et al., 2011 for weak emotion effects in the acoustic structure, indicating a low emotional content in general), which highlights that previous studies that used preselected stimuli overemphasized the informative value of expressions by the vocal domain (e.g., Scherer et al., 2001). Vocal expressions can be distinct and intense, but seemingly in daily life they are used more ambigous.

Participants were especially inconsistent in classifying fear expressions. Yet, recognition was above chance level (Chapter 3). Regarding its function to give warnings about an approaching threat, the ability to reach more distant individuals by using the vocal channel would be of advantage, (K. R. Scherer, 2003). In case of our stimulus set, the question arises why fear is not recognized strongly by the participants (see also K. R. Scherer, Banse, Wallbott, & Goldbeck, 1991), especially while other studies did find high recognition accuracies (Pell et al., 2009; K. R. Scherer et al., 2001). The label fear often comprises a high aroused version – panic - or alternatively a less aroused version – anxiety (cf. Banse & Scherer, 1996) and is thus not consistent (see also Russell, 2003). It seems plausible that panic is more frequently expressed in short exclamations such as single words, short phrases or by nonverbal affect bursts (such as screams, see Sauter et al., 2010) than by longer texts as used in our studies. The fear stimuli included in our stimulus set were recorded in social interactions. As fear is characterized by a lack of control and power (Ellsworth & Scherer, 2003), it seems plausible that people did not want to communicate this inner state to people other than their intimates. Considering that even in the acted conditions fear was inconsistently recognized, anxiety is seemingly not only suppressed in social situation, but is generally transmitted weakly via the vocal channel.

The recognition of joy was inconsistent in our studies and generally low for both conditions (i.e., Drolet et al., 2012, see also Pell et al., 2009; K. R. Scherer et al., 2001). Considering the use of social smiles and other situations in which humans express happiness, although they might feel different (e.i., Fernández-Dols & Ruiz-Belda, 1995), it is interesting that our findings result in such a low recognition of play-acted joy. I would assume that the face, with the smile as

clear expression pattern, is much more important when expressing, or play-acting joy than the voice.

Anger and sadness are of special interest, as they are consistently perceived differently in our studies when being play-acted or spontaneous respectively. The encoding condition was not more frequently recognized for anger and sadness compared to the other emotion categories (Chapter 3). The function of anger might explain the high recognition rates for this emotion (Banse & Scherer, 1996; K. R. Scherer, 2003), which includes threatening opponents as well as signaling dominance and power (Fischer & Manstead, 2008; K. R. Scherer, 2003). Being able to threat people from a distance as well as to recognize the threat in time is an advantage, promoting the evolution of anger vocal expressions (K. R. Scherer, 2003). Anger stimuli in general attract high attention (cf. Bayer & Schacht, 2014; see also the quick and distinct pupil increase towards angry speech in Chapter 4), possibly caused by the importance for the individual's well-being to recognize a possible threat. Display rules might reduce the intensity of anger in spontaneous situations, as its expression can have strong negative effects on social relations (Fischer & Manstead, 2008). In daily life, people might supress their anger, which results in mild versions that are recognized infrequently. When professional actors are asked to play-act anger, this social control is not necessary and they might produce a more intense expression that is easier to recognize. It could be speculated that non-actors cannot abandon the social constriction while acting. The ambiguity of spontaneous anger is seen in the recognition bias against judging authentic expressions as anger by participants from collectivistic societies. As proposed by Elfenbein et al. (2002), cultural biases might only be effective when the expression is less distinct.

Expressions of sadness frequently showed high recognition rates (K. R. Scherer, 2003), which is not surprising considering that it is the only low aroused negative emotion (with the excepion of low aroused fear; Laukka et al., 2005; Russell, 1980). The social function of sadness is to call for help and support (Fischer & Manstead, 2008; Hendriks & Vingerhoets, 2006), and while also characterized by a lack of control, it does not possess the submissive character of fear

(Ellsworth & Scherer, 2003). Instead of suppressing, people might rather express this emotion to gain support. Play-acting sadness includes actively reducing arousal and activity, which might simply be difficult to obtain.

However, since these explanations are done post-hoc empirical support is still needed. Additionally, rating studies reveal the ability of the listener to judge a stimulus, and they only allow a restricted inference about the production. To account for production differences, an analysis of the acoustic structure is necessary. In my acoustic analysis, I found no interaction between emotion and encoding condition to support the assumptions that play-acted anger expressions were more intense than spontaneous ones or that spontaneous sadness is encoded differently than acted sadness (Chapter 3, Jürgens et al., 2011). Differences in the acoustic structure should exist as the recognition accuracies are stimulus-based (Drolet et al., 2013); these might be too subtle to be detected in the acoustic analysis, or are found in different parameters other than speech rate, fundamental frequency, energy distribution, and harmonic-to-noise ratio (Jürgens et al., 2011).

Alternatively, the speech melody might interact with emotion recognition. More aroused expressions (e.g. anger) are characterized by more vivid intonation while low aroused emotions possess a more monotonous speech melody (Drolet et al., 2014; Juslin & Laukka, 2001). In line with these observations acted expressions might be perceived as anger, while spontaneous expressions are more strongly associated with sadness. This is in line with the observation that people generally were biased to judge professionally acted expressions as anger (Chapter 2).

Last, individual variation in expressivity of our speakers should also be taken into account (Hildebrandt, Olderbak, Sommer, & Wilhelm, 2014; Spackman, Brown, & Otto, 2009). Our stimulus set consisted of recorded by a high number of speakers (78 for the authentic stimuli, 41 for the professionally acted and 39 for the non-professional expressions). Nevertheless, to confirm our findings and to ensure that the emotion recognition pattern is caused not by individual differences in encoding ability (cf. Ekman & Oster, 1979; Hildebrandt et al., 2014), but

by the process of acting in general, a replication using a second, independent stimulus set of spontaneous expressions would be helpful.

5.1.3 Reliability of vocal expressions

Do emotional expressions reliably indicate the affective state of the sender? In my studies, I showed that acting-inexperienced people were able to produce convincing vocal expressions. This finding is in conflict with reliability and suggests that play-acting emotional expressions (via the voice) is a common human ability. Following the argumentation of Goffman (1959), people play-act their emotions every day. He described human interactions as a theatre analogy and according to his approach, humans are social actors, who play a role and who adapt their behavior and their appearance according to the situation and the social expectations. This speaks for the exceeding use of deliberate expressions in daily life. Studies on emotion regulation also demonstrated the ability to change expressive behavior to adapt to social necessities (Gross, 1998; Kappas, 2013). Getting back to the model proposed by U. Scherer et al. (1980), explaining expressions of emotions by push and pull factors, this result should not be surprising (see also K. R. Scherer & Bänziger, 2010). It is however surprising considering that the lack of underlying emotion is not recognized, in light of the importance of deception detection in stable communication systems.

In the case that emotional expressions have a predictive value for the receiver, deception should be minimized (R. A. Johnstone & Grafen, 1993). As summarized in Chapter 5.1.2, there is only evidence that sadness might be difficult to produce deliberately. For the other emotions, although listeners cannot explicitly name the level of authenticity or showed difficulties in recognizing the play-acted versions, mental processes are at least different when attending to acted or spontaneous expressions (Drolet et al., 2012), which in a next step might lead to different behaviors in response to the authenticity. Speech melody (pitch variability), might be difficult to adapt deceptively (see also Audibert et al., 2010; Williams & Stevens, 1972). Yet it is

interesting that trained actors are not more capable of adjusting their behavior and not more convincing than non-trained people, as would be assumed when play-acted expressions would need specific attention or effort during production (Mehu et al., 2012; Schmidt & Cohn, 2001)

I advance the view that pitch variability might be a misleading marker for detecting acted expressions and that under certain conditions expressions are not distinguishable in terms of their authenticity. A look at word stress and the use of pitch variability for differentiating sentence types (such as exclamatory, or interrogative sentence) (Kent & Read, 1992; Lehiste & Peterson, 1961) demonstrates the ability to deliberately manipulate this marker quite precisely. Additionally, if this parameter would represent a common differentiation between spontaneous and play-acted expressions, listeners should be more attentive to it and make use of this parameter, but recognition rates do not indicate this. Pitch variability has been proposed to be positively related to arousal and intensity (Laukka et al., 2005; K. R. Scherer, 2003). Higher pitch variability, that is more vivid speech, in play-acted expressions might be explained by acted expressions being more intense (K. R. Scherer et al., 2011), but interestingly this intensity effect is not consistent in the other acoustic parameters (Jürgens et al., 2011). The origin of the differences in pitch contour is thus still unknown. Notably, both acted conditions deviated from the spontaneous recordings in the fact that actors did not learn the sentences by heart, but read them aloud from a sheet of paper. Acting was thus not the only difference between the speech samples. Speech melody is influenced by reading, although the literature is not consistent on the direction of this effect, namely whether reading increases or diminishes variability of the fundamental frequency (Batliner et al., 1995; Laan, 1997). The differences in pitch variability might therefore not be caused by acting but by reading. This line of thought is however purely speculative. The necessary next step should be to disentangle the effect of reading and acting. A comparison between script-based and improvised acting might uncover which acoustic characteristics are caused by reading and whether improvised portrayals influence response

behavior similarly to spontaneous expressions. A study focusing on this effect is currently under preparation.

Independent of the implicit perception and the acoustic differences, the response on the behavioral level is puzzling. Why are people so capable in play-acting vocal expressions or so poor in detecting the deception respectively (see Dezecache et al., 2013; R. A. Johnstone & Grafen, 1993)? For listeners it should always be of importance to be able to detect emotional deception, not only in the case of sadness. Be taken in by false sadness and supporting the wrong individual might cost important resources, but detecting false anger might also be of advantage. Authentic anger indicates an actual threat and a more powerful opponent, while play-acted anger does not.

One explanation for this lack of deception detection might be that it is more costly to miss an authentic emotion than to mistakenly attend to a faked one (Ekman, 1996). I would state that this explanation is not sufficient, as even attending to false emotional expressions has negative consequences, like mentioned above. Understanding whether an expression is used deceptively or not, might comprise attending to the whole body, including face, voice, body posture and speech content (Ekman & O'Sullivan, 2006; Mortillaro et al., 2013). The synchronization between all expressive channels might be more unmasking than a single channel (Mortillaro et al., 2013). Interestingly, the ability to detect lies on the basis of whole body shots and verbal content has been found to be poor as well (Ekman & O'Sullivan, 1991; Warren et al., 2008; Zuckerman, Koester, & Colella, 1985). A compelling explanation comes from Schmidt and Cohn (2001), who proposed that the detection of faked emotional expressions might be most operative within close social entities, such as friends or romantic partners. Familiar people are of highest significance and so is their deceptive behavior. In these social entities, individuals have a clear knowledge on the emotional expressivity of their partner and might be more attentive (Young & Hugenberg, 2010; Zhang & Parmley, 2010). Familiarity with the expressive pattern of others may be necessary for a successful evaluation of deceptive behavior (but see Levine et al., 1999). However, I could not confirm the notion that social connectedness influenced emotion recognition; at least as

experimentally manipulated similarity did not improve the recognition of vocal emotional expressions (see Chapter 4). In less experimental settings, social connection might have a positive effect (see 5.2.1).

C. D. Frith and Frith (2007) stated that human communication emerge "when both sender and receiver are aware that they are exchanging signals" (p. R724). This knowledge and the ability to mentalize others' internal states make human emotion communication - although evolutionally rooted (Scheiner & Fischer, 2011) - more complex than any non-human communication. Humans do not base their behavioral responses on the perception of expressions alone, but also include situational appraisal, own experience, prior behavior as well as knowledge on intentions and beliefs of the other person in their decisions (cf. for involvement of mentalizing during authenticity perception Drolet et al., 2012; Drolet et al., 2013). Additionally, the act of emotional deception is also done intentionally (for involvement of the mentalizing network during deception, see Lisofsky, Kazzer, Heekeren, & Prehn, 2014), which – in combination with being aware of the production and the effect of emotional expressions - might allow humans to imitate these signals precisely. Single channels, like the voice, might have lost their predictive value. For our study set, I can at least summarize that the voice alone is not sufficient to reliably predict whether an individual undergoes an emotional episode, or not. To strengthen this finding, we need, however, knowledge on the speaker's autonomic reactions to fully reveal his/her affective state during voice recording. Studies on the coherence of emotional components should be of priority in emotion research, as they are essential to understand the nature of emotions and might answer the question about *what is* being communicated (see Fernández-Dols & Crivelli, 2013; Reisenzein et al., 2013).

5.1.4 Implications for using acted expressions in research

Play-acted emotional expressions are the common stimulus material in emotion research. Barrett (2011) stated the question, whether these (facial) emotion portrayals do represent facial

behavior in daily situations, or whether they are merely symbols of emotional expressions. The same question applies for commonly used vocal recordings. Although K. R. Scherer and Bänziger (2010) pointed out that actors' portrayals are not studied to gain knowledge about spontaneous behavior but about expressive codes, implications for daily expressions are nevertheless made on the basis of this research.

I argued that vocal emotional expressions can be play-acted in a realistic fashion, perhaps with the exception of sadness. However, this assumption of realism does not hold for the commonly used emotion portrayals that are produced by the standard recording procedure, in which speakers are asked to express one word (Hammerschmidt & Jürgens, 2007; Leinonen, Hiltunen, Linnankoski, & Laakso, 1997) or one sentence (Banse & Scherer, 1996; Juslin & Laukka, 2001) in a given emotion. Afterwards, speech tokens are preselected for being highly recognizable and thus rather represent a particular set of expressions that overestimates the emotional content of speech recordings. Our studies indicate that acted expressions can be regarded as similar to spontaneous ones, but only under specific recording conditions. We achieved realistic expressions by using longer texts, providing context information, and abstaining from pre-selection (see Enos & Hirschberg, 2006 for their call for more realistic recording conditions). This procedure reduced the recognition accuracies of stimuli compared to other studies (Banse & Scherer, 1996; K. R. Scherer et al., 2001; Van Bezooijen et al., 1983), but just because of this, increased similarity to spontaneous expressions. The assumption that acted expressions are more stereotypical and intense (i.e., K. R. Scherer et al., 2011), is most likely true for all play-acted stimulus sets that were created using preselected recordings. The stereotypy and high intensity is merely founded by the recording conditions than by acting per se. Our emotion portrayals clearly do not reflect the majority of existing stimulus sets, and might be regarded as an example how realistic stimulus corpora can be produced, namely by recording longer text segments, by providing speakers with context situations, and by abstaining from pre-selection.

The idea of adapting recording conditions for producing more realistic expressions, without accessing daily life recordings, is not new (see Enos & Hirschberg, 2006; K. R. Scherer & Bänziger, 2010). Moreover, the importance of realistic data sets has recently been pointed out (Schuller, Batliner, Steidl, & Seppi, 2011). In this context, it has been assumed that the use of specific acting techniques, in which the actor creates actual emotions during his performance, improves authenticity of emotion depiction (Gosselin et al., 1995; K. R. Scherer & Bänziger, 2010). These techniques go back to Stanislavskij (1989) and Strasberg (1987), who highlight the importance of creating inner affective states via recollection of experienced emotional situations (see also Goldstein & Winner, 2010). However, whether these techniques actually are responsible for more realistic acting is still under debate and empirical evidence is lacking (Goldstein & Bloom, 2011; Goldstein & Winner, 2010; Konijn, 1995). The recognition of emotion expressions by acting-inexperienced people closely resembled the spontaneous expressions, indicating that specifically trained actors are not necessary in emotion research. On the contrary, acting- and speech training seems to have even negative effects on the credibility of emotion portrayals. This conclusion does not hold for acting in general, but only for the transmission of (vocally expressed) emotions.

5.2 Processing of emotional expressions

5.2.1 Influences on emotion recognition

Emotion recognition was supposed to be altered by the social connection between interaction partners (Weisbuch & Ambady, 2008). In Chapter 2, I found a positive effect of group-membership on both emotion and authenticity recognition. This in-group effect might be explained by the fact that German participants understood the language and felt more secure in their judgments. However, the stimuli might also have more social relevance for the German participants, and cause an attention-shift or an increase in empathic concern (Thibault et al., 2006; Weisbuch & Ambady, 2008). In Chapter 4, social relevance of the vocal expressions was increased by biographical similarity manipulation that was thought to create a fictitious social link

between speaker and receiver (cf. Burger et al., 2004; Guéguen & Martin, 2009). However, similarity did not alter emotion recognition in the predicted way or intensified emotional engagement. Additionally, vocal expressions in the similar condition did not attract more attention as was assumed (Ackerman et al., 2006; Thibault et al., 2006), which was indicated by the unaffected pupil size. It is possible that the manipulation was not realistic or strong enough to affect the listeners, indicating that an actual social link between both partners is necessary for a similarity manipulation to function, as for example when interacting with one another in person (i.e., Walton et al., 2012). The fact that group membership improved emotion recognition while similarity did not, might be caused by cultural group-membership being more realistic or credible compared to the similarity manipulation in Chapter 4. People speaking an unintelligible language obviously belong to an out-group, while the similarity manipulation via fictitious biographical profiles was apparently constructed.

One interesting explanation why similarity did not result in improved emotion recognition comes from Todd and colleagues (2011). They described that similarity triggers an egocentric perspective. Accordingly, listeners might incorporate their own expectations, situational evaluations and moods into their ratings when listening to expressions by more similar speakers. The emotion recognition would in this case be individually affected. Similarity and group-member ship might have very different influences on human behavior other than increasing social relevance (see Bijlstra et al., 2010; Mussweiler & Ockenfels, 2013).

5.2.2 Sharing emotions

Attending to others' emotions is thought to evoke congruent affects in the receiver (Hatfield et al., 2011; Magnée et al., 2007; Preston & de Waal, 2002). This process has been explained to be of importance for understanding the inner affective states of others (Carr et al., 2003; Goldman & Sripada, 2005; Niedenthal & Maringer, 2009; Stel & van Knippenberg, 2008). Autonomic responses are not affected by the similarity manipulation (Chapter 5.2.1). Moreover,

listening to vocal expressions did not trigger strong autonomic reactions at all, speaking against an evocation of emotional episodes during emotion recognition (for similar results concerning facial expressions see Alpers et al., 2011; Wangelin et al., 2012). Vocal expressions are thus not processed and understood via – measurable - bodily sensations. Autonomic reactions as responses to emotional states in others were found in lifelike situations, namely when a real interaction between both sender and receiver existed (Cwir et al., 2011; Hein et al., 2011; Levenson & Rueff, 1992), but seemingly not when a single expression channel is attended to. The similarity manipulation in Chapter 4 presumably did not strengthen social relevance of the stimulus sufficiently for an effective emotional engagement. It would be a promising experimental approach to look at emotional engagement of naïve participants during manipulated phone calls. In these cases participants would also listened to only one expression channel, but the situation would be more socially relevant. I would additionally assume that biographical similarity in this example would affect emotion processing.

Moreover, the stimulus selection might diminish realism of the experiment additionally as I used intense and preselected emotion expressions to investigate affect sharing. This was done to evoke strong and controllable effects (see K. R. Scherer & Bänziger, 2010). I summarized in Chapter 5.1.4 that emotional expressions can be convincingly play-acted, but that this is most often not the case. Stel and Vonk (2009) demonstrated that whether an emotional situation was perceived as real or not at least facilitated perspective taking, although every emotion in this study was play-acted and only participants' believes varied. There is the possibility that emotional expressions used to investigate affective processing are simply not believed in and thus do not cause any necessity to involve. A follow-up study using real life expressions or realistically acted expressions (see Chapter 3), would disclose this issue.

Emotional expressions did not trigger affective responses when presented in isolation. This result is difficult to bring in line with studies on facial mimicry that are conducted on context free facial expressions or voices (Dimberg et al., 2011; Künecke et al., 2014; Magnée et al., 2007;

Sato et al., 2013; Sonnby-Borgström, 2002). Thus, the question is whether mimicry is an actual indication of emotional episodes during emotion perception (Dimberg, 1997) or a "cognitive mechanism that is not associated with or dependent on any particular affective or emotional state" (Chartrand & Bargh, 1999, p. 904). Moody, McIntosh, Mann, and Weisser (2007) demonstrated that emotion induction in the receiver influenced facial mimicry, which they interpreted as an indication for underlying affective processes during mimicry (Blairy et al., 1999). Emotion-congruent facial reactions to vocal expressions (Hawk, Fischer, & Van Kleef, 2012; Magnée et al., 2007), to affective sounds (Bradley & Lang, 2000) or a boosted mimicry effect when attending to facial expressions of friends (McIntosh, 2006), also contradicts a pure motor mimicry behavior. Chapter 4 of this thesis did not include analyses of facial mimicry, but the absence of arousal-related skin conductance responses when listening to vocal expressions (or looking at facial expressions, see Alpers et al., 2011) speaks against a common affective basis of facial mimicry and autonomic responses. Future studies should include the simultaneous recording of both facial EMG and autonomic reactions, to enlighten this issue on emotion evocation.

Generally, the formation of empathy, i.e. sharing an emotion of others, seems to be a holistic process, and simultaneous processing of all communication channels (speech, gestures, face, voice) as well as context might be necessary for a complete empathic reaction (see Cwir et al., 2011; Hess & Fischer, 2013; Regenbogen et al., 2012). Presumably, emotional expressions are simply recognized first (cf. Britton et al., 2006; Zahavi, 2008), while empathic reactions are formed in a later processing step, after exhaustive evaluation of the situation, including evaluating the relevance of the sender (de Vignemont & Singer, 2006). Empathy is then responsible for costly helping and prosocial behavior (Batson et al., 1981; Hein et al., 2011).

My results are not in contrast with the assumption that emotional expressions are processed via a shared representation between experiencing and attending to an emotional episode (cf. Carr et al., 2003; Singer, 2006). The pupil data presented in Chapter 4 clearly depicts a

level of affective processing. In contrast to SCRs (skin conductance responses), pupil size is a more sensitive marker indicating the slightest differences during stimulus processing (Laeng et al., 2012). The affective reaction is just not intense enough to trigger emotional episodes including autonomic reactions (Laeng et al., 2012; Levenson, 2014). However, my findings indicates that if simulation and mirroring are important for understanding emotions (Carr et al., 2003; Morris et al., 1996; Wicker et al., 2003), the process is merely limited to shared mental representations in the central nervous system and does not automatically radiate to the peripheral physiology and induces "bodily sensations" (Singer, 2006).

5.2.3 Emotional content of vocal expressions

Situations that are of relevance for our well-being and future prospects possess an emotional meaning that elicits emotional reactions in the beholder including autonomic reactions (Lang, 1995), preparing the individual for rapid action. In fact, in Chapter 4 of this thesis I could confirm this notion by revealing reactions of the autonomic nervous system as response to affective sounds (such as dentist drills, bombs, baby babble, or slot machines) (see also Bradley et al., 2001; Bradley & Lang, 2000). Emotional expressions are not only thought to be shared between social partners (Chapter 5.2.2) but also are regarded as possessing emotional content itself, for example seeing an angry looking man approaching may induce fear. As to their social importance and high ecological relevance, emotional expressions are described to have high emotional value (see Okon-Singer et al., 2013). In this context it is even more surprising that sounds induced autonomic reactions, while vocal prosody did not. The same pattern was found for faces and pictures as well, strengthening our findings (Alpers et al., 2011; Wangelin et al., 2012). On closer inspection, emotional expressions and affective sounds or pictures are however quite different in their meaning. Sounds/pictures might simply be more intense (Bayer & Schacht, 2014; Okon-Singer et al., 2013), but presumably this is only one part of the explanation. Emotional expressions are highly social and context dependent. According to Walla and Panksepp (2013),

they do not depict the emotion of the situation itself, but rather the situational evaluation of the sender. Hence to understand the expression completely, the context is crucial for the receiver. Listening to a bomb on the other hand already depicts the whole situation. Beyond that, listening to the explosion of a bomb should involve everyone in proximity, while expressions are social, communicative entities. These examples emphasized the importance of regarding contexts when evaluating emotional expressions. Emotion research has made extensive use of context free expressions. They were used were used in laboratory experiments as emotion elicitors (Porter et al., 2012; Westermann, Spies, Stahl, & Hesse, 1996) and regularly function as stimuli triggering emotion reactions in studies on affective neuroscience (Okon-Singer et al., 2013; Tamietto & de Gelder, 2010). My findings speak for a careful use of emotional expressions as affective stimuli, and are thus in line with Walla and Panksepp (2013), who highlighted the differences in emotional meaning between expressions and scenic stimuli (such as picture, or sound). As a whole, an expression (be it vocal or facial) alone does not have a strong emotional impact on the receiver.

5.3 Conclusion and Outlook

Studying spontaneous, daily life expressions showed how weak and ambiguous the emotional content of speech excerpts might be. The use of intensely and preselected play-acted expressions in prior studies overemphasized the importance of the vocal channel in emotion communication (e.g. K. R. Scherer et al., 2001) and masked emotion-specific difference in recognition accuracies. The voice, nevertheless, clearly transmits emotional information independent of verbal content. However, emotionality of prosodic excerpts is not invariably connected to an underlying emotional episode, but rather controlled by the speaker. Vocal expressions putatively do not inform the receiver about the emotional state of the sender, while receivers do not react with strong affective responses to vocal expressions. The exchange of

emotional information by the vocal channel seems more strongly based on cognitive, rather than on affective processes.

Barrett and colleagues (Barrett, 2011; Barrett, Mesquita, & Gendron, 2011), argued that humans automatically integrate context information into their emotion classification; a fact that demonstrates a meaningful ambiguity of emotional expressions, per se. They spoke for a necessary paradigm shift in emotion research, away from the assumption that expressions (in their case facial expressions) can be read to infer the emotional state of others (see Mulligan & Scherer, 2012), towards viewing them contextualized. Considering the low intrinsic emotional content of vocal expression (Chapter 4) and the inconsistent emotion recognition of realistic emotion expressions (Chapter 2, Chapter 3) my findings are in line with this notion.

Taking a single channel out of the whole picture is artificial, but it is nevertheless necessary before understanding the complex interplay of all expressive channels. Research so far has concentrated on single expression (cf. Cowie et al., 2001; Porter & ten Brinke, 2009; K. R. Scherer et al., 2001), but it is now time to move to more context-based, multimodal and realistic approaches, in order to understand the complexities of emotion communication (see also K. R. Scherer et al., 2011). Future studies should, in line with my findings, focus on two aspects. First, the coherence between the different emotional components should be investigated further. This should be done to investigate whether the subjective feeling, emotion expressions and physiological responses build one unit or whether they are decoupled from each other. We only understand *what* is being communicated when the interplay of autonomic reactions, subjective experience and emotional expressions is disclosed (Reisenzein et al., 2013). Second, research should shift the focus away from single channels to multimodal communication (cf. Mortillaro et al., 2013). It is the next challenge in emotion research to investigate the usage and recognition of emotional expressions regarding all modalities at once, in realistic, but controllable settings (see Bänziger & Scherer, 2007; Kotz & Paulmann, 2007; Regenbogen et al., 2012; Regenbogen et al., 2012; Van den Stock, Righart, & de Gelder, 2007). There are already existing corpora which can

serve to provide stimuli for future studies. For example, the GEMEP – Corpus (Geneva Multimodal Emotion Portrayals, Bänziger & Scherer, 2007) consists of audiovisual recordings allowing the comparison of vocal, facial and bodily expressions. However, this corpus includes only nonsense verbal content and the recordings were acted and preselected, leaving intense and easily recognizable expression. The GEMEP corpus represents a promising start to investigate the interplay of all channels, but closeness to daily life is in this case still questionable. Recording multimodal emotional expression in spontaneous situations is especially difficult due to technical and ethical reasons, but realistic expressions should nevertheless be sought.

In closing, my findings provide two methodological recommendations for future studies on vocal expression. First I demonstrated that acting-inexperienced people could even be more suitable to produce credible vocal expressions than professional actors, which is of importance for the creation of future speech corpora. Second, vocal expressions are less strongly processed by the autonomous nervous system than emotional sounds. This result might be of importance for studies on affective neurosciences that use vocal expressions as affective stimulus material.

6 References

Ackerman, J. M., Shapiro, J. R., Neuberg, S. L., Kenrick, D. T., Becker, D. V., Griskevicius, V., Manner, J. K., Schaller, M. (2006). They all look the same to me (unless they're angry) - From out-group homogeneity to out-group heterogeneity. *Psychological Science, 17*, 836-840.

Agresti, A. (2007). *An introduction to categorical data analysis*. New Jersey: Wiley.

Alpers, G. W., Adolph, D., & Pauli, P. (2011). Emotional scenes and facial expressions elicit different psychophysiological responses. *International Journal of Psychophysiology, 80*, 173-181. doi: 10.1016/j.ijpsycho.2011.01.010

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

Aubergé, V., Audibert, N., & Rilliard, A. (2004). *E-Wiz: A trapper protocol for hunting the expressive speech corpora in Lab*. Paper presented at the 4th LREC, Lisbon, Portugal.

Audibert, N., Aubergé, V., & Rilliard, A. (2008). *How we are not equally competent for discriminating acted from spontaneous expressive speech.* Paper presented at the Speech Prosody 2008, Campinan, Brazil.

Audibert, N., Aubergé, V., & Rilliard, A. (2010). *Prosodic correlates of acted vs. spontaneous discrimination of expressive speech: A pilot study*. Paper presented at the 5th International Conference on Speech Chicago, USA.

Aue, T., Cuny, C., Sander, D., & Grandjean, D. (2011). Peripheral responses to attended and unattended angry prosody: a dichotic listening paradigm. *Psychophysiology, 48*, 385-392. doi: 10.1111/j.1469-8986.2010.01064.x

Averill, J. R. (1980). A constructivist view of emotion. In R. Plutchik & H. Kellermann (Eds.), *Emotion, Theory, Research and Experience* (pp. 305-339). New York: Academic Press.

Bach, D. R., Flandin, G., Friston, K. J., & Dolan, R. J. (2009). Time-series analysis for rapid event-related skin conductance responses. *Journal of Neuroscience Methods, 184*, 224-234. doi: 10.1016/j.jneumeth.2009.08.005

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

Bänziger, T., & Scherer, K. R. (2007). Using actor portrayals to systematically study multimodal emotion expression: the GEMEP corpus. In A. Paiva, R. Prada & R. W. Picard (Eds.), *Affective computing and intelligent interaction*. Berlin, Heidelberg: Springer.

Barkhuysen, P., Krahmer, E., & Swerts, M. (2007). *Cross-modal perception of emotional speech*. Paper presented at the ICPhS, Saarbrücken, Germany.

Barrett, L. F. (2009). Variety is the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition & Emotion, 23*, 1284-1306.

Barrett, L. F. (2011). Was Darwin wrong about emotional expressions? *Current Directions in Psychological Science, 20*, 400-406. doi: 10.1177/0963721411429125

Barrett, L. F., Mesquita, B., & Gendron, M. (2011). Context in emotion perception. *Current Directions in Psychological Science, 20*, 286-290. doi: 10.1177/0963721411422522

Bates, D. (2005). Fitting linear mixed models in R using the lme4 package. *R News, 5*, 27-30.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes (Version R package version 0.999375-42). Retrieved from http://CRAN.R-project.org/package=lme4

Batliner, A., Kompe, R., A., K., Nöth, E., & Niemann, H. (1995). Can yout tell apart spontaneous and read speech if you just look at prosody? *Speech Recognition and Coding, 147*, 321-324.

Batliner, A., Fischer, K., Huber, R., Spilker, J., & Nöth, E. (2000). *Desperately seeking emotions or: Actors, wizards and human beings.* Paper presented at the ISCA Workshop on Speech and Emotion, Newcastle, Northern Ireland.

Batson, C. D., Duncan, B. D., Ackerman, P., Buckley, T., & Birch, K. (1981). Is empathic emotion a source of alturistic motivation? *Journal of Personality and Social Psychology, 40*, 290-302.

Bayer, M., & Schacht, A. (2014). Event-related brain responses to emotional words, pictures, and faces – A cross-domain comparison. *Frontiers in Psychology, 5*(1106).

Bayer, M., Sommer, W., & Schacht, A. (2011). Emotional words impact the mind but not the body: Evidence from pupillary responses. *Psychophysiology, 48*, 1554-1562. doi: 10.1111/j.1469-8986.2011.01219.x

Benedek, M., & Kaernbach, C. (2010a). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190*, 80-91. doi: 10.1016/j.jneumeth.2010.04.028

Benedek, M., & Kaernbach, C. (2010b). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology, 47*, 647-658.

Bijlstra, G., Holland, R. W., & Wigboldus, D. H. J. (2010). The social face of emotion recognition: Evaluations versus stereotypes. *Journal of Experimental Social Psychology, 46*, 657-663. doi: 10.1016/j.jesp.2010.03.006

Blairy, S., Herrera, P., & Hess, U. (1999). Mimicry and the judgment of emotional facial expressions. *Journal of Nonverbal Behavior, 23*, 5-41. doi: 10.1023/a:1021370825283

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.1.11) [Computer program]. Retrieved August 4, 2009 from http://www.praat.org/

Bogart, K. R., & Matsumoto, D. (2010). Facial mimicry is not necessary to recognize emotion: Facial expression recognition by people with Moebius syndrome. *Social Neuroscience, 5*, 241-251.

Borod, J. C., Yecker, S. A., Brickman, A. M., Moreno, C. R., Sliwinski, M., Foldi, N. S.., . . . Welkowitz, J. (2004). Changes in posed facial expression across the adult life span. *Experimental Aging Research, 30*, 305-333.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion, 1*, 276-298. doi: 10.1037//1528-3542.1.3.276

Bradley, M. M., Codispoti, M., Sabatinelli, D., & Lang, P. J. (2001). Emotion and motivation II: sex differences in picture processing. *Emotion, 1*, 300.

Bradley, M. M., & Lang, P. J. (1999). *The International affective digitized sounds (IADS)[: stimuli, instruction manual and affective ratings*: NIMH Center for the Study of Emotion and Attention.

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology, 37*, 204-215.

Bradley, M. M., & Lang, P. J. (2007). *The international affective digitized sound (2nd Edition; IADS-2): Affective ratings of sound and struction manual*. Gainsville, Fl: University of Florida.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*, 602-607. doi: 10.1111/j.1469-8986.2008.00654.x

Britton, J. C., Taylor, S. F., Sudheimer, K. D., & Liberzon, I. (2006). Facial expressions and complex IAPS pictures: common and differential networks. *Neuroimage, 31*, 906-919. doi: 10.1016/j.neuroimage.2005.12.050

Brothers, L. (1990). The social brain: A project for integrating primate behavior and neurophysiology in a new domain. *Concepts in Neuroscience, 1*, 27-51.

Brown, L. M., Bradley, M. M., & Lang, P. J. (2006). Affective reactions to pictures of ingroup and outgroup members. *Biol Psychol, 71*, 303-311. doi: 10.1016/j.biopsycho.2005.06.003

Burger, J. M., Messian, N., Patel, S., del Prado, A., & Anderson, C. (2004). What a coincidence! The effects of incidental similarity on compliance. *Personality and Social Psychology Bulltin, 30*, 35-43. doi: 10.1177/0146167203258838

Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., & Weiss, B. (2005). *A database of German emotional speech.* Paper presented at the Interspeech, Lissabon, Portugal.

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas.

Proceedings of the National Academy of Sciences, 100, 5497-5502. doi: 10.1073/pnas.0935845100

Carroll, J. M., & Russell, J. A. (1997). Facial expressions in Hollywood's protrayal of emotion. *Journal of Personality and Social Psychology, 72*, 164-176. doi: doi: 10.1037/0022-3514.72.1.164.

Carruthers, P. (1996). Simulation and self-knowledge: A defense of theory-theory. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 22-38). Cambridge: Cambridge University Press.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology, 76*, 893-910. doi: 10.1037/0022-3514.76.6.893

Christensen, R. H. B. (2012). ordinal --- Regression models for ordinal data R package version 2012.09-11. Retrieved from http://www.cran.r-project.org/package=ordinal website:

Coricelli, G. (2005). Two levels of mental state attribution: From auto-maticity to voluntariness. *Neuropsychologia, 43*, 294-300.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., & Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine, 18*(1), 32-80.

Cwir, D., Carr, P. B., Walton, G. M., & Spencer, S. J. (2011). Your heart makes my heart move: Cues of social connectedness cause shared emotions and physiological states among strangers. *Journal of Experimental Social Psychology, 47*, 661-664. doi: 10.1016/j.jesp.2011.01.009

Dalgleish, T. (2006). The emotional brain. *Nature Reviews Neuroscience, 5*, 582-589.

Darwin, C. (1872). *The expression of emotions in man and animal*. London: John Murray.

Davies, M., & Stone, T. (1998). Folk Psychology and mental simulation. *Royal Institute of Philosophy Supplement, 43*, 53-82.

Dawson, M. E., Schell, A. M., & Filion, D. L. (2007). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 159-181). Cambridge: University Press.

De Gelder, B. (2009). Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B, 364*, 3475-3484.

De Gelder, B., & Van den Stock, J. (2011). The bodily expressive action stimulus test (BEAST). Construction and validation of a stimulus basis for measuring perception of whole body expression of emotions. *Frontiers in Psychology, 2*. doi: doi: 10.3389/fpsyg.2011.00181

de Vignemont, F., & Singer, T. (2006). The empathic brain: how, when and why? *Trends in Cognitive Sciences, 10*, 435-441. doi: 10.1016/j.tics.2006.08.008

Dezecache, G., Mercier, H., & Scott-Phillips, T. C. (2013). An evolutionary approach to emotional communication. *Journal of Pragmatics, 59*, 221-233. doi: 10.1016/j.pragma.2013.06.007

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events - A neurophysiological study. *Experimental Brain Research, 91*, 176-180.

Dimberg, U. (1982). Facial reactions to facial expressions. *Psychophysiology, 19*, 643-647. doi: 10.1111/j.1469-8986.1982.tb02516.x

Dimberg, U. (1997). Facial reactions: Rapidly evoked emotional responses. *Journal of Psychophysiology, 11*, 115-123.

Dimberg, U., Andréasson, P., & Thunberg, M. (2011). Emotional empathy and facial reactions to facial expressions. *Journal of Psychophysiology, 25*, 26-31. doi: 10.1027/0269-8803/a000029

Dimberg, U., & Thunberg, M. (2012). Empathy, emotional contagion, and rapid facial reactions to angry and happy facial expressions. *PsyCh Journal, 1*, 118-127. doi: 10.1002/pchj.4

Douglas-Cowie, E., Campbell, N., Cowie, R., & Roach, P. (2003). Emotional speech: Towards a new generation of databases. *Speech Communication, 40*, 33-60.

Drolet, M., Schubotz, R. I., & Fischer, J. (2012). Authenticity affects the recognition of emotions in speech: behavioral and fMRI evidence. *Cognitive Affective & Behavioral Neuroscience, 12*, 140-150. doi: 10.3758/s13415-011-0069-3

Drolet, M., Schubotz, R. I., & Fischer, J. (2013). Explicit authenticity and stimulus features interact to modulate BOLD response induced by emotional speech. *Cognitive Affective & Behavioral Neuroscience, 13*, 318-329. doi: 10.3758/s13415-013-0151-0

Drolet, M., Schubotz, R. I., & Fischer, J. (2014). Recognizing the authenticity of emotional expressions: F0 contour matters when you need to know. *Frontiers in Human Neuroscience, 8*, 144. doi: 10.3389/fnhum.2014.00144

Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to pro-social and related behaviors. *Psychological Bulletin, 101*, 91-119. doi: 10.1037//0033-2909.101.1.91

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*, 384-392.

Ekman, P. (1996). Why dont't we catch liars? *Social Research, 63*, 801-817.

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion*. Sussex: John Wiley & Sons, Ltd.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology II. *Journal of Personality and Social Psychology, 58*, 342-353.

Ekman, P., & Friesen, W. V. (1969a). Nonverbal leakage and clues to deception. *Psychiatry*, 32, 88-105.

Ekman, P., & Friesen, W. V. (1969b). The repertoire of nonverbal behavior: Categories, origins, usage and coding. *Semiotica, 1*, 49-98.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in face and emotion. *Journal of Personality and Social Psychology, 17*, 124-129.

Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial cues*. Englewood Cliffs, NJ: Prentice Hall.

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system: Part two*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior, 6*, 238-252.

Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist, 46*, 913-920.

Ekman, P., & O'Sullivan, M. (2006). From flawed self-assessment to blatant whoppers: The utility of voluntary and involuntary behavior in detecting deception. *Behavioral Sciences & the Law, 24*, 673-686. doi: 10.1002/bsl.729

Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual Review of Psychology, 30*, 527-554.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science, 164*, 86-88.

Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*, 203-235. doi: 10.1037//0033-2909.128.2.203

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. doi: 10.1037/1528-3542.7.1.131

Elfenbein, H. A., Mandal, M., Ambady, N., Harizuka, S., & Kumar, S. (2002). Cross-cultural patterns in emotion recognition: Highlighting design and analytical techniques. *Emotion, 2*, 75-84.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 572-595). New York: Oxford University Press.

Enos, F., & Hirschberg, J. (2006). *A framework for eliciting emotional speech: Capitalizing on the actor's process.* Paper presented at the LREC Workshop on Corpora for Research on Emotion and Affect. , Genoa, Italy.

Ethofer, T., De Ville, D. V., Scherer, K., & Vuilleumier, P. (2009). Decoding of emotional information in voice-sensitive cortices. *Current Biology, 19*, 1028-1033. doi: 10.1016/j.cub.2009.04.054

Eskenazi, M. (1992). *Changing speech styles: Strategies in read speech and casual and careful spontaneous speech.* Paper presented at the International Conference on Spoken Language Processing, Banff, Alberta, Canada.

Fernández-Dols, J.-M., & Crivelli, C. (2013). Emotion and expression: Naturalistic studies. *Emotion Review, 5*, 24-29.

Fernández-Dols, J.-M., & Ruiz-Belda, M.-A. (1995). Are smiles a sign of happiness? Gold medal winners at the Olympic Games. *Journal of Personality and Social Psychology, 69*, 1113-1119. doi: doi: 10.1037/0022-3514.69.6.1113

Fichtel, C., Hammerschmidt, K., & Jürgens, U. (2001). On the vocal expression of emotion. A multi-parametric analysis of different states of aversion in squirrel monkey. *Behaviour, 138*, 97-116.

Fischer, A. H., & Manstead, A. S. R. (2008). Social functions of emotion. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of Emotions* (Third Edition ed., pp. 456-468). New York: The Guilford Press.

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences, 4*, 258-267.

Frick, R. W. (1985). Communicating emotion - the role of prosodic features. *Psychological Bulletin, 97*, 412-429.

Fridlund, A. J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology, 60*, 229-240. doi: doi: 10.1037/0022-3514.60.2.229

Frijda, N. H. (1986). *The emotions*. Cambridge, England: Cambridge University Press.

Frijda, N. H., & Scherer, K. R. (2009). Emotion definitions (psychological perspectives). In D. Sander & K. R. Scherer (Eds.), *The Oxford Companion to Emotion and Affective Sciences* (pp. 142-144). New York, NY: Oxford Universitiy Press.

Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology, 17*, R724–R732.

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences, 358*, 459-473. doi: 10.1098/rstb.2002.1218

Gallese, V. (2003). The roots of empathy: The shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology, 36*, 171-180. doi: 10.1159/000072786

Gobl, C., & Ni Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication, 40*, 189-212.

Goffman, E. (1959). *The presentation of self in everyday life*. Oxford England: Doubleday.

Goldman, A. I., & Sripada, C. S. (2005). Simulationist models of face-based emotion recognition. *Cognition, 94*, 193-213. doi: 10.1016/j.cognition.2004.01.005

Goldstein, T. R., & Bloom, P. (2011). The mind on stage: why cognitive scientists should study acting. *Trends in Cognitive Sciences, 15*, 141-142.

Goldstein, T. R., & Winner, E. (2010). A new lens on the development of social cognition - The study of acting. In C. Milbrath & C. Lightfood (Eds.), *Art and Human Development*. New York: Psychology Press.

Gosselin, P., Kirouac, G., & Dore, F. Y. (1995). Components and recognition of facial expression in the communication of emotion by actors. *Journal of Personality and Social Psychology, 68*, 83-96.

Goudbeek, M., & Scherer, K. (2010). Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *The Journal of the Acoustical Society of America, 128*, 1322. doi: 10.1121/1.3466853

Greasley, P., Sherrard, C., & Waterman, M. (2000). Emotion in language and speech: Methodological issues in naturalistic approaches. *Language and Speech, 43*, 355-375.

Grimm, M., Kroschel, K., & Narayanan, S. S. (2008). *The Vera am Mittag German audio-visual emotional speech database*. Paper presented at the IEEE International Conference on Multimedia and Expo, Hannover, Germany.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271.

Guéguen, N., & Martin, A. (2009). Incidental similarity facilitates behavioral mimicry. *Social Psychology, 40*, 88-92. doi: 10.1027/1864-9335.40.2.88

Gunnary, S. D., & Hall, J. A. (2014). The Duchenne smile and persuasion. *Journal of Nonverbal Behavior, 38*, 181-194.

Halberstadt, A. G., Dennis, P. A., & Hess, U. (2011). The influence of family expressiveness, individuals' own emotionality, and self-expressiveness on perceptions of others' facial expressions. *Journal of Nonverbal Behavior, 35*, 35-50. doi: 10.1007/s10919-010-0099-5

Hammerschmidt, K., & Jürgens, U. (2007). Acoustical correlates of affective prosody. *Journal of Voice, 21*, 531-540.

Hatfield, E., Rapson, R. L., & Le, Y. L. (2011). Emotional contagion and empathy *The social neuroscience of empathy* (pp. 19-28). Boston, MA: MIT Press.

Hawk, S. T., Fischer, A. H., & Van Kleef, G. A. (2012). Face the noise: embodied responses to nonverbal vocalizations of discrete emotions. *Journal of Personality and Social Psychology, 102*, 796-814. doi: 10.1037/a0026234

Hein, G., Lamm, C., Brodbeck, C., & Singer, T. (2011). Skin conductance response to the pain of others predicts later costly helping. *PLoS ONE, 6*, e22759. doi: 10.1371/journal.pone.0022759

Hendriks, M. C. P., & Vingerhoets, A. J. J. M. (2006). Social messages of crying faces: Their influence on anticipated person perception, emotions and behavioural responses. *Cognition & Emotion, 20*, 878-886.

Hess, U., & Blairy, S. (2001). Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. *International Journal of Psychophysiology, 40*, 129-141.

Hess, U., & Fischer, A. (2013). Emotional mimicry as social regulation. *Personality and Social Psychology Review, 17*, 142-157. doi: 10.1177/1088868312472607

Hess, U., & Kleck, R. E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology, 20*, 369-385.

Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science, 132*, 349-350.

Hildebrandt, A., Olderbak, S., Sommer, W., & Wilhelm, O. (2014). Modeling individual differences in facial expressivity. *Personality and Individual Differences, 60*, S36.

Hochschild, A. R. (1979). Emotion work, feeling rules, and social structure. *The American Journal of Sociology, 85*, 551-575.

Hofstede, G. (1980). *Cultures's consequences*. Berverly Hills, CA: Sage.

Hofstede, G. (1996). The Nation-state as a source of common mental programming: Similarities and differences across Eastern and Western Europe. In S. Gustavsson & L. Lewin (Eds.), *The Future of Nation State - Essays on cultural pluralism and political integration* (pp. 2-20). London, New York: Routledge.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal, 50*, 346-363.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science, 14*, 640-643.

Hunt, W. (1941). Recent development in the field of emotions. *Psychological Bulletin, 38*, 249-276.

Izard, C. E. (1971). *The face of emotion*. New York: Appelton-Century-Crofts.

Izard, C. E. (1992). Basic emotions, relations among emotions, and emotion-cognition relations. *Psychological Review, 99*, 561-565.

Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 7241-7244. doi: 10.1073/pnas.1200155109

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *NeuroImage, 24*, 771-779. doi: 10.1016/j.neuroimage.2004.09.006

James, W. (1884). What is an emotion? *Mind, 9*(34), 188-205.

Johnstone, R. A., & Grafen, A. (1993). Dishonesty and the handicap principle. *Animal Behaviour, 46*, 759-764.

Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions.* (Second Edition ed., pp. 220-235). New York, London: The Guilford Press.

Johnstone, T., van Reekum, C. M., Hird, K., Kirsner, K., & Scherer, K. R. (2005). Affective speech elicited with a computer game. *Emotion, 5*(4), 513-518. doi: 10.1037/1528-3542.5.4.513

Jones, J. T., Pelham, B. W., Carvallo, M., & Mirenberg, M. C. (2004). How do I love thee? Let me count the Js: Implicit egotism and interpersonal attraction. *Journal of Personality and Social Psychology, 87*(5), 665-683. doi: 10.1037/0022-3514.87.5.665

Jürgens, R., Drolet, M., Pirow, R., Scheiner, E., & Fischer, J. (2013). Encoding conditions affect recognition of vocally expressed emotions across cultures. *Frontiers in Psychology, 4*. doi: 10.3389/fpsyg.2013.00111

Jürgens, R., Hammerschmidt, K., & Fischer, J. (2011). Authentic and play-acted vocal emotion expressions reveal acoustic differences. *Frontiers in Psychology, 2*, 180.

Juslin, P. N. (2012). Are musical emotions invariant across cultures? *Emotion Review, 4*, 283-284. doi: 10.1177/1754073912439773

Juslin, P. N., & Laukka, P. (2001). Impact of intended emotion intensity on cue utilization and decoding accuracy in vocal expression of emotion. *Emotion, 1*, 381-412.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129*, 770-814. doi: 10.1037/0033-2909.129.5.770

Kappas, A. (2013). Social regulation of emotion: Messy layers. *Frontiers in Psychology, 4,* 51.

Kent, R. D., & Read, C. (1992). *The acoustic analysis of speech*. San Diego, London Singular Publishing Group.

Kilner, J. M., & Lemon, R. N. (2013). What we know currently about mirror neurons. *Current Biology, 23*, R1057-R1062.

Konijn, E. (1995). Actors and emotion: A psychological perspective. *Theater Research International, 20*, 132-140.

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

Kracauer, S. (2005). Remarks on the actor. In K. Knopf (Ed.), *Theater and film: A comparative anthology* (pp. 323-333). New York: Yale University Press.

Krahmer, E., & Swerts, M. (2008). *On the role of acting skills for the collection of simulated emotional speech.* Paper presented at the INTERSPEECH, Brisbane, Australia.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*, 394-421. doi: 10.1016/j.biopsycho.2010.03.010

Kret, M. E., & De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia, 50*, 1211-1221. doi: 10.1016/j.neuropsychologia.2011.12.022

Krumhuber, E. G., & Manstead, A. S. R. (2009). Can Duchenne Smiles be feigned? New evidence on felt and false smiles. *Emotion, 9*, 807-820. doi: 10.1037/a0017844

Kuchinke, L., Schneider, D., Kotz, S. A., & Jacobs, A. M. (2011). Spontanous but not explicit processing of positive sentences impaired in Asperger's syndrom: Pupillometric evidence. *Neuropsychologia, 49*, 331-338.

Künecke, J., Hildebrandt, A., Recio, G., Sommer, W., & Wilhelm, O. (2014). Facial EMG responses to emotional expressions are related to emotion perception ability. *PLoS One, 9*), e84053. doi: 10.1371/journal.pone.0084053

Laan, G. P. M. (1997). The contribution of intonation, segmental duration , and spectral features on perception of a spontaneous and a read speaking style. *Speech communication, 22*, 43-65.

Laeng, B., Sirois, S., & Gredeback, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science, 7*, 18-27. doi: 10.1177/1745691611427305

Lang, P. J. (1995). The emotion probe: Studies of motivation and attention. *American Psychologists, 50*(5), 372-385.

Laukka, P., Audibert, N., & Aubergé, V. (2012). Exploring the determinants of the graded structure of vocal emotion expressions. *Cognition & Emotion, 26*, 710-719. doi: 10.1080/02699931.2011.602047

Laukka, P., Elenius, K., Fredrikson, M., Furmark, T., & Neiberg, D. (2008). *Vocal expression in spontaneous and experimentally induced affective speech: Acoustic correlates of anxiety, irritation and resignation.* Paper presented at the International Conference of Language, Resources and Evaluation, Marrakech, Morocco.

Laukka, P., Juslin, P. N., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition & Emotion, 19*, 633-653. doi: 10.1080/02699930441000445

Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, Ö., Faria, V., . . . Furmark, T. (2008). In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech. *Journal of Nonverbal Behavior, 32*, 195-214. doi: 10.1007/s10919-008-0055-9

Lazarus, R. S. (1991). *Emotion and adaption*. New York: Oxford University Press.

Lazarus, R. S., Averill, J. R., & Opton, E. M. (1970). Towards a cognitive theory of emotion. In M. B. Arnold (Ed.), *Feelings and emotion* (pp. 207-232). New York: Academic Press.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America, 33*, 419-425.

Leinonen, L., Hiltunen, T., Linnankoski, I., & Laakso, M. L. (1997). Expression of emotional-motivational connotations with a one-word utterance. *Journal of the Acoustical Society of America, 102*, 1853-1863.

Levenson, R. W. (2011). Basic emotion questions. *Emotion Review, 3*, 379-386. doi: 10.1177/1754073911410743

Levenson, R. W. (2014). The autonomic nervous system and emotion. *Emotion Review, 6*, 100-112.

Levenson, R. W., & Rueff, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology, 63*, 234-246.

Levine, T. R., Park, H. S., & McCornack, S. A. (1999). Accuracy in detecting truths and lies: Documenting the "veracity effect". *Communication Monographs, 66*, 125-144. doi: 10.1080/03637759909376468

Lindquist, K., Siegel, E. H., Quigly, K. S., & Barrett, L. F. (2013). The hundred-year emotion war: Are emotions natural kinds or psychological constructions? Comment on Lench, Flores, and Bench (2011). *Psychological Bulletin, 139*, 255-263.

Lipps, T. (1903). Einfühlung, innere Nachahmung und Organempfindung. *Archive für die gesamte Psychologie* (Vol. 1, pp. 185-204). Leipzig: Engelmann.

Lisofsky, N., Kazzer, P., Heekeren, H. R., & Prehn, K. (2014). Investigating socio-cognitive processes in deception: A quantitative meta-analysis of neuroimaging studies. *Neuropsychologia, 61*, 113-122.

Lithari, C., Frantzidis, C., Papadelis, C., Vivas, A. B., Klados, M., Kourtidou-Papadeli, C., . . . Bamidis, P. (2010). Are females more responsive to emotional stimuli? A neurophysiological study across arousal and valence dimensions. *Brain topography, 23*, 27-40.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

Luce, R. D. (1963). A threshold theory for simple detection experiments. *Psychological Review, 70*, 61-79.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing, 10*, 325-337.

Maccoby, E. E., & Wilson, W. C. (1957). Identification and observational-learning from films. *Journal of Abnormal and Social Psychology, 55*, 76-87.

Magnée, M. J., Stekelenburg, J. J., Kemner, C., & de Gelder, B. (2007). Similar facial electromyographic responses to faces, voices, and body expressions. *Neuroreport, 18*, 369-372.

Masten, C. L., Morelli, S. A., & Eisenberger, N. I. (2011). An fMRI investigation of empathy for 'social pain' and subsequent prosocial behavior. *NeuroImage, 55*, 381-388. doi: 10.1016/j.neuroimage.2010.11.060

Master, S., Debiase, N., Chiari, B., & Laukkanen, A. (2008). Acoustic and perceptual analyses of Brazilian male actors' and nonactors' voices: Long-term average spectrum and the “actor's formant”. *Journal of Voice, 22*, 146-154. doi: 10.1016/j.jvoice.2006.09.006

Mathur, V. A., Harada, T., Lipke, T., & Chiao, J. Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *NeuroImage, 51*, 1468-1475. doi: 10.1016/j.neuroimage.2010.03.025

Matsumoto, D. (1989). Cultural influences on the perception of emotion. *Journal of Cross-Cultural Psychology, 20*, 92-105.

Matsumoto, D. (1992). American-japanese cultural differences in the recognition of universal facial expressions. *Journal of Cross-Cultural Psychology, 23*, 72-84. doi: 10.1177/0022022192231005

Matsumoto, D., & Hwang, H. S. (2011). Culture and emotion: The integration of biological and cultural contributions. *Journal of Cross-Cultural Psychology, 43*, 91-118. doi: 10.1177/0022022111420147

Matsumoto, D., Olide, A., & Willingham, B. (2009). Is there an ingroup advantage in recognizing spontaneously expressed emotions? *Journal of Nonverbal Behavior, 33*, 181-191. doi: 10.1007/s10919-009-0068-z

Matsumoto, D., Seung Hee, Y., & Fontaine, J. (2008). Mapping expressive differences around the world: The relationship between emotional display rules and individualism versus collectivism. *Journal of Cross-Cultural Psychology, 39*, 55-74. doi: 10.1177/0022022107311854

Mauss, I. B., Levenson, R. W., McCarter, L., Wilhelm, F. H., & Gross, J. J. (2005). The tie that binds? Coherence among emotion experience, behavior, and physiology. *Emotion, 5*, 175-190. doi: 10.1037/1528-3542.5.2.175

Maynard Smith, J. (1991). Honest signalling: The Philip Sidney game *Animal Behaviour, 42*, 1034-1035.

Maynard Smith, J., & Harper, D. G. C. (1995). Animal signals: Models and terminology. *Journal of Theroretical Biology, 177*, 305-311.

McIntosh, D. N. (2006). Spontaneous facial mimicry, liking and emotional contagion. *Polish Psychological Bulletin, 37*, 31-42.

Mehu, M., Mortillaro, M., Bänziger, T., & Scherer, K. R. (2012). Reliable facial muscle activation enhances recognizability and credibility of emotional expression. *Emotion, 12*, 701-715. doi: 10.1037/a0026717

Mehu, M., & Scherer, K. R. (2012). A psycho-ethological approach to social signal processing. *Cogn Process, 13 Suppl 2*, 397-414. doi: 10.1007/s10339-012-0435-2

Mier, D., Lis, S., Neuthe, K., Sauer, C., Esslinger, C., Gallhofer, B., & Kirsch, P. (2010). The involvement of emotion recognition in affective theory of mind. *Psychophysiology, 47*, 1028-1039. doi: 10.1111/j.1469-8986.2010.01031.x

Miller, D. T., Downs, J. S., & Prentice, D. A. (1998). Minimal conditions for the creation of a unit relationship: The social bond between birthdaymates. *European Journal of Social Psychology, 28*, 475-481.

Moody, E. J., McIntosh, D. N., Mann, L. J., & Weisser, K. R. (2007). More than mere mimicry? The influence of emotion on rapid facial reactions to faces. *Emotion, 7*, 447-457.

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the art and future development. *Emotion Review, 5*, 119-124.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383*, 812-815.

Mortillaro, M., Mehu, M., & Scherer, K. (2013). The evolutionary origin of multimodal synchronization and emotional expression. In E. Altenmüller, S. Schmidt & E. Zimmermann (Eds.), *Evolution of Emotional Communication: From Sounds in Nonhuman Mammals to Speech and Music in Man (2013): 3-25* (pp. 3-25). Oxford: Oxford University Press.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology, 20*, 750-756. doi: 10.1016/j.cub.2010.02.045

Mulligan, K., & Scherer, K. R. (2012). Toward a working definition of emotion. *Emotion Review, 4*, 345-357. doi: 10.1177/1754073912445818

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech - a review of the literature on human vocal emotion. *Journal of the Acoustical Society of America, 93*, 1097-1108.

Mussweiler, T., & Ockenfels, A. (2013). Similarity increases altruistic punishment in humans. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 19318-19323.

Nawka, T., Anders, L. C., Cebulla, M., & Zurakowski, D. (1997). The speaker's formant in male voices. *Journal of Voice, 11*, 422-428.

Niedenthal, P. M. (2007). Embodying emotion. *Science, 316*, 1002-1005. doi: 10.1126/science.1136930

Niedenthal, P. M., & Maringer, M. (2009). Embodied emotion considered. *Emotion Review, 1*, 122-128. doi: 10.1177/1754073908100437

Norscia, I., & Palagi, E. (2011). Yawn contagion and empathy in *Homo sapiens*. *PLoS ONE, 6*, e28472.

Okon-Singer, H., Lichtenstein-Vidne, L., & Cohen, N. (2013). Dynamic modulation of emotional processing. *Biological Psychology, 92*, 480-491.

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

Ortony, A., & Turner, T. J. (1990). Whats basic about basic emotions. *Psychological Review, 97*, 315-331.

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies, 59*, 185-198. doi: 10.1016/s1071-5819(03)00017-x

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology, 4*, 345. doi: 10.3389/fpsyg.2013.00345

Paulmann, S., & Kotz, S. A. (2007). Early emotional prosody perception based on different speaker voices. *NeuroReport, 19*, 209-213.

Pell, M. D., & Kotz, S. A. (2011). On the time course of vocal emotion recognition. *PLoS ONE, 6*, e27256. doi: 10.1371/journal.pone.0027256

Pell, M. D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S. A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics, 37*, 417-435. doi: 10.1016/j.wocn.2009.07.005

Pell, M. D., Jaywant, A., Monetta, L., & Kotz, S. A. (2011). Emotional speech processing: disentangling the effects of prosody and semantic cues. *Cognition & Emotion, 25*, 834-853. doi: 10.1080/02699931.2010.516915

Pell, M. D., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication, 50*, 519-530. doi: 10.1016/j.specom.2008.03.006

Pinker, S. (1997). *How the mind works*. New York: Norton.

Porter, S., & ten Brinke, L. (2009). Reading between the lies: Identifying concealed and falsified emotions in universal facial expressions. *Psychological Science, 19*, 508-514.

Porter, S., ten Brinke, L., & Wallace, B. (2012). Secrets and lies: Involuntary leakage in deceptive facial expressions as a function of emotional intensity. *Journal of Nonverbal Behavior, 36*, 23-37.

Preis, M. A., & Kroener-Herwig, B. (2012). Empathy for pain: The effects of prior experience and sex. *European Journal of Pain*, n/a-n/a. doi: 10.1002/j.1532-2149.2012.00119.x

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind. *The Behavioral and Brain Siences, 4*, 515-526.

Preston, S. D. (2007). A perception action model for empathy. In T. Farrow & P. Woodruff (Eds.), *Empathy in mental illness* (pp. 428-447). Cambridge, UK: Cambridge University Press.

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*, 1-72.

Prinz, J. (2004). Which emotions are basic? In D. Evans & P. Cruse (Eds.), *Emotion, Evolution, and Rationality* (pp. 69-88). Oxford, UK: Oxford University Press.

R Developmental Core Team. (2012). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Ramachandra, V., Depalma, N., & Lisiewski, S. (2009). The role of mirror neurons in processing vocal emotions: evidence from psychophysiological data. *International Journal Neuroscience, 119*, 681-690. doi: 10.1080/00207450802572188

Raney, A. A. (2004). Expanding Disposition Theory: Reconsidering Character Liking, Moral Evaluations, and Enjoyment. *Communication Theory, 14*, 348-369. doi: 10.1111/j.1468-2885.2004.tb00319.x

Regenbogen, C., Schneider, D. A., Finkelmeyer, A., Kohn, N., Derntl, B., Kellermann, T., . . . Habel, U. (2012). The differential contribution of facial expressions, prosody, and speech content to empathy. *Cognition & Emotion*, 1-20. doi: 10.1080/02699931.2011.631296

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*, 2346-2356. doi: 10.1016/j.neuroimage.2012.02.043

Reisenzein, R., Bördgen, S., Holtbernd, T., & Matz, D. (2006). Evidence for strong dissociation between emotion and facial displays: The case of surprise. *Journal of Personality and Social Psychology, 91*, 295-315. doi: 10.1037/0022-3514.91.2.295

Reisenzein, R., Studtmann, M., & Horstman, G. (2013). Coherence between emotion and facial expression: Evidence from laboratory experiments. *Emotion Review, 5*, 16-23.

Riediger, M., Voelkle, M. C., Ebner, M. C., & Lindenberger, U. (2011). Beyond "happy, angry or sad?" Age-of-poser and age-of-rater effects on multi-dimensional emotion perception. *Cognition & Emotion, 25*, 968-982.

Riediger, M., Studtmann, M., Westphal, M., Rauers, A., & Weber, H. (2014). No smile like another: Adult age differences in identifying emotions that accompany smiles. *Frontiers in Psychology, 5*.

Rinn, W. E. (1984). The neuropsychology of facial expression: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin 95*, 52-77.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131-141. doi: 10.1016/0926-6410(95)00038-0

Roy, N., Ryker, K. S., & Bless, D. M. (2000). Vocal violence in actors: An investigation into its acoustic consequences and the effects of hygienic laryngeal release training. *Journal of Voice, 14*, 215-230. doi: 10.1016/s0892-1997(00)80029-6

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161-1178.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145-172. doi: 10.1037/0033-295x.110.1.145

Russell, J. A., Bachorowski, J.-A., & Fernández-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54*, 329-349. doi: 10.1146/annurev.psych.54.101601.145102

Sato, W., Fujimura, T., Kochiyama, T., & Suzuki, N. (2013). Relationships among facial mimicry, emotional experience, and emotion recognition. *PloS ONE, 8*, e57889.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences of the United States of America, 107*, 2408-2412. doi: 10.1073/pnas.0908239106

Scarantino, A. (2012). How to Define Emotions Scientifically. *Emotion Review, 4*, 358-368. doi: 10.1177/1754073912445810

Scheiner, E., & Fischer, J. (2011). Emotion expression - the evolutionary heritage in the human voice. In W. Welsch, W. Singer & A. Wunder (Eds.), *Interdisciplinary anthropology: The continuing evolution of man.* . Heidelberg & New York: Springer.

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2002). Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of nfants. *Journal of Voice, 16*, 509-529. doi: 10.1016/s0892-1997(02)00127-3

Scheiner, E., Hammerschmidt, K., Jürgens, U., & Zwirner, P. (2004). The influence of hearing impairment on preverbal emotional vocalizations of infants. *Folia Phoniatrica et Logopedica, 56*, 27-40.

Scherer, K. R. (1984). On the nature and function of emotion: a component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293-318). Hillsdale, NJ: Erlbaum.

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

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

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition & Emotion, 23*, 1307-1351.

Scherer, K. R. (2013). Vocal markers of emotion: Comparing induction and acting elicitation. *Computer Speech & Language, 27*, 40-58. doi: 10.1016/j.csl.2011.11.003

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

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15*, 123-148. doi: 10.1007/bf00995674

Scherer, K. R., & Bänziger, T. (2010). On the use of actor portrayals in research on the emotional expression. In K. R. Scherer, T. Bänziger & E. Roesch (Eds.), *A blueprint for an affectively competent agent: Cross-fertilization between emotion psychology, affective neuroscience, and affective computing* (pp. 166-176). Oxford: Oxford University Press.

Scherer, K. R., Clark-Polner, E., & Mortillaro, M. (2011). In the eye of the beholder? Universality and cultural specificity in the expression and perception of emotion. *International Journal of Psychology, 46*, 401-435. doi: 10.1080/00207594.2011.626049

Scherer, K. R., & Ellgring, H. (2007). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion, 7*, 113-130. doi: 10.1037/1528-3542.7.1.113

Scherer, U., Helfrich, H., & Scherer, K. R. (1980). Internal push or external pull? Determinants of paralinguistic behavior. In H. Giles, P. Robinson & P. Smith (Eds.), *Language: Social psychological perspectives* (pp. 279-282). Oxford, England: Pergamon Press.

Scherer, K. R., & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology, 66*, 310-328.

Schirmer, A., Chen, C. B., Ching, A., Tan, L., & Hong, R. Y. (2013). Vocal emotions influence verbal memory: neural correlates and interindividual differences. *Cognitve Affective & Behavioral Neuroscience, 13*, 80-93. doi: 10.3758/s13415-012-0132-8

Schmid, P. C., & Schmid Mast, M. (2010). Mood effects on emotion recognition. *Motivation and Emotion, 34*(3), 288-292. doi: 10.1007/s11031-010-9170-0

Schmidt, K. L., & Cohn, J. F. (2001). Human facial expressions as adaptations: Evolutionary questions in facial expression research. *American Journal of Physical Anthropology, 116*, 3-24. doi: 10.1002/ajpa.20001

Schrader, L., & Hammerschmidt, K. (1997). Computer-aided analysis of acoustic parameters in animal vocalisations: A multi-parametric approach. *Bioacoustics - The International Journal of Animal Sound and its Recordings, 7*, 247-265.

Schuller, B., Batliner, A., Steidl, S., & Seppi, D. (2011). Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication, 53*, 1062-1087. doi: 10.1016/j.specom.2011.01.011

Shariff, A. F., & Tracy, J. L. (2011). What are emotion expressions for? *Current Directions in Psychological Science, 20*, 395-399. doi: 10.1177/0963721411424739

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience and Biobehavioral Reviews, 30*, 855-863. doi: 10.1016/j.neubiorev.2006.06.011

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences, 1156*, 81-96. doi: 10.1111/j.1749-6632.2009.04418.x

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., & Frith, C. D. (2006). Empathic neural responses are modulated by the perceived fairness of others. *Nature, 439*, 466-469. doi: 10.1038/nature04271

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & D., F. C. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*, 1157-1162. doi: 10.1126/science.1093535

Sirois, S., & Brisson, J. (2014). Pupillometry. *WIREs Congnitive Sciences*. doi: 10.1002/wcs.1323

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48*, 813-838.

Smith, J. E. K. (1982). Recognition models evaluated: A commentary on Keren and Baggen. *Perception and Psychophysics, 31*, 183-189.

Sneddon, I., McKeown, G., McRorie, M., & Vukicevic, T. (2011). Cross-cultural patterns in dynamic ratings of positive and negative natural emotional behaviour. *PLoS ONE, 6*, e14679. doi: 10.1371/journal.pone.0014679

Sobin, C., & Alpert, M. (1999). Emotion in speech: The Acoustic attributes of fear, anger, sadness and joy. *Journal of Psycholinguistic research, 28*, 347-365.

Sonnby-Borgström, M. (2002). Automatic mimicry reactions as related to differences in emotional empathy. *Scandinavian Journal of Psychology, 43*, 433-443. doi: 10.1111/1467-9450.00312

Spackman, M. P., Brown, B. L., & Otto, S. (2009). Do emotions have distinct vocal profiles? A study of idiographic patterns of expression. *Cognition & Emotion, 23*, 1565-1588. doi: 10.1080/02699930802536268

Stanislavskij, K. S. (1989). *An actor prepares*. New York: Routledge.

Stanners, R. F., Coulter, M., Sweet, A. W., & Murphy, P. (1979). The pupillary response as an indicator of arousal and cognition. *Motivation and Emotion, 3*, 319-340.

Stel, M., & van Knippenberg, A. (2008). The role of facial mimicry in the recognition of affect. *Psychological Science, 19*, 984-985. doi: 10.1111/j.1467-9280.2008.02188.x

Stel, M., & Vonk, R. (2009). Empathizing via mimicry depends on whether emotional expressions are seen as real. *European Psychologist, 14*, 342-350. doi: 10.1027/1016-9040.14.4.342

Strasberg, L. (1987). *A dream of passion: The development of the Method*. Boston: Little Brown.

Sturtz, S., Ligges, U., & Gelman, A. (2005). R2WinBUGS: A package for running WinBUGS from R. *J Stat Softw 12*, 1-16.

Tamietto, M., & de Gelder, B. (2010). Neural bases of non-conscious perception of emotional signals. *Nature Reviews Neuroscience, 11*, 697-709.

Tanner, J., Wilson, P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61*, 401-409.

Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception and Psychophysics, 27*, 24-27.

Trimbitas, O., Lin, Y., & D., C. K. (2007). Arta de a cere scuze in cultura romaneasca: Use of apology in ethinic Romanian culture. *Human Communication, 10*, 401-420.

Thibault, P., Bourgeois, P., & Hess, U. (2006). The effect of group-identification on emotion recognition: The case of cats and basketball players. *Journal of Experimental Social Psychology, 42*, 676-683. doi: 10.1016/j.jesp.2005.10.006

Todd, A. R., Hanko, K., Galinsky, A. D., & Mussweiler, T. (2011). When focusing on differences leads to similar perspectives. *Psychological Science, 22*, 134-141. doi: 10.1177/0956797610392929

Van Bezooijen, R., Otto, S. A., & Heenan, T. A. (1983). Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *Journal of Cross-Cultural Psychology, 14*, 387-406. doi: 10.1177/0022002183014004001

Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion, 7*, 487-494. doi: 10.1037/1528-3542.7.3.487

van Lankveld, J. J., & Smulders, F. T. (2008). The effect of visual sexual content on the event-related potential. *Biol Psychol, 79*, 200-208. doi: 10.1016/j.biopsycho.2008.04.016

Vandenbergh, S. G. (1972). Assortative mating, or who marries whom. *Behavior Genetics, 2*, 127-157.

Velten, E. (1968). A laboratory task for induction of mood states. *Behaviour Research and Therapy, 6*, 473-482.

Völlm, B. A., Taylor, A. N. W., Richardson, P., Corcoran, R., Stirling, J., McKie, S., . . . Elliott, R. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *Neuroimage, 29*, 90-98.

Vrana, S. R., & Gross, D. (2004). Reactions to facial expressions: effects of social context and speech anxiety on responses to neutral, anger, and joy expressions. *Biological Psychology, 66*, 63-78. doi: 10.1016/j.biopsycho.2003.07.004

Vrana, S. R., & Rollock, D. (1998). Physiological response to a minimal social encounter: Effects of gender, ethnicity, and social context. *Psychophysiology, 35*, 462-469.

Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior, 17*, 3-28.

Walla, P., & Panksepp, J. (2013). Neuroimaging helps to clarify brain affective processing without necessarily clarifying emotions. In F. Fountas (Ed.), Winchester: InTch. doi: 10.5772/51761

Walton, G. M., Cohen, G. L., Cwir, D., & Spencer, S. J. (2012). Mere belonging: The power of social connections. *Interpersonal Reactions and Group Processes, 102*, 512-532.

Wambacq, I. J., & Jerger, J. F. (2004). Processing of affective prosody and lexical-semantics in spoken utterances as differentiated by event-related potentials. *Cognitive Brain Research, 20*(3), 427-437.

Wangelin, B. C., Bradley, M. M., Kastner, A., & Lang, P. J. (2012). Affective engagement for facial expressions and emotional scenes: the influence of social anxiety. *Biological Psychology, 91*, 103-110. doi: 10.1016/j.biopsycho.2012.05.002

Warren, G., Schertler, E., & Bull, P. (2008). Detecting deception from emotional and unemotional cues. *Journal of Nonverbal Behavior, 33*, 59-69. doi: 10.1007/s10919-008-0057-7

Weisbuch, M., & Ambady, N. (2008). Affective divergence: Automatic responses to others' emotions depend on group membership. *Journal of Personality and Social Psychology, 95*, 1063-1079. doi: 10.1037/a0011993

Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of Social Psychology, 26*, 557-580.

Wheeler, B. C., Searcy, W. A., Christiansen, M. H., Corballis, M. C., Fischer, J., Grüter, C., . . . Wild, M. (2011). Communication. In R. Menzel & J. Fischer (Eds.), *Animal thinking: Contemporary issues in comparative cognition* (pp. 187-205). Cambridge, MA: MIT Press.

Wicker, B., Keysers, C., Plailly, J., Royet, J.-P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron, 4*, 655-664.

Wild, B., Erb, M., & Bartels, M. (2001). Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: quality, quantity, time course and gender differences. *Psychiatry Research, 102*, 109-124.

Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America, 52*, 1238-1250.

Wilting, J., Krahmer, E., & Swerts, M. (2006). *Real vs. acted emotional speech*. Paper presented at the INTERSPEECH-2006, Pittburgh PA, USA.

Young, S. G., & Hugenberg, K. (2010). Mere social categorization modulates identification of facial expressions of emotion. *Journal of Personality and Social Psychology, 99*, 964-977. doi: 10.1037/a0020400

Zahavi, D. (2008). Simulation, projection and empathy. *Consciousness and Cognition, 17*, 514-522. doi: 10.1016/j.concog.2008.03.010

Zhang, F., & Parmley, M. (2010). What your best friend see that I don't see: Comparing female close friends and casual aquaintances on the perception of emotional facial expressions of varying intensities. *Personality & Social Psychology Bulletin, 37*, 28-39.

Zuckerman, M., Koestner, R., Colella, M. J., & Alton, A. O. (1984). Anchoring in the detection of deception and leakage. *Journal of Personality and Social Psychology, 47*, 301-311.

Zuckerman, M., Koester, R., & Colella, M. J. (1985). Learning to detect deception from three communication channels. *Journal of Nonverbal Behavior, 9*, 188-194.

7 Acknowledgments

First of all my gratitude goes to Julia Fischer, who made this thesis possible and who accompanied me on this - sometimes quite stony - path. You gave me an academic home but also encouraged me to go my own way. Thank you for your confidence, your support and for enlightening my thoughts whenever necessary.

Second I would like to express my special thanks to Annekathrin Schacht, without your enthusiasm and your ideas this thesis would have never been finished. Thank you for integrating me so spontaneously in your group and for supporting my research and my academic development.

My gratitude goes to Hannes Rakoczy, for co-supervising this thesis as well as to Margarete Boos, Igor Kagan and Bernhard Fink for agreeing to be on my thesis committee. There were other people, who mentored me on my way, and I would like to express my thanks to Julian Klein, who showed me what can be accomplished with enthusiasm and a pressing deadline, and to Jens Scheiner, who reminded me to ask the right questions to myself.

Working and feeling at home in two research groups enlarges the list of people I have to thank extensively. I would like to thank all coggies, for great coffee breaks on the roof, funny evenings in St. Andreasberg, and for all the lunch breaks. Although thematically an outsider, I never felt like one. Thanks for that. I already miss seeing the baboon distribution map twice a month. Special thanks goes to Laura (you know why!), Chris (for knowing everything, and for helping every time), Ludwig (for always providing me with technical stuff), Kurt (for all the acoustic knowledge, and all the wine), Adeelia, Philip, and Max (for helping in times of need), Rasmus (for being a friend) and Mechthild (for explaining me the Dienstreiseantrag again and again).

I would like to thank all “Psycholinguists” for forming such a great research group. I felt like home, the moment I entered your lab. You broadened my view with your linguistic-psychological-philological-neuroscientific backgrounds. Specifically, I would like to thank Annika (for being such a wonderful office neighbor), Anna, Lena and Billa (for supporting me in the lab and during data cleaning), Mareike (for quick email responses, and analytical support) and Sarah, Olga, and Katrin (for giving me such a great welcome). I will miss you all!

There are a lot of other people (too many to mention here), who supported me with relaxing coffee and lunch breaks, who encouraged or distracted me whatever was necessary and who just made my days. Thank you all! I would like to express my special thanks to Ivy, the best friend I could image; to Anna, my voice of reason who supplied me with self-made food when needed most urgently, and to Justus, Peter and Bob who made my sleepless nights bearable. Additionally, I thank my family for their support. My biggest gratitude goes to Ingo. Words cannot express my thoughts and my feelings, so I better do not try. Thank you for all!