SERIES IN AFFECTIVE SCIENCE

Evolution of EMOTIONAL COMMUNICATION

From Sounds in Nonhuman Mammals to Speech and Music in Man

EDITED BY
Eckart Altenmüller
Sabine Schmidt
Elke Zimmermann

OXFORD

Evolution of Emotional Communication

Series in Affective Science

Series Editors
Richard J. Davidson
Paul Ekman
Klaus Scherer

The Evolution of Emotional Communication Eckart Altenmüller, Sabine Schmidt, and Elke Zimmnermann (eds.)

The Neuropsychology of Emotion John C. Borod

Persons, Situation, and Emotions: An Ecological Approach Herman Brandstätter and Andrzej Eliasz

Handbook of Emotion Elicitation and Assessment James A. Coan and John J.B. Ellen (eds.)

Anxiety, Depression, and Emotion Richard J. Davidson

What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS) 2e Paul Ekman and Erika L. Rosenberg (eds.)

The Nature of Emotion: Fundamental Questions Paul Ekman and Richard J. Davidson

The Psychology of Gratitude Robert A. Emmons and Michael E. McCullough (eds.)

Who Needs Emotions? The brain meets the robot Jean-Marc Fellous and Michael A. Arbib (eds.)

Emotions in Psychopathology: Theory and Research William F. Flack and James D. Laird (eds.)

Shame: Interpersonal Behaviour, Psychopathology, and Culture Paul Gilbert and Bernice Andrews (eds.)

Pleasures of the Brain Martin L. Kringelbach and Kent C. Berridge

Infant Chimpanzee and Human Child: A Classic 1935 Comparative Study of Ape Emotions and Intelligence N.N. Ladygina-Kohts (deceased) and Frans B.M. de Waal (eds.) Boris Vekker (translator)

Feelings: The Perception of Self James D. Laird

Cognitive Neuroscience of Emotions Richard D. Lane and Lynn Nadel (eds.)

The Development of Social Engagement: Neurobiological Perspectives Peter J. Marshall and Nathan A. Fox (eds.)

Science of Emotional Intelligence: Knowns and Unknowns Gerald Matthews, Moshe Zeidner, and Richard D. Roberts (eds.)

Affective Neuroscience: The Foundations of Human and Animal Emotions Jaak Panskepp

Nonverbal Behaviour in Clinical Settings Pierre Philippot, Robert S. Feldman, and Erik J. Coats (eds.)

Emotion in Memory and Development: Biological, Cognitive, and Social Considerations Jodi Quas and Robyn Fivush (eds).

Memory and Emotion Daniel Reisberg and Paula Hertel (eds.)

Emotion Explained Edmund T. Rolls

Emotion, Social Relationships, and Health Carol D. Ryff and Burton Singer (eds.)

Oxford Companion to Emotion and the Affective Sciences David Sander and Klaus Scherer

A Blueprint for Affective Computing: A sourcebook and manual Klaus R. Scherer, Tanja Bänzinger, and Etienne Roesch

Appraisal Processes in Emotion: Theory, Methods, Research K. Scherer, A. Schorr, and T. Johnstone (eds.)

Evolution of Emotional Communication

From Sounds in Nonhuman Mammals to Speech and Music in Man

Edited by

Eckart Altenmüller

University of Music, Drama and Media, Hanover, Germany

Sabine Schmidt

University of Veterinary Medicine, Hanover, Germany

Elke Zimmermann

University of Veterinary Medicine, Hanover, Germany

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First edition published in 2013

Impression: 1

British Library Cataloguing in Publication Data
Data available

ISBN 978–0–19–958356–0

Library of Congress Control Number: 2012945092

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

Contents

Contributors

Eckart Altenmüller
Institute of Music Physiology and Musicians' Medicine, University of Music, Drama and Media Hanover, Hanover, Germany

Kai Alter
Newcastle University, Newcastle, UK

Jorge L. Armony
International Laboratory for Brain, Music and Sound research (BRAMS), University of Montreal, Montreal; Douglas Mental Health University Institute and Department of Psychiatry, McGill University, Montreal, QC, Canada

William Aubé
International Laboratory for Brain, Music and Sound research (BRAMS), University of Montreal, Montreal, QC, Canada

Jo-Anne Bachorowski
Department of Psychology, Vanderbilt University, Nashville, TN, USA

Stefan M. Brudzynski
Department of Psychology, Brock University, St. Catharines, ON, Canada

Reinhard Dengler
Department of Neurology, Medical School, Hanover, Hanover, Germany

Günter Ehret
Institute of Neurobiology, University of Ulm, Ulm, Germany

W. Tecumseh Fitch
Department of Cognitive Biology, University of Vienna, Austria

Thomas Fritz
Max Planck Institute for Human Cognitive and Brain Science, Leipzig, Germany; Institute for Psychoacoustics and Electronic Music, Ghent, Belgium

Sascha Frühholz
Neuroscience of Emotion and Affective Dynamics (NEAD) Laboratory, Department of Psychology, University of Geneva; Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

Didier Grandjean
Neuroscience of Emotion and Affective Dynamics (NEAD) Laboratory, Department of Psychology, University of Geneva; Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

Oliver Grewe
University of Music, Drama and Media Hanover, Hanover, Germany

Anna S. Hasting
Max Planck Institute for Human Cognitive and Brain Sciences, Neurocognition of Rhythm in Communication Group, Leipzig, Germany

Patrik N. Juslin
Music Psychology Group, Uppsala University, Uppsala, Sweden

Stefan Koelsch
Cluster of Excellence "Languages of Emotion," Freie Universität, Berlin, Germany

Reinhard Kopiez
University of Music, Drama and Media Hanover, Hanover, Germany

Sonja A. Kotz
Max Planck Institute for Human Cognitive and Brain Sciences, Neurocognition of Rhythm in Communication Group, Leipzig, Germany

Lisette Leliveld
Institute of Zoology, University of Veterinary Medicine Hanover, Hanover, Germany

Marc Mehu
Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

Marcello Mortillaro
Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

Michael J. Owren
Department of Psychology, Georgia State University, Atlanta, GA, USA

Silke Paulmann
Department of Psychology, University of Essex, Colchester, UK

Isabelle Peretz
International Laboratory for Brain, Music and Sound research (BRAMS), University of Montreal, Montreal, QC, Canada

Michael Philipp
Department of Psychology, University of Queensland, Brisbane, QLD, Australia

Josef P. Rauschecker
Department of Neuroscience, Georgetown University School of Medicine, Washington, DC, USA

Simone Schehka
Institute of Zoology, University of Veterinary Medicine Hanover, Hanover, Germany

Klaus R. Scherer
Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland

Sabine Schmidt
Institute of Zoology, University of Veterinary Medicine Hanover, Hanover, Germany

Christine Schröder
Department of Neurology, Medical School Hanover, Hanover, Germany

Allison Schulman
Department of Psychology, Cornell University, Ithaca, NY, USA

Charles T. Snowdon
Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA

Joseph Soltis
Department of Education and Science, Disney's Animal Kingdom,® Lake Buena Vista, FL, USA

Diana P. Szameitat
University of Tübingen, Tübingen, Germany

David Teie
School of Music, University of Maryland, College Park, MD, USA

Niyati Trivedi
Department of Psychology, Cornell University, Ithaca, NY, USA

Eric Vanman
Department of Psychology, University of Queensland, Brisbane, QLD, Australia

Dirk Wildgruber
University of Tübingen, Tübingen, Germany

Philip Sanford Zeskind
Department of Pediatrics, Levine Children's Hospital at Carolinas Medical Center, Charlotte, NC, USA
and Department of Pediatrics, University of North Carolina at Chapel Hill, NC, USA

Elke Zimmermann
Institute of Zoology, University of Veterinary Medicine Hanover, Hanover, Germany

Klaus Zuberbühler
University of Neuchâtel, Cognitive Science Centre, Neuchâtel, Switzerland

Evolution of emotional communication: An introduction

Eckart Altenmüller, Sabine Schmidt, and
Elke Zimmermann

Emotions are integral decision mechanisms in the brain of humans and animals. They are universal in mammals, originally evolved for governing fitness-relevant behavioral and physiological responses towards a particular stimulus or situation. Emotional brain assessment mechanisms are dependent on motivation and vary according to homeostatic needs, for example, feelings of hunger, coldness, or sexual urges. They may be either solely inherited or additionally modified by experience and, in humans, by memory and tradition. Emotions can be measured on the behavioral level by the type of a behavioral response, reflected, for example, in vocalizations, bodily displays, facial mimics, gestures, or simply in avoidance-approach tendencies. Furthermore, emotions are characterized by changes in intensity of responses toward a specific stimulus or situation. These intensity changes are reflected in different levels of arousal, which in turn can be objectively assessed by measuring reactions of the autonomous nervous system, for example, heart rate, blood pressure, piloerection, or epinephrine (adrenaline) secretion. As further means to gain objective information on emotions, neural circuits and neurohormonal correlates can be investigated and specific brain representations can be visualized with modern imaging methods. In humans, self-reports may provide additional important insights into the nature, quality, and time course of emotions.

Communication of emotions is crucial for organizing all aspects of social life. To name but a few, parent–offspring relations, partnership and reproduction, competition for resources, establishment of group coherence and hierarchies, coalition formation and cooperation, and finally—in agonistic contexts—signaling threat or readiness for defense rely on efficient communication of emotions. Signaling may occur in all sensory domains, e.g., via acoustic, visual, tactile, olfactory, and gustatory modalities.

As a research topic, communication of emotions is important for many fields such as animal behavior, anthropology, evolutionary biology, human psychology, linguistics, musicology, neurology, and psychiatry. So far, in ethology and psychology, most research has been conducted on the role of emotion in visual communication, concentrating on observation of behaviors such as facial displays, postures, or piloerection. Research on the role of emotions in acoustic communication and its evolution has been neglected for a long time. Recent developments of technical tools such as digital acoustic signal analysis, neuropharmacology, and neuroimaging as well as new theoretical approaches together with current comparative studies of nonhuman mammals and humans are now providing new and surprising insights into this area of emotion research. The nineteen contributions within this book examine manifold facets of emotional communication by acoustic signals. Our aim was to provide a collection of state of the art articles considering a broad range of mammalian taxa and a broad range of affective signaling, encompassing affect bursts, emotional signaling in animals, laughter, infant cry, prosody in healthy and diseased humans, and music as the most acculturated way of conveying emotions. Furthermore, we strove to integrate the data presented in the book in order to address the central and yet unresolved question whether universal principles rule the encoding and decoding of emotions from animals to man. Therefore, by analysing shared and unique principles we attempted to reconstruct evolutionary pathways for emotions in the acoustic domain, for example, when considering the strong emotions whilst listening to music and experiencing chill responses and shivers down the spine.

We organized the book in five parts and added a summarizing chapter at the end. The first, introductory part sets the ground for the volume and emphasizes the evolutionary pathways of acoustic communication and its neurophysiological basis in three review chapters. Mortillaro et al. start with a theoretical outline, defining emotions as dynamic episodes, which are characterized by a high degree of coordination between the components elaborated in Scherer's Component Process Model. The model's fundamental assumption is that organisms constantly scan and evaluate their environment and the incoming stimuli. Appraisal of these stimuli—stimulus evaluation checks or SECs—determine their relevance, implications, and ability to cope with them. Particularly significant events give rise to reactions involving functional changes in subsystems of the organism, such as the facial, vocal, and bodily motor systems. A constitutive feature of emotion is the synchronization of response channels, for example, in the visual-gestural and the acoustic domain. Since this is an essential aspect of spontaneous affect expressions in animals and humans, it may point towards the evolutionary origin of emotional expressions.

Fitch and Zuberbühler explore in their chapter the evolution of human language using a comparative approach. The language ability can be subdivided into several components, namely the ability to produce, perceive, and learn complex signals and the ability to interpret and control these signals depending on current context and prior knowledge. When comparing these components with primate vocalizations and their anatomical and neurophysiological prerequisites, the authors come to interesting conclusions: Primate social intelligence and context-dependent signal interpretation involve multiple cognitive mechanisms that are homologous with those employed in human semantics and pragmatics. In contrast, there is little evidence for homology between human learned vocalization, such as language or song, and primate vocalizations. Accordingly, primate calls are better understood as homologs of our innate species-typical vocalizations, such as laughs, cries, and screams that every human child is born with. Furthermore, human syntax goes far beyond the combinatorial complexity so far observed in nonhuman primate communication and therefore seems to be a unique feature for humans developed recently in evolution.

The last chapter in this section by Rauschecker opens a different perspective based on comparing the functional anatomy and neurophysiology of auditory processing hierarchies in primates and humans. One of these processing streams, extending from auditory core areas in the middle superior temporal cortex rostrally and ventrally, is involved in the decoding and identification of complex auditory patterns or "objects." Species-specific communication sounds, including speech in humans, are one such category of complex sounds. Accordingly, their decoding happens in this anterior-ventral stream. Categorization of these sounds occurs most probably in ventrolateral prefrontal cortex. The superior temporal cortex gives rise to another projection stream into medial prefrontal areas, where the emotional context of vocalizations is processed. Neurophysiological analysis of this emotional-acoustic network of the primate brain, including the anterior insula and anterior cingulate cortex, in addition to the amygdala, nucleus accumbens, and other subcortical stations, is still not well understood. The medial prefrontal network ultimately projects to the hypothalamus, through which it contributes to emotional actions and reactions and to the experience of emotions as such.

Part B focuses on a comparative approach on the role of emotions in acoustic communication in various mammalian taxa. Similarities and differences are highlighted and eventual "candidates" for universal components of affective signaling are discussed. In the first article of this section, Ehret starts with emotional communication in mouse pups, which produce six acoustically different types of sounds. These sounds express by their frequency bandwidths and noisiness emotions of fear, submissiveness, distress, and comfort. They are perceived according to their basic emotional meanings, since adult mice specifically respond to them. Fear and submissiveness is perceived as attraction, distress as aversion, and comfort as cohesion.

Subsequently, Brudzynski, opens the complex fields of neuroanatomy and neurotransmitter systems related to emotional vocalizations in rats and cats. The ascending cholinergic system can initiate a negative state with accompanying autonomic and somatic symptoms, and the ascending dopaminergic system initiates a positive state also with characteristic physiological changes. These emotional states are signaled by vocalization to conspecifics. Negative, alarming/threatening vocalizations are initiated by release of acetylcholine in the medial cholinoceptive vocalization strip, originating from hypothalamic nuclei. The vocalizations are characterized by a low and relatively constant sound frequency and long duration of individual calls. Positive, appetitive calls are initiated by release of dopamine in the shell of the nucleus accumbens, and at least in rats, are characterized by high and often modulated sound frequency and short duration of individual calls. High repeatability of relevant acoustic features of positive or negative vocalizations makes them useful indicators of animal emotional states.

Schmidt addresses sound emission and perception during social interactions in bats. Since bats are a highly vocal group evolutionarily remote from primates, they are of particular interest for a comparative approach to emotional acoustic communication. In a model species, *Megaderma lyra*, attractive, cohesive, or aversive behaviors are characterized. Distinct call types, typically consisting of several syllables of similar or different frequency-time contours occur in these specific behavioral situations. Call types of similar syllable structure emitted in different situations are compared with respect to emotional context. In addition, the variability in call structure is related to affect intensity. In sum, calls emitted by bats during social interactions may reflect the emotional relevance of the situation.

Soltis widens the perspective to African elephants and reviews the evidence for vocal expression of their emotions. In doing so, he also speculates as to how complex vocal communication systems in elephants and other animals could have evolved from simple vocal expressions of basic emotions.

Zimmermann and coauthors continue in this line and explore a "prehuman origin hypothesis of affective prosody" using a comparative approach with nonhuman mammals that focuses on the intensity component of affective prosody. Based on their own work and an extensive literature review, they extract four general behavioral contexts in which empirical data on the expression of acoustically conveyed affect intensity can be compared across the most-studied mammalian orders. These contexts are alarm/disturbance, agonism, social isolation, and affiliation. Vocalizations in eight taxonomic orders of mammals (Primates, Scandentia, Chiroptera, Rodentia, Carnivora, Artiodactyla, Proboscidea, and Cetacea) are analyzed. Although these groups diverge greatly in size, ecological niche, habitat, and social system, comparative data reveal that affect intensity is similarly encoded across contexts and species. Furthermore, acoustically conveyed affect intensity is of significant relevance for listeners in five of the mammalian orders studied. This work supports the hypothesis that affective prosody in human acoustic communication has deep-reaching phylogenetic roots, deriving from precursors already present and relevant in the vocal communication systems of nonhuman mammals.

The last contribution of this section by Snowdon and Teie is of particular interest with respect to a comparative approach concerning the biological roots of music. The authors argue that, although human music is unique to our species, the emotional aspects of music have a long evolutionary history. They show that music/emotional features can be seen in the vocalizations of cotton-top tamarins, a small monkey from the New World. In a second step, they composed music comprising their hypothesized emotional features in the frequency range and tempos of tamarin vocalizations, and played these compositions to the monkeys. This way they could induce calming and arousing emotional states in the monkeys, although human-based music with similar features had little or no effect on the monkeys. Their results suggest that other species incorporate similar emotional features into their calls as are used in music to influence emotions. The results also suggest that calls are not merely communicating the caller's emotional state but may be used to induce emotional responses in listeners.

Part C outlines affective signaling in nonverbal human communication including infant cry and laughter. Zeskind examines the communicative and emotional significance of the cry of the newborn and young infant within a conceptual framework that focuses on a synchrony of arousal between infant and caregiver behavioral systems. He demonstrates that infant crying has its basis in anatomical, physiological, and neurobehavioral mechanisms that produce nonspecific changes in the intensity of infant arousal, which are reflected in a graded and dynamic acoustic signal comprised of a complex array of spectral and temporal characteristics. These graded cry sounds in turn affect the intensity of the arousal system of the caregiver in a synchronous graded manner. However, changes in the intensity of caregiver arousal are mediated by the receiver's subjective affective state, thus providing the basis for the same cry sound to elicit different responses from different caregivers. These basic elements of emotional communication are further illustrated by examining responses to the accentuated sounds of hyperphonation, an unusually high-pitched cry sound found in infants at risk for poor intellectual and social development, including physical child abuse.

Owren and colleagues focus on different aspects of human laughter. Human laughter can trigger affect in listeners, including both positive and negative emotions. In a series of experiments the authors demonstrate that voiced laughs are rated positively while unvoiced versions are not. These results are mirrored in automatic evaluation by listeners and in felt emotion, thus providing evidence that listeners can experience both conscious and unconscious emotional responses upon hearing laughter. The authors further elaborate the view that listener reactions develop as learned emotional responses to laugh sounds. According to them, associative positive emotional value is likely accruing more readily to voiced than to unvoiced versions.

In a similar vein, Szameitat and colleagues suggest that laughter is primarily a signal for regulating group structure. Based on their experiments on emotional valence ratings of different types of laughter, they hypothesize that on one hand laughter may integrate other group members and thus reinforce within-group relations, e.g., in the form of joyous laughter as a greeting. On the other hand, laughter may segregate others by rejecting or excluding individuals from group context, e.g., in the form of taunting laughter in a mobbing situation. Concerning ontogeny, these two forms of laughter seem to emerge consecutively during child development. Initially, when laughter functions as a reward for the carer, it aims exclusively at integrating the child into the social group. From a certain age onward, the child starts to exercise group hierarchy, e.g., through certain forms of play such as unruly play and mock fighting. Such play often imitates domineering and aggressive encounters, i.e., behavior that would in the everyday social context pose a threat to the social partner.

Part D is dedicated to human prosody and its alteration in disorders of the frontal lobe and the basal ganglia. Grandjean and Frühholz propose a specific neuroanatomical and functional network involved in decoding of emotional prosody in humans which is based on brain imaging and lesion studies. This model includes a first pathway of acoustical and emotional analysis of auditory stimuli in the primary and non-primary auditory cortex in collaboration with the subcortical gray nuclei and the amygdala, respectively. This information is then fed forward to anterior brain regions in the medial and lateral inferior frontal cortex, where elaborative stimulus evaluations take place that guide decision processes and action tendencies in response to the stimulus.

The second contribution of this section by Kotz and colleagues provides a comprehensive overview of the functional roles of the orbitofrontal cortex and the basal ganglia in the processing of vocal emotional expressions in humans. The authors suggest that the two structures are critically involved in the integration and evaluation of vocal emotional expressions. Furthermore, the orbitofrontal cortex is functionally differentiated and may respond to emotional expressions during both early automatic and late controlled, evaluative processing stages. Based on comparative neuroanatomical evidence, the authors speculate that the orbitofrontal-basal ganglia network may be involved in the processing of emotional vocalizations across different species.

In the final chapter of this section, Juslin delivers a scholarly analysis of the problems and promises of research into vocal affect expression. He concludes with a tentative agenda for future research, broadening our perspectives by using a wider range of emotion labels based on theoretical analyses. Juslin is in favor of obtaining natural speech samples in field studies that use experience sampling methods and ambulatory physiological measurement. Interestingly, he proposes to investigate instances of "mixed emotions," which reflect combinations of spontaneous and posed emotion.

Part E highlights the role of music as a powerful means to communicate emotions, its neurobiological foundations, and factors contributing to strong emotional responses in humans. Peretz and coauthors present current knowledge about musical emotions from a neuropsychological perspective. They provide evidence that musical emotions depend on a specialized emotional pathway that may recruit various subcortical and cortical structures that might be shared, at least in part, with other biologically important systems. This emotional pathway is not simple. There is not a single, unitary emotional system underlying all emotional responses to music. For instance, most of the neural pathway underlying emotional responses to dissonance has been delineated and involves a complex and distributed system in the brain.

Fritz and Koelsch focus on different aspects of music and evolution. They review two traits that have previously been discussed as possible homologous traits of the human capacity for music passed on to us by a great ape common ancestor: drumming and song. Furthermore they point out another possible homology of the human capacity for music: a neurological mechanism for acoustically mediated emotional contagion. According to cross-cultural research, contagious emotional expression in music exist as a universal feature. The authors then discuss how emotional contagious behavior in our closest phylogenetic relatives, such as play panting and pant-hoots in African great apes, are structurally and functionally related to laughter and music in humans. They conclude that the musical capacity in humans may have evolved as a response to selective pressures for increased group size for its effect of synchronizing group motivation and emotional experience through emotional contagion, and as such would have promoted group gatherings, social functions, and the establishment of rituals.

In the final article of this section, Altenmüller and colleagues discuss the adaptational value of the chill response to music linked to strong emotions, feelings of pleasure, and nostalgia.

They argue that emotions induced by music refer to different evolutionary origins. Aesthetic emotions may have developed relatively late in human evolution potentially in the context of the invention of the first musical instruments some 35,000 years ago. In contrast, strong emotions such as chill responses to music are linked to an activation of the sympathetic nervous system and the brain reward circuits. Chill responses occur in the presence of novel and unexpected musical events. Furthermore, they depend on individual listening biographies and personality factors. Since chill responses support memory consolidation, the authors speculate that they may have reinforced the development of human auditory perception and fine-tuned auditory pattern recognition. Finally a hypothetical "mixed origins of music" theory is presented: Early roots of music may lie in an ancient affective signaling system, common to many socially living mammals. Later, music was further developed; it induced aesthetic emotions and provided a safe playground for auditory learning in general and promoted social cohesion and well-being.

In the summarizing chapter in Part E, the editors of this book propose an ethological framework by which acoustically conveyed emotions could be explored experimentally across nonhuman mammals and humans. They develop an individualized, context-related, concept of emotion for which fitness-related social scenarios have to be defined to determine the emotion of an individual in a given context. Based on this framework, they review the evidence for universal features common to humans and nonhuman mammals provided in the contributions of this book. Furthermore they disclose unique species-specific adaptations, and shed light on selective factors shaping the evolution of speech and music. However, in the future, further systematic comparative analyses are needed to disentangle universal from taxa-specific components in emotional prosody and the respective processing networks in the brain.

Finally, we thank the many contributors of this volume, not only by providing their articles, but also by cross-reviewing the work of their colleagues and thus improving quality and depth of thoughts. We also would like to thank the generous support of Oxford University Press, specifically of Martin Baum and Charlotte Green. Furthermore, we thank the German Research Foundation for financing the research group FOR 499 "Acoustic communication of emotions in nonhuman mammals and man." All members of this research group contributed to this volume, and furthermore were engaged in stimulating discussions, which were the starting point for the present work.

Part A

Introductory Section

This page intentionally left blank

Chapter 1

The evolutionary origin of multimodal synchronization and emotional expression

Marcello Mortillaro, Marc Mehu, and Klaus R. Scherer

Abstract

Emotions are defined as dynamic episodes characterized by a high degree of coordination between several organismic subsystems in the interest of optimal adaptation to relevant events. We argue that the constitutive feature of emotions—synchronization of response channels—is linked to the evolutionary origin of emotional expressions, by suggesting that it is an essential aspect of spontaneous affect expressions in animals and humans. The significance of this mechanism to the signaling function in emotional expression communication is explored. In particular, the importance of the perception of a sender's authenticity in speech communication and music is highlighted. These theoretical notions are then exemplified with a concrete illustration from ongoing research.

What exactly is communicated in emotional communication?

We hold that research on "emotional communication" should start with a clear notion of the nature and function of emotion, because in order to understand the mechanisms of communication, we need to define the content. Unfortunately, so far there is no agreed upon definition of emotion, even in the central domain of the psychology of emotion (Frijda 2007a; Scherer 2005). However, there is some convergence on at least some central elements of emotion which Frijda and Scherer (2009) have summarized in four points. First, an emotion occurs only in consequence to an event that is deemed relevant for the organism's needs, goals, values, and general well-being. The degree of relevance of the event is concretely determined by how the event is appraised on a number of stimulus evaluation criteria (e.g., if it is appraised as novel rather than expected, if it is intrinsically pleasant or unpleasant, and if it is motivationally consistent or inconsistent; Scherer 2001). Second, emotions evolved as adaptive mechanisms because they have a strong motivational force and prepare the organism to action (*action readiness*; Frijda 2007b). Third, emotions urge the person to act or to suspend action and thus they are accompanied by appropriate preparatory regulation of somatovisceral and motor systems. The functioning of these systems tends to cohere during emotion episodes and becomes synchronized at different levels (Scherer 2005). Fourth, because of the related emotion, action readiness states have control precedence and thus demand (successfully or not) priority in the control of behavior and experience (Frijda 2007b).

According to Frijda and Scherer (2009) these four features define what an emotion is for both researchers and lay people and they distinguish emotions from other affective states (e.g., moods). Emotion, thus, can be defined as a brief, delimited time period or episode in the life of an organism when there is an emergent pattern of synchronization between several components (and related bodily subsystems) in the preparation of adaptive responses to relevant events as defined by their behavioral meaning (Scherer 2001).

Appraisal-driven functional subsystem coherence as a central defining criterion

In this contribution we will focus on the central feature of the coherence of bodily subsystems and consequent synchronization of emotional response channels, with special emphasis on the relevance of this synchronization phenomenon for the evolution of motor expression and emotional communication. In order to explain the fundamental assumption underlying this notion, we will briefly describe Scherer's Component Process Model (CPM) of emotion (Scherer 2001, 2009).

The model's fundamental assumption is that organisms constantly scan and evaluate their environment, and particularly significant events give rise to a process of reactive adaptation that involves functional changes in most, if not all, subsystems of the organism (Fig. 1.1). The CPM entails a functional approach—as first explicitly suggested by Darwin (1872/1998)—for which emotion prepares action readiness and different action tendencies, though without necessarily enacting them (Frijda 2007b). For example, emotions like anger or fear prepare the body for executing emotional behaviors such as aggression or flight, but emotion is not a sufficient condition for their execution. This functional account which is implicit in many emotion theories has been explicitly developed by Nesse (2009) in his discussion of the evolution of emotion.

The CPM describes five functional components of the theoretical construct "emotion" (the five "rows" of Fig. 1.1): (1) appraisal processes, i.e., the evaluation of the stimulus; (2) autonomic physiology or physiological arousal, i.e., the regulation of the organism through the activation of the sympathetic and parasympathetic nervous systems; (3) action tendencies, i.e., the physiological and behavioral preparation towards action; (4) motor expression, i.e., the communication of reaction and behavioral intention in terms of facial, vocal, and bodily expression; (5) subjective feeling, i.e., the monitoring of internal state and organism–environment interaction through the conscious experience of the emotion state (see Scherer 2005, 2009 for an overview).

Components and related bodily subsystems are assumed to function in a highly interdependent fashion through multidirectional physiological mechanisms of feedback and feedforward (e.g., Gellhorn 1964). Recent studies support this functional interdependence: for example, Bonanno and Keltner (2004) identified associations between facial expression (motor expression component) and appraisal themes (cognitive component) and showed that coherence among systems raises the level of experienced emotion (subjective feeling component).

In the case of emotion it seems reasonable that the origin of a recursive chain of changes should be located in the information processing system, i.e., the appraisal of the event (Arnold 1960). According to the CPM, organisms continuously appraise events and their consequences on a number of evaluation checks producing a motivational effect. This occurs on multiple levels of processing—from automatic sensorimotor or schematic to effortful representational levels (Leventhal and Scherer 1987)—which explains why many appraisal mechanisms can also be demonstrated in animals or babies (Désiré et al. 2004; Scherer et al. 2004). The basic principle of the theory is that the appraisals (conscious or not) of a relevant event produce efferent effects in all subsystems and thus in all the other components.

The CPM details four main groups of appraisals or *stimulus evaluation checks* (SECs) organized in a fixed sequence (see the upper part of Fig. 1.1): (1) relevance of the event (e.g., is the event unexpected? Is the event intrinsically pleasant?); (2) implications for major needs, goals, and values; (e.g., is the event conducive to reaching individual's goals?); (3) ability to deal with event's consequences or coping potential (e.g., do I have the power to change the event?); and (4) normative significance of the event (e.g., is the event compatible with my standards?). Each SEC has efferent effects on all emotion subsystems (descending arrows in Fig. 1.1) so that each SEC produces changes that cumulate on changes produced by earlier SECs.

The appraisals and the motivational change will directly cause efferent effects in the autonomic physiology component (i.e., in the autonomic nervous system, for example, in the form of cardiovascular changes) and on the motor expression component (i.e., in the somatic nervous system); these modifications are represented in the central nervous system and fused in a multimodal integration area that is continuously updated. When this central integrated representation (or parts of it) becomes conscious to the person (subjective feeling component), then it can be emotionally labeled with emotion words, linguistic expressions, or metaphors (Scherer 2001, 2005, 2009).

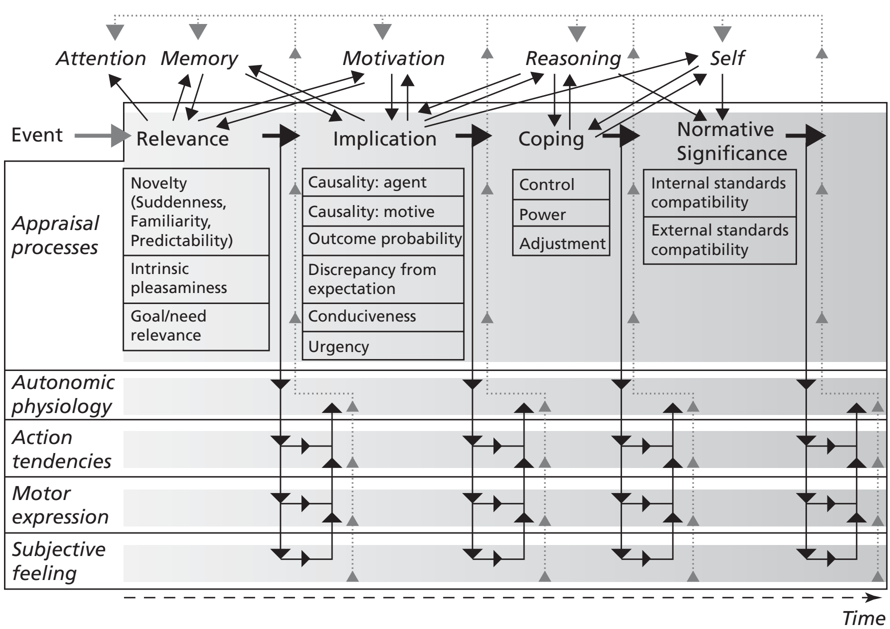

Figure 1.1 Graphical representation of the Component Process Model. The five components are listed vertically on the left of the picture. The appraisal processes are organized in four subsequent groups of stimulus (event) evaluation checks: Relevance, Implication, Coping, and Normative Significance. The appraisal component triggers the emotion episodes and has efferent effects on all the other components (descending arrows). These components have reciprocal effects between them and feedback effects on the appraisal processes (ascending arrows). Reproduced from Scherer, K.R., Appraisal considered as a process of multilevel sequential checking, in Sherer, K.R., Schorr, A., and Johnstone, T. (eds.) *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 92–120. © 2001, Oxford University Press, with permission.

While there is a wide agreement among appraisal theorists on most of these evaluation checks, the sequential hypothesis is more controversial (Roseman and Smith 2001).[1] The existence of a fixed sequence implies that the pattern of changes in the components is specific to the particular sequence of modifications produced by the succession of SECs. Each SEC result has an added value in a complex sequential interaction, so that any specific pattern of component states can occur if there is a corresponding pattern of SEC results (see Scherer 2001, 2009).[2]

Motor expression as signal in emotional communication

The CPM includes detailed predictions about the changes in all subsystems due to each specific SEC result (Scherer 2009). These predictions reflect the functional approach of the CPM, in terms of both general functions of emotion and specific function of each SEC. It is important to highlight that for social species the adaptive response enacted by emotion entails motor action for both instrumental and social purposes, i.e., interaction and communication. In this chapter, we will explore the model predictions that concern the function of communication of reaction and behavioral intention, which centrally concerns emotional expression in face, voice, and gestures. Further discussion of the predictions as well as pertinent empirical findings can be found for facial expression (Kaiser and Wehrle 2001; Dael, Mortillaro, & Scherer, 2012; Mortillaro, Mehu, & Scherer, 2011; Scherer 1992; Scherer and Ellgring 2007a; Smith and Scott 1997; van Reekum et al. 2004; Wehrle et al. 2000) and vocal expression (Banse and Scherer 1996; Goudbeek and Scherer 2010; Johnstone et al. 2001; Juslin and Scherer 2005). An important premise is that only approximate predictions can be made because the activation of muscles involved in the emotion expression does not depend on emotion only but on many different factors, particularly the prevalent goal states of the organism and the environmental demands at the specific time.

Predictions for facial and vocal expressions of individual SEC outcomes are reported in Table 1.1. The sequence of appraisal outcomes is extremely variable so that the cumulative effects of each SEC determine innumerable complex response patterns. Nevertheless, there are some emotions that are characterized by frequently recurrent patterns of appraisals—such as anger, fear, and joy—which have been called modal emotions (Scherer 1994a). Predictions about the expression of these emotions can be formulated on the basis of their theoretically predicted appraisal profiles (see Table 5.4 in Scherer 2001) and the assumptions concerning the efferent effects of each SEC (Table 1.1). Examples of predictions about expressive configurations for some modal emotions are reported in Table 1.2 (for further details, see Scherer 2001, 2009).

The efferent effects of the various appraisal checks affect all expressive modalities—voice, face, and body movements—determining multimodal coherence and synchronization between expressive channels. Although multimodality is a key element for emotion communication for both production and recognition, unfortunately, as most research groups have specialized in a

[1] Recent evidence on mental chronography and on odor processing showed that some SECs are usually processed earlier (and faster) than others. In particular, there is strong experimental evidence that the stimulus is first processed for its novelty and then for its pleasantness, thus supporting the hypothesis that at least some appraisals are organized in a fixed sequence (Delplanque et al. 2009; Grandjean and Scherer 2008).

[2] This principle is also at the basis of the patterns which seem to characterize emotions like anger and fear. For this reason, generally, componential theorists do not endorse the idea that there are some basic emotions, but rather a high number of differentiated emotions: some of these emotions occur more frequently because of the high frequency of certain sequences of SEC results (Scherer 2001, 2009).

Table 1.1 CPM's predictions for facial and vocal expressions of individual SEC outcomes

Appraisal dimension	SEC outcome	Facial expression	Vocal expression
Novelty	Novel	Brows up, lids up; or brows lowered, jaw drop, open mouth, open nostrils, gaze directed	Interruption of phonation, ingressive (fricative) sound with glottal stop (noise-like spectrum)
	Not novel	No change	No change
Intrinsic pleasantness	Pleasant	Lids up, jaw drop, open mouth, open nostrils; or lip corners pulled upwards, lips parted, gaze directed	Faucal and pharyngeal expansion, relaxation of tract walls, vocal tract shortened due to lip corners pulled upwards (wide voice)
	Unpleasant	Brows lowered, lid tightened, eye closed, nose wrinkling, upper lip raised, lip corner depression, chin raised, lip press, nostril compression; or lower lip depressed, tongue thrust, lips parted, jaw drop; gaze aversion	Faucal and pharyngeal constriction, tensing of tract walls, vocal tract shortened due to lip corner depression (narrow voice)
Goal conduciveness	Relevant and consistent	Relaxation of facial muscle tone	Relaxation of vocal apparatus (relaxed voice)
	Relevant and discrepant	Brows lowered, lids tightened, lips tightened, chin raising; gaze directed	Tensing of vocal apparatus (tense voice)
Coping potential	No control	Hypotonus of facial musculature, lip corner depression, lips parted, jaw drop, lids droop, eyes closed; if tears, inner brow raised, brows lowered, gaze aversion	Hypotonus of vocal apparatus (lax voice)
	Control and high power	Brows lowered, lids up; or lids tightened, lips tightened and parted; or lips tightened and pressed together, nostril dilatation; stare	Chest register in phonation (full voice)
	Control and low power	Brows up, lids up, jaw drop, mouth stretched and corner retraction, nostril dilatation, switching between gaze direction and aversion	Head register in phonation (thin voice)
Norms	Respect	Elements of pleasantness and high power response	Elements of pleasantness and high power response
	Violation	Elements of unpleasantness and low power response	Elements of unpleasantness and low power response

Sources: Data from Scherer, K.R. (2001). Appraisal considered as a process of multilevel sequential checking, in K.R. Scherer, A. Schorr, and T. Johnstone (ed.) *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 92–120, New York, NY: Oxford University Press. Scherer, K.R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, 23, 1307–51.

Table 1.2 CPM's predictions of representative expressive features for some modal emotions

Behavior	**ENJ/ HAP**	**ELA/ JOY**	**DISP/ DISG**	**SAD/ DEJ**	**DESP**	**ANX/ WOR**	**FEAR**	**IRR/ COA**	**RAGE/ HOA**
Inner brow raiser		>		>	>>		>>		
Outer brow raiser		>					>>		
Brow lowerer			>	>	>>	>	>	>	>>
Upper lid raiser	>	>				>	>		>
Cheek raiser	>	>							
Nose wrinkler			>						
Upper lip raiser			>						>
Lip corner puller	>	>>							
Lip corner depressor			>	>	>>				
Chin raiser			>		>			>	>>
F0 Perturbation	<=	>		>	>		>		>
F0 Mean	<	>	>	<>	>		>>	><	><
F0 Variability	<	>		<	>		>>	<	>>
F1 Mean	<	<	>	>	>	>	>	>	>
F2 Mean			<	<	<	<	<	<	<
Intensity mean	<	<	>	<<	>		>	>	>>
Intensity variability	<	<		<			>		>
High-frequency energy	<	<>	>	><	>>	>	>>	>>	>>
Speech rate	<	>		<	>		>>		>

> indicates increase; (>) indicates potential increase; double symbols indicate the predicted strength of the change; joint use of two symbols pointing in the opposite direction refer to cases where antecedent voice types exert influence in opposite direction. ANX/WOR = Anxiety/Worry; DISP/DISG = Displeasure/Disgust; ELA/JOY = Elation/Joy; ENJ/HAP = Enjoyment/Happiness; IRR/COA = Irritation/Cold anger; RAGE/HOA = Rage/Hot anger; SAD/DEJ = Sadness/Dejection.

particular modality, mostly the face, there are very few empirical data available for *multimodal* expression of emotion (but see Bänziger, Mortillaro, & Scherer, 2011; Hess et al. 1988; Pell 2005; Scherer and Ellgring 2007b; van den Stock et al. 2007). Furthermore, multimodal research is extremely challenging for both theoretical and methodological reasons: it requires precise specification of the hypothesized production mechanisms and implies working with different kinds of data. A recent effort to adopt a multimodal perspective in studying emotion expression was made by Scherer and Ellgring (2007b) who analyzed a set of acted emotion portrayals looking for the existence of stable multimodal configurations and their degree of emotion specificity. Facial expressions were coded using the Facial Action Coding System (Ekman and Friesen 1978), body movements, postures, and gestures were coded using the coding scheme developed by Wallbott (1998), and several acoustic parameters were computed. By means of a cluster analysis, the authors found three multimodal patterns of behavior. The first cluster, labeled Multimodal Agitation, grouped elements indicative of an aroused emotional state (e.g., arms stretched sideways, mouth stretched, high fundamental frequency, and high amplitude).

The second cluster consisted of elements typical of positive emotional states (e.g., smiles, fast speech rate) and of surprise (e.g., jaw drop), so it was labeled Multimodal Joyful Surprise. The third multimodal cluster—called Multimodal Resignation—combined vocal parameters of low arousal (e.g., low amplitude, low fundamental frequency), facial movements like eyelids drop, a collapsed body posture and self manipulation gestures. These multimodal patterns could be meaningfully interpreted on the basis of the appraisal checks that are supposedly at the origin of the emotions expressed through those patterns. Thus, emotions portrayed with behaviors belonging to the multimodal agitation cluster (e.g., hot anger, panic fear) were characterized by an appraisal of high urgency; conversely, emotions portrayed with behaviors belonging to the multimodal resignation cluster (e.g., sadness, boredom) seemed characterized by an appraisal of loss of control, suggesting a probable reason for the absence of action preparation.

Perception and the Brunswikian lens model

Expression is one half of communication, perception is the other, and the two must always be considered together when studying any communication process. Accordingly, different authors have consistently suggested the Brunswikian lens model (Fig. 1.2) to study the process of emotional communication (Juslin and Laukka 2003; Juslin and Scherer 2005; Scherer 2003). The original purpose

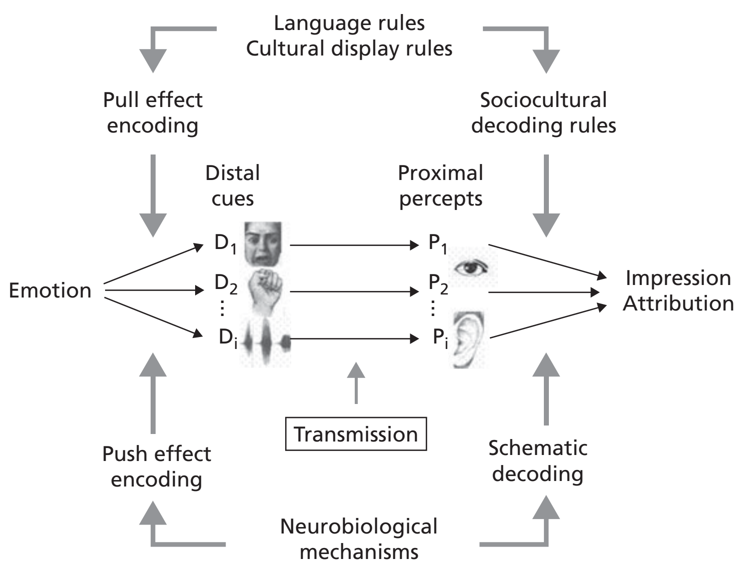

Figure 1.2 Modified version of the Brunswikian Lens Model. Starting from the left of the picture, an emotion is encoded by the sender through a number of multimodal cues which are the result of push effects (due to neurobiological mechanisms) and pull effects (due to language and cultural rules). These distal cues are transmitted through a channel that, along with the positioning of the receptors, determines how similar to the original distal cues the proximal percepts will be. The receiver employs sociocultural and schematic rules to decode the proximal cues and to form an impression and attribute an emotion intention to the sender. Reproduced from Scherer, K.R. and Bänziger, T., On the use of actor portrayals in research on emotional expression, in Scherer, K.T., Bänziger, T., and Roesch, E.B. (eds.) *Blueprint for Affective Computing: A Sourcebook*, pp. 166–76. © 2010, Oxford University Press, with permission.

of Brunswik's Lens Model was to frame the study of the visual perception process (Brunswik 1956; Hammond and Stewart 2001) but its central principles can be almost readily applied to any form of communication, including the multimodal emotion communication process.

According to the lens model, the emotion communication process begins when the sender expresses an emotion state by means of several cues—likely pertaining to different expressive modalities—which are labeled distal cues (i.e., they are remote from the observer). These distal cues are transmitted through a channel and they are perceived by the observer in a more or less modified form as proximal cues (percepts): their degree of similarity to the distal cues depends on the quality of the transmission channel and of the sensory organs. The observer has access only to these proximal cues and based on them he or she probabilistically attributes an emotion intention to the sender. Scherer (1985) suggested that distal cues are the product of the interaction of *push* and *pull* effects.

Push effects are motor response patterns resulting from physiological changes and from the preparation of instrumental motor actions in consequence of information processing activity and behavioral preparation. In other words, push effects refer to expressive cues that are the direct by-products of neurobiological mechanisms that affect the expressive motor system. Three instrumental functions of the facial organs (lips, nose, ears) and the vocal tract (mouth, pharynx, larynx) contribute to push effects on expression: (1) transferring matter (air, liquid, light) from the body to the environment and vice versa (e.g., in the service of nutrition); (2) searching for optimal reception of stimulation (e.g., wide opening eyelids); and (3) acting directly on objects and other organisms (e.g., biting). Push effects usually have a rapid onset and are direct and uncontrolled externalizations of internal processes. Examples of expressions exclusively due to push effects are affect bursts (i.e., brief, discrete, sudden expressions as a consequence of emotionally charged events; Krumhuber & Scherer, 2011, Scherer 1994b) or infant grunts. Push effects are supposed to occur universally but their concrete appearances are relatively idiosyncratic, and thus subject to individual differences.

Pull effects are expressive configurations that are part of a socially shared communication code and so they are sociocommunicative signals used to inform, or influence other group members. Individuals learn through socialization to employ specific patterns of responses for communicating effectively, or deceptively, internal states and behavioral intentions to other people. In this sense, pull effects exclusively refer to cultural and linguistic rules. Examples of pure pull effects are merely conventionalized emotion expressions, symbolic coding, and affect emblems (similar to visual emblems; Ekman and Friesen 1969, 1972). As a consequence of their highly conventionalized forms, pull effects show little interindividual variations: these responses can be decoded effectively only if they respect social rules and adhere to the fixed socially shared symbolic code.

Production mechanisms

Pure push or pull expressions are extremely rare. Instead, we hypothesize that several mechanisms jointly operate to determine the final expression, that can be mapped on a continuum with pure push and pure pull effects at the extremes (Table 1.3, first column).

- First, emotion expression may be partly determined by reaction-specific *hard-wired neuro-motor program commands*, which are at the basis of push effects. In reaction to specific stimuli these adaptive neuro-motor programs automatically activate all neuro-biological subsystems in a very abrupt way (Izard 2007), leading to highly synchronized patterns of multimodal responses (e.g., affect bursts; Scherer 1994b).

Table 1.3 Mechanisms involved in emotion expression production and perception

Production mechanisms	Perception mechanisms
Hard-wired neuro-motor programs	Hard-wired patterns of feature detection
Appraisal-driven responses	Inference of underlying appraisals
Response regulation	Detection of regulation strategies. Inference of intention
Symbolic signaling	Schematic decoding of symbolic meaning
	Motor-mimicry
	Appraisal of the contextual information

- Second, changes in different response modalities are produced by the *complete appraisal* of the emotion eliciting event. Adaptive sequential and cumulative modifications in the states of the biological subsystems occur as a result of the information processing activity and resulting behavioral tendency. These modifications are reflected in expressive behavior in the form of appraisal-driven changes that appear gradually (via a sequence of cumulative changes) and in a synchronized fashion (coherence between modalities at each step of the appraisal sequence) in all the different expressive modalities (Scherer 2001, 2009).
- Third, the expresser can voluntarily manipulate the expression by applying emotion *regulation strategies*. Regulation strategies aim at modifying or suppressing the expression or parts of it in order to control the intention that is perceived; for achieving this result, regulation strategies modify the course of emotion expression likely introducing perturbations in the temporal unfolding of the appraisal-driven expressive changes (de-synchronization).
- Fourth, the expression is consciously and voluntarily produced by the sender to communicate a specific *symbolic meaning* or intention. The expresser makes use of vocal, facial, or bodily expressions that correspond to symbolic conventional signals that can be understood by the observer on the basis of a socially shared communication code. This latter mechanism is the source of pure pull effects, which are especially visible in vocal, facial, or bodily emblems (Ekman and Friesen 1969, 1972; Scherer 1994b).

These four mechanisms are at the basis of any emotion communicative behavior either in isolation or combined. However, there are only very few cases in which a single mechanism is at work; typically, all these mechanisms jointly contribute to the resulting emotional expression which is practically always multidetermined.

Perception mechanisms

The complexity of the production process has its counterpart in the multiplicity of perception mechanisms that the receiver employs to infer an emotion intention and attribute an emotion state to the expresser (Table 1.3, second column). Emotion perception happens on the basis of the proximal cues which are processed by the receiver by means of varying combinations of different recognition mechanisms—which can be invoked automatically or (in some cases) voluntarily.

- First, hard-wired neuro-motor production commands may be accompanied by *hard-wired detection mechanisms*. This is an innate form of recognition shaped by evolution that the perceiver employs automatically. It can be considered the human analog of animal instinctive mechanisms for releasing species-specific behaviors in response to appropriate stimulation (innate releasing mechanism)—e.g., the courtship behavior of sticklebacks described by Tinbergen (1951, 1952).

- Second, the receiver can execute a rudimentary *reflex-like process of motor mimicry* of fixed nonverbal expression patterns. The receiver unconsciously simulates part of the expression of the sender, and uses the proprioceptive information that becomes available to infer the emotion that is expressed (Chartrand and Bargh 1999; Lipps 1907; Niedenthal 2007).
- Third, the perceiver can also employ *controlled cognitive decoding* processes aimed at inferring, from the behavior, the information processing activity and behavioral tendencies of the sender. In other words, the receiver uses the nonverbal information to infer the underlying appraisals and consequently he or she attributes the emotion to the sender (Scherer and Grandjean 2008).
- Fourth, the perceiver uses the *sociocultural communication code* that he or she has in common with the sender to interpret the emotion intention. The pull effects are decoded by the perceiver through a schema-based strategy which is learnt in the process of socialization.
- Fifth, the perceiver analyzes the likelihood that *regulation strategies* have been adopted by the sender to mask, attenuate, or exaggerate the emotion intention. In other words, the perceiver evaluates whether the final expression results from deliberate efforts by the sender to modify the expression driven by push effects, or whether the expression was strategically produced on the basis of pull factors.
- Finally, the perceiver can evaluate the expression based on situation, task, needs, or goals, as well as interpersonal *context*.

These mechanisms constitute an impressive battery of strategies available to the perceiver for decoding the emotion expression and inferring the sender's intention. On the other side, this array of mechanisms illustrates the complexity of the emotion communication process (on both sides of production and recognition) and why it cannot be reduced to a simple one-to-one correspondence between one expression and one meaning.

The automaticity of synchronized expression

Adaptive behavioral responses require coordination at the endocrine, physiological, and motor level, and emotional processes are believed to achieve such coordination (Scherer 2009). Coordination at the motor level involves the synchronization of different muscle systems and results in specific behavioral patterns aimed at modifying physical environmental conditions (like running away from a danger) or acting on a receiver's perceptual system (social signaling). The coordination of the different muscles can have effects on multiple expressive modalities producing multimodal signals. Multimodality has evolved to make signals more efficient, for example, by adapting to the constraints imposed by transmission in variable physical environments and by the receiver's psychology (Guilford and Dawkins 1991; Rowe 1999). As argued earlier, multimodal synchronization is a fundamental aspect of emotional expression and it could originate in the need of the organism to rapidly prepare coordinated behavioral readiness.

Although the pairing of sound and facial expression evolved to produce more efficient signals, all associations one can observe between the two might not have the same signaling function. While a vocalization is generally accompanied by facial movements (sound production usually requires a particular configuration of mouth, jaw, and lips), the reverse is not true, as facial expressions can be produced silently as unimodal signals. The combination of vocalization and facial expression may therefore change the structure of the signal and increase the repertoire of vocalizations. On the other hand, the addition of sound to a facial expression may have different consequences such as making the visual signal more conspicuous to receivers and facilitate the localization of the signaler. The temporal association between a vocalization and a facial expression could also tell us whether the two components are redundant or not (Partan and Marler 1999). For example, when the vocalization comes first it could function to locate and direct attention to the signaler. In a synchronized multimodal signal, the components could provide backups to each other and not convey different information. Sequential components could transfer different types of information whereas simultaneous components could be redundant and enhance each other's effect, or simply provide backup.

An example of emotional expression that involves the integration and synchronization of visual and auditory modalities is the *affect burst* (Scherer 1994b). Affect bursts are due to push effects and are considered some of the most basic instances of facial and vocal expression integration that human beings have in common with many other mammalian species (Scherer 1994b). These expressions are commonly made of brief affective vocalizations accompanied by specific patterns of facial expressions, and they are considered as mostly automatic and spontaneous behaviors.

Automaticity in behavior is to be contrasted with voluntary actions, which depend on the ability to form internal representation of desired goal states, to predict the future effects of one's actions and to select actions based on these anticipated consequences (Goschke 2003). The distinction between spontaneously produced expressions and intentional displays is a crucial issue in research on emotional communication. Although both types of signals serve a communicative function, the difference between the two may lie in the amount of conscious and strategic planning inherent to these signals. The question is further complicated by the fact that an expression typically involves both automatic and controlled production mechanisms. Although spontaneous and intentional expressions may have a similar form, dynamic features are certainly important to disentangle these two aspects—for example, deliberate facial expressions have more irregularities and involve shorter offsets than spontaneous facial expressions (Ekman and Friesen 1982; Hess and Kleck 1990). Cognitive control imposed on an expression may indeed disturb the time course of its components and therefore appear more inconsistent than spontaneous expressions. Another difference that distinguishes spontaneous versus intentional expressions is the synchronized timing of their multimodal components. It is indeed unlikely that conscious processes are capable of timely coordinating the production of the subtle muscular movements involved in the production of sound and facial expression.

The idea that synchronized expressions are not under conscious behavioral control is corroborated by the observation that different neurobiological pathways are involved in the control of spontaneous and voluntary expressions. Spontaneous facial behavior originates in a phylogenetically older motor system known as the extrapyramidal circuits, mostly located in subcortical areas (Rinn 1984). Volitional motor behavior is usually controlled by cortical regions—the supplementary motor cortex and the cingulate gyrus—because lesions in these regions suppress voluntary movement while electrical stimulation results in coordinated movement (Halgren and Marinkovic 1995). Rinn (1984) also reported, on the basis of neurological disorders, that voluntary facial behavior is controlled by the cortical motor control areas through the pyramidal system. The double dissociation between voluntary and spontaneous, emotional, movements was demonstrated by the fact that either kind of movement can be disrupted by neurological damage while the other stays intact (DeJong 1979). In the vocal domain, spontaneous and voluntary expressions may also be controlled by separate mechanisms. Clinical studies suggest that vocal and articulatory speech control can be impaired separately (Aronson 1990, cited in Davis et al. 1996), as emotional reactions with a complete or partial loss of voice were observed in patients who nevertheless maintained the ability to articulate or to use a whispered form of speech.

The separation between spontaneous and voluntary expressions should not imply that these two types of reactions always occur in a distinct fashion. It is unlikely that expressive motor behavior will be entirely volitional, or entirely spontaneous, as, functionally, there are more mechanisms jointly operating at the basis of emotional expression (see the discussion on push and pull effects in the 'Perception and the Brunswikian lens model' section) and, physiologically, there could be coupling between the limbic system (responsible for spontaneous, emotional expressions) and cortical motor control areas (Hutton 1972).

Automaticity produces honest signals: A source for the impression of authenticity

For the reasons already discussed in this chapter, synchronization is inherent to integrated emotional processes. Synchronization at the expressive level (synchronization of the different channels involved in the expressive component of emotion) could reflect the activity, or be under the control of, other emotion components. Synchronization could therefore constitute a "trademark" of emotional processes and, at the level of expression, a guarantee of authenticity. The absence of voluntary control on the production of synchronized expressions would therefore prevent deception based on strategic and conscious planning. In this sense automaticity and synchronized expressions could represent honest signals of internal states hence, indirectly, of future behavior.

The impression of authenticity about a person results from an evaluation by the perceiver that the information derived from that person can be used to take adaptive decisions. Such information is available through two categories of indicators: cues and signals. Cues are features of the environment that have not necessarily evolved for communication purposes but that nonetheless convey information. In other words, a social cue is an opportunity for receivers to extract information about an individual, possibly with the aid of contextual indicators and some basic rules about the relationships between these indicators. The second source of information about individuals is social signals. Social signals are actions or structures that evolved as a result of their effect on other organisms (Dawkins and Krebs 1978; Maynard Smith and Harper 2003). Even though signals might not always contain information, or act in the benefit of the receiver (Owren and Bachorowski 2003; Owren and Rendall 1997), signaling is expected to be reliable whenever the costs of deceptive signals exceed the benefits of providing honest cues.

Synchronization of the different channels of emotional expression could be a by-product of the synchronization of emotional processes and may not serve a communicative function. Synchronization would thus be considered as a cue that authentically reflects emotional processes. This, however, requires a one-to-one relationship between the experience of an emotion and behavioral expression, implying that each time an emotion is experienced, a consistent expression occurs. Different authors argued that this is not the case for most emotion expression in everyday interactions where the social context plays an essential role (Fridlund 1991; Russell and Fernadez-Dols 1997; Scherer 1992), and, furthermore, the unrestricted availability of emotional processes to others could be a disadvantage in groups where individuals have different interests and make use of deceptive tactics.

The importance of the social context and the neurobiological specialization in the production and perception of multimodal emotional expressions (Adolphs 2002; Hess et al. 1995; Rinn 1984) strongly suggests that synchronization between channels of the expressive component could have a signaling function. The reliability of multimodal synchronized emotional expressions could be ensured by either one of three factors: strategic costs (handicap), mutual goals between sender and receiver (as in minimal-costs signals), and physical association between the signal and the advertised attribute (index; Maynard Smith and Harper 2003). It is not expected that all emotional expressions enter the same category, since their evolution may result from different selective pressures.

First, the signal is reliable because it is too costly for individuals who do not possess the underlying quality that the signal is meant to advertise. This is also called the "handicap principle" (Zahavi 1975; Zahavi and Zahavi 1997). Costs can relate to signal production (sender-related costs) or to receiver's aversive response to the signal (receiver-imposed costs, e.g., punishment or exploitation). In the case of emotional expressions, the signal is related to the underlying state of the individual and is reliable because it can only be produced when an emotion is experienced. This perspective implies that emotional aspects of communicative behavior could be considered as added costs. Here, the word "added" means that the costs of emotional expression are not incurred to transfer the information accurately but to ensure the reliability of the signal. For example, emotional expression such as the Duchenne smile has been considered an honest signal of altruistic dispositions because it involves a physiological component that could constitute added cost to signal production (Brown and Moore 2002; Mehu et al. 2007). Recent research showed that the activation of facial muscles that are difficult to control voluntarily leads to increased perceived authenticity of emotional expression (Mehu, Mortillaro, Bänziger, & Scherer, 2012).

Costly signals are usually displayed in situations in which social exploitation is possible, for example when the goals of sender and receiver are in contradiction. The costs of multimodal communication and synchronization are not very well studied. Producing multimodal signals could be more costly if the sender has to increase energy expenditure to supply multiple communication channels. On the other hand, multimodal signals could be as costly as unimodal ones if the same amount of energy invested in signal production is distributed across channels (Partan and Marler 2005). However, costs related to signal detection by eavesdroppers or predators are increased in multimodal signals since they are more conspicuous than unimodal signals (Roberts et al. 2007). Multimodal synchronized signals should therefore be more costly than unimodal signals. The question remains whether these costs participate to making the signal more reliable (strategic costs) or whether they are just costs necessary to convey the information (efficiency costs) (Guilford and Dawkins 1991).

Second, a signal can be expected to be reliable because there is no incentive for a sender to cheat: nothing can be gained by "lying," even if the signal has no cost. This often occurs when signaler and receiver have a common interest, i.e., they assess the possible outcomes of the interaction in the same order of preferences (Maynard Smith and Harper 2003). These signals are also called minimal-cost signals because their cost is as low as possible while at the same time transferring the necessary information (Maynard Smith and Harper 1995). Expressions of low-arousal emotions are usually low in intensity and could enter this category. It would imply that in situations where senders and receivers have common interests they tend to express more low-arousal emotions (e.g., sadness) than what they do in situations where they have conflicting interests.

Third, a signal could be reliable because it cannot possibly be faked. Such signal is typically called an index (Maynard Smith and Harper 1995). Indices are physically related to the attribute that is of interest to the receiver and it is this physical association that makes them reliable. Indices are not costly and demonstrate some quality that cannot be bluffed like, for instance, the capacity to resist parasites (Hamilton and Zuk 1982). A behavior that indicates body size can be considered an index because the signal is directly related to size and cannot be faked. The form of such signals is determined by its content. Examples of indices include the erect posture, like in the pride expression (Tracy and Robins 2004), that has the effect of increasing perceived body size; and some vocal parameters such as speech rate or fundamental frequency could act as indices of mood changes (Ellgring and Scherer 1996). Since emotion is a multicomponent process (Scherer 2005), the expressive component could act as an index of the appraisal component or the physiological component, which would in itself be a cue for future behavior.

An illustration of the effect of synchronization: Perceived spontaneity in singing

As illustrated previously, two determinants of emotional expression are push and pull mechanisms. Push effects are due to internal psychobiological processes that affect the different expressive modalities; as a consequence, we can expect that these multimodal behaviors are highly synchronized since they are produced by the same biologically rooted, hard-wired mechanisms. Pull effects, on the contrary, are conventionalized expressions which are voluntarily controlled and more loosely synchronized. The subsystems' synchronization typical of emotion states is very costly for the organism, because it changes the normal functioning of all subsystems requiring many fundamental physiological activities to modify their course of action to serve the action tendency generated by the emotion. As a consequence, the symbolic representation of an emotion (pull effects) is unlikely to recruit all subsystems and make them work in a synchronized way. More adaptively, only some visible or audible markers should be activated as pull effects, letting all the physiological subsystems continue working for the normal functioning of the organism. Furthermore, even in case of a deliberate attempt at faking synchronization, strategic expressions produced by untrained people should mostly result in artificial and probably less synchronized exemplars than what could be expected for expressions due to push effects. On the basis of this reasoning, it can be argued that the larger the role of push effects, the more synchronized the multimodal expression is likely to be.

Our hypothesis is that, given its adaptive significance, the perceiver detects, consciously or not, the degree of synchronization between the different components of the multimodal signal, and that he or she uses this information to judge the spontaneity of an emotional expression. To investigate this hypothesis we conducted an exploratory study on emotion expression in singing consisting of two parts; first, a rating study to determine perceived spontaneity of the multimodal expression studied, with the hypothesis that judgments of spontaneity based on audiovisual information are not reducible to those based on either facial or vocal information; second, behavioral coding to explore the relationship between the synchronization of different expressive channels and ratings of spontaneity. As we already mentioned, there are very few multimodal studies available in literature, with practically none that include temporal information, which is a necessary feature for studying synchronization. Given the absence of established procedures, the second part of our study was exploratory in nature, mainly aimed at showing the feasibility of the study of synchronization in multimodal expressive behavior.

We decided to investigate our hypothesis in the context of staged operatic singing. On the one hand, stage performances represent a particular condition in which viewers are aware of the fact that they are watching professional actors, but nevertheless react to what they see as if it was not acted. On the other hand, actors employ different techniques for being judged credible and believable in their emotion behavior by the viewers. In other words, in stage performances everybody is aware of the acting framework but, nevertheless, some performers can be judged as more authentic than others, in the sense that they are perceived as feeling at least the rudiments of the emotions that they portray on stage (as expected for those actors who adopt techniques like Stanislavsky or method acting techniques).

We chose two audio-video recordings of two well-known professional tenors performing the same song, “La Danza” (composed by G. Rossini), with clearly different expressive behaviors. One was much more expressive and active than the other in the way of performing: for the sake of clarity we labeled them Dynamic and Static, respectively. The song is the narration of a summer dancing party at the seaside and the general emotion tone of both lyrics and music is of happiness and enjoyment (*tarantella*). From the two performances the same five segments were extracted based on content and technical requirements.[3]

The rating study involved 36 participants recruited at the University of Geneva who were randomly assigned to one of three perceptual conditions: audio-video (AV; n=12), audio only (A; n=12), or video-only (V; n=12).[4] Participants rated the content of each of the ten clips (two singers times five segments) on three continuous dimensions: *Spontaneity*, i.e., “How spontaneous do you think the singer’s behavior was?”; *Agreeableness of the singer*, i.e., “How much did you like the singer in this clip?”; *Agreeableness of the performance*, i.e., “How much did you like this performance?” The two questions about agreeableness were included to control that the judgments of spontaneity were not judgments of “liking.” In other words, we wanted to check that participants rated spontaneity—at least partly—independently of their preference for one tenor. Spontaneity mainly referred to the emotional representation while liking referred to the singing and acting skills as they were appreciated by the person. In addition to these dimensional ratings, participants answered some control questions about their knowledge of opera (“How well do you know opera?,” on a 5-point scale from “not at all” to “very well”), of the song (“Did you already know this song?”), and of the tenors (“Did you already know either one of the two tenors?”).[5]

We analyzed the ratings with respect to the effects due to the two tenors, the different Segment of the song, and the perceptual Condition for the ratings.[6] As it can be seen in Fig. 1.3, the Dynamic

3 The original two audio video files were converted into avi files (divX codec) for their presentation through a computer-based interface programmed in Cogent.

4 The participant sat in front of a 17 inch computer screen and headphones were available when the experimental condition required it. Instructions were provided on both a paper sheet and the computer screen. They could watch and or listen to each segment only once, and they interacted with the interface through mouse and keyboard.

5 The ratings of one participant were excluded from further analysis because they consisted of outlying values in the AV condition and, as the control questions showed, she was the only participant who knew the singers and the song. Thus, this previous knowledge might have influenced her ratings through an established attitude.

6 Three-way mixed-design ANOVAs (analyses of variance) with Segment (5) and Tenor (Dynamic vs. Static) as within-subjects factors and Condition (A vs. V vs. AV) as between-subjects factor were performed on Spontaneity, Agreeableness of singer, and Agreeableness of performance. Concerning *Spontaneity*, we found a main effect of Tenor ($F(1,32) = 9.70$, $p < 0.01$), qualified by an interaction effect of Tenor × Condition ($F(2,32) = 5.28$, $p < 0.01$). Concerning *Agreeableness of Singer*, participants preferred some segments for one tenor, and some others for the other tenor—main effect of Segment ($F(4,128) = 4.02$, $p < 0.01$), which was qualified by the interaction effect Segment × Tenor ($F(4,128) = 2.89$, $p < 0.05$)—and in general they liked AV stimuli most—main effect of Condition ($F(2,32) = 3.70$, $p < 0.05$). Concerning *Agreeableness of Performance* we found only a main effect of Segment ($F(4,128) = 4.27$, $p < 0.01$).

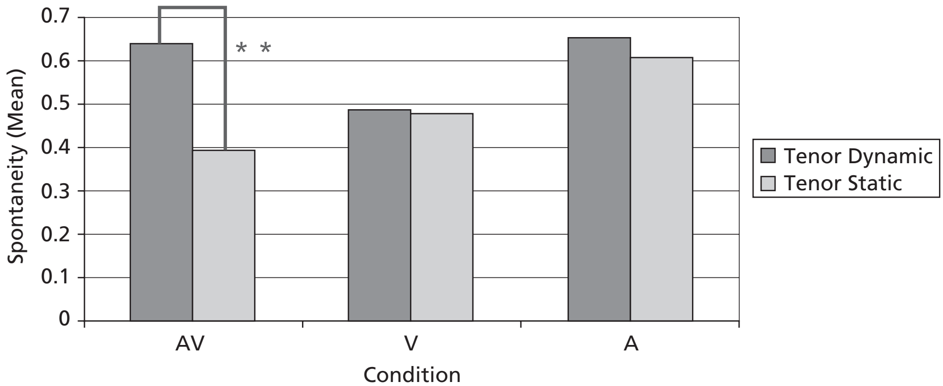

Figure 1.3 Perceived spontaneity for song segments in three perceptual conditions. Mean ratings (n=35) in three perceptual conditions: Audio and Video (AV), Video only (V), and Audio only (A). The Dynamic tenor was rated significantly more spontaneous (**p <0.01) than the Static tenor in the multimodal condition.

tenor was rated more spontaneous than the Static one, but in the AV condition only—the judgments of spontaneity for the two tenors did not differ in either the V or the A condition. Neither the way of singing (audible acoustic characteristics) alone nor the way of acting (e.g., visible gestures and body movements) alone differentiated the two tenors in terms of spontaneity in the eyes of the raters. Rather, it was an interaction of both factors, i.e., the way of using jointly the two expressive modalities that influenced the judgments of the spontaneity of the tenors' performances.

With the two other variables (Agreeableness of the singer and Agreeableness of the performance) we wanted to check that the ratings of spontaneity were not a by-product of a more general preference for one tenor. We found that judges preferred some segments for the Static tenor and others for the Dynamic tenor, and that they liked AV stimuli most. These results indicate that judges did not rate each segment in the same way and that they preferred alternatively one of the two tenors depending on the segment: there was no systematic preference for one of the two tenors. Furthermore, even though judges preferred the performance when it was shown with both audio and video information, this preference was not depending on the tenor. All in all, these results confirmed that the interaction effect between Tenor and Condition on the judgments of spontaneity was not due to a general preference for one singer in the AV condition. Rather, the multimodal information had a crucial influence on judgments of spontaneity.

Our hypothesis was that synchronization between expressive modalities was the main reason why judges rated the Dynamic tenor as being more spontaneous than the other. Clearly, the available material did not allow any direct test of the hypothesized link between synchronization and judgments of spontaneity; nevertheless, we could check whether the singer who was rated more spontaneous in the rating study, showed more synchronization between the different modalities.

For this purpose, we analyzed vocal and facial behavior. Given the exploratory nature of this study, we decided to focus only on one particular multimodal behavior pattern that may be used to convey a salient symbolic meaning but may also be used without any voluntary control: eyebrow raise and accents in speech. Eyebrows are salient features of almost any facial expression (including emotion expressions) and one of the most important communication devices, serving several pragmatic purposes (Ekman 1979). An abundant literature showed their relevance for emotional communication, and some studies in particular found that raising the eyebrows is used to signal prominence of certain elements of the speech (Ekman 1979; Flecha-Garcia 2010; House et al. 2001). At the same time it has been shown that a speaker frequently raises the eyebrows without any voluntary control or awareness (Rinn 1991). We wanted to analyze how the two singers used this facial movement with respect to their vocal expressions. According to our hypothesis, indeed, when an "accent" in face appears in a synchronous way with an "accent" in voice, the behavior will be judged more spontaneous, because it is implicitly seen as a push effect. On the contrary, when the eyebrow raise movement does not co-occur with a vocal accent it would be perceived as voluntarily used for conveying a symbolic meaning, so is implicitly seen as a pull effect.

Video files were FACS (Facial Action Coding System) coded continuously for the presence of the raising the eyebrows movement (AU1, inner brow raising, and AU2, outer brow raising; Ekman and Friesen 1978).[7] In terms of prosodic properties, we extracted the pitch contour for each file (fundamental frequency, F0), using the "Praat" software (Boersma and Weenink 1996). The target facial movement (eyebrow raise) was operationalized as the simultaneous presence of AU1 and AU2 at their apexes (i.e., eyebrows are held raised). At the same time we needed to define what an accent in speech would be. We assumed that some elements of a vocal expression may be particularly apt for signaling the prominence of specific pieces of information: Start of vocalization, end of vocalization, and pitch accents. i.e., peaks in the fundamental frequency contour (recent findings corroborate this assumption; Flecha-Garcia 2010).

Segments during which AU1 and AU2 were simultaneously present at their apexes were isolated and became our units of analysis (n =46; 25 for the Dynamic tenor), fundamental frequency contours were drawn and the segment was labeled synchronous if either one of the following criteria was met (see Fig. 1.4): (1) The beginning of *AU1+AU2 apex* coincides with the beginning of a vocalization; (2) the end of *AU1+AU2 apex* coincides with the end of a vocalization; (3) either the beginning or the end of *AU1+AU2 apex* corresponds to an accent in speech (a peak in the fundamental frequency contour).

Figure 1.5 shows the distribution of these segments (synchronous vs. nonsynchronous) between the two tenors: as predicted, the Dynamic tenor showed a significantly higher proportion of synchronous segments than the Static tenor (respectively, 15 out of 25, and 6 out of 21).

All in all, this exploratory study suggested that multimodality and synchronization between expressive channels are relevant to observers' evaluation of the spontaneity (or authenticity) of an expression. In agreement with the hypothesis that this relevance is due to the perceived dynamic coherence between signals pertaining to different modalities (i.e., synchronization), our analysis of tenors' behavior confirmed that the tenor who was rated higher for spontaneity was also the one showing patterns of synchronized behavior more frequently. Unfortunately, our study did not allow any direct test of the hypothesis and thus we cannot draw a causal link between behavioral synchronization and judgments of spontaneity; nevertheless, this study is a first attempt at investigating this hypothesis empirically and at operationalizing the concept of multimodal behavior synchronization in humans.

[7] The presence of the movement (separately for AU1 and AU2) was coded for each video frame, for three stages: onset, apex, and offset. The onset starts with the frame in which the first appearance change characteristic for the movement is observed. The apex starts with the frame that follows the last increase in intensity observed for the particular movement. The offset starts with the frame in which the first evidence of a decrease in intensity is observed. The offset terminates with the disappearance of the movement from the face or with a new onset.

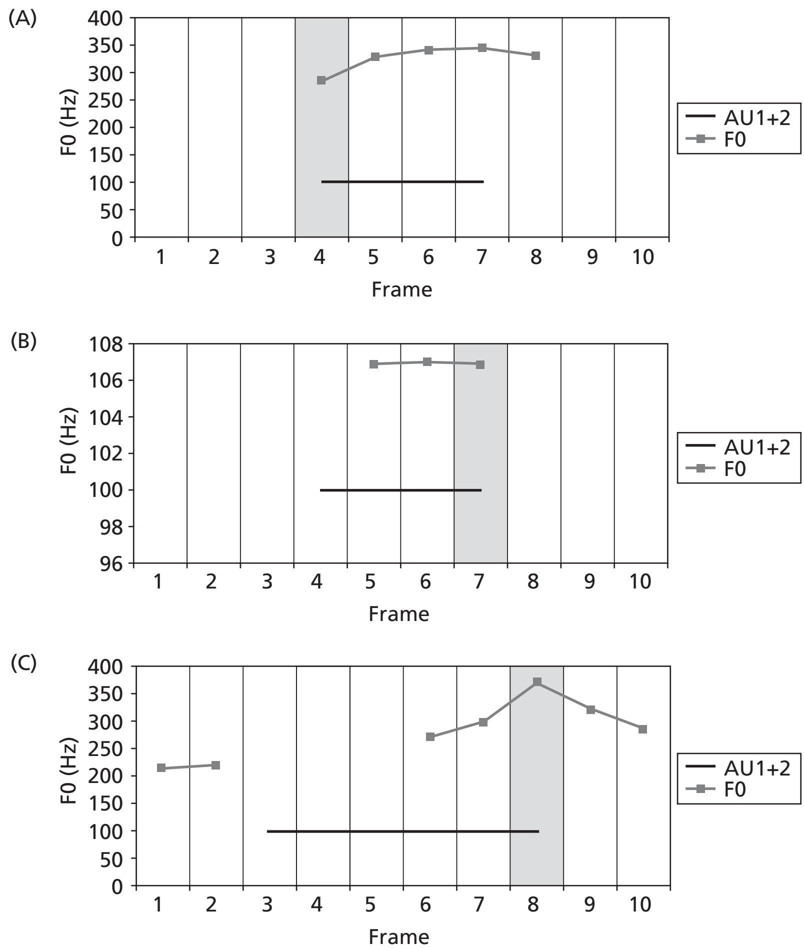

Figure 1.4 Examples of synchronized occurrences of "raised eyebrows" with vocal events. (a) Start of vocalization co-occurs with the beginning of the apex of AU1+AU2. (b) End of vocalization co-occurs with the end of the apex of AU1+AU2. (c) The end of the apex of AU1+AU2 co-occurs with a peak in the fundamental frequency contour.

Conclusion

What we tried to show is that multimodal synchronization is an important area of research for emotion psychology, especially when adopting an evolutionary account. Emotion evolved in humans as instrumental processes with an important role in the preparation of adaptive action (action readiness) and in the management of social relationships (through emotional communication and influence of conspecifics), expression is therefore an essential aspect of emotion. Emotion expression is the product of biological, cognitive, and social determinants that affect a number of expressive channels in different ways. Coherence and synchronization between modalities bring important information about how the expression has been produced and the receiver can apply different strategies to infer adaptive information from it. Unfortunately, with

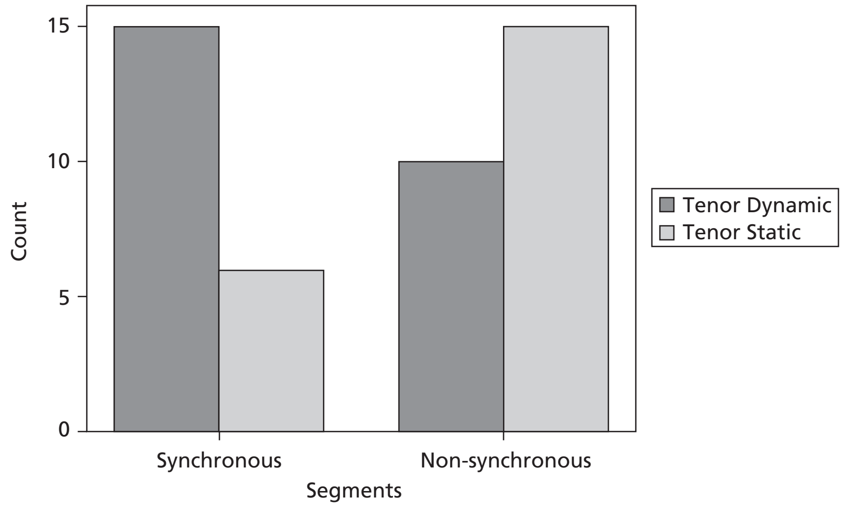

Figure 1.5 Frequency of synchronous and nonsynchronous segments for the two tenors. The distribution of synchronous and nonsynchronous segments was significantly different between the two tenors (Pearson's Chi-square = 4.54, $p < 0.05$). Dynamic tenor had a significantly higher proportion of segments containing a multimodal synchronous pattern than the Static tenor.

few exceptions, most studies are unimodal, mainly because of theoretical and methodological reasons. First, multimodal research requires the identification of the mechanism that determines the synchronized multimodal behavior. We suggest that the CPM offers a suitable theoretical framework for this kind of research, as it proposes a concrete production mechanism and predicts that response patterning including motor expression is driven by the sequential unfolding of appraisal outcomes (Scherer 2001, 2009). Second, apart from the urgent need to study all modalities or channels of expression in an integrated fashion, the study of the psychobiological and cultural coevolution of the expression of emotion in particular, requires a research paradigm adapted to this purpose. To disentangle biopsychological push and sociocultural pull effects, we suggest that researchers adopt a Brunswikian perspective, which combines both the production and perception aspects of emotional communication. The illustrative case study showed how observers' inference, based on the proximal cues, could be linked to micro-coded features of behavioral expression, allowing the identification of the distal cues involved, and the degree of synchronization between them that mediate impression formation. We are quite confident that investigating the two sides of communication by analyzing multimodal distal and proximal cues will greatly augment our ability to understand the relationships between production and recognition mechanisms and to determine the respective roles of push and pull features.

Acknowledgments

This research was supported by the Swiss National Science Foundation (FNRS 101411–100367 and FNRS 100014–122491/1) and by the National Centre of Competence in Research (NCCR) Affective Sciences (51NF40–104897) hosted by the University of Geneva. The authors thank Martijn Goudbeek for his precious contribution. Parts of the study included in this chapter have been presented at the XXIX International Congress of Psychology, Berlin, Germany, 20–25 July 2008.

References

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, **12**, 169–77.

Arnold, M.B. (1960). *Emotion and Personality. Vol. I. Psychological Aspects.* Oxford: Columbia University Press.

Aronson, A.E. (1990). *Clinical Voice Disorders: an Interdisciplinary Approach*. New York, NY: Thieme.

Banse, R . and Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, **70**, 614–36.

Bänziger, T., Mortillaro, M., and Scherer, K.R. (2011). Introducing the Geneva Multimodal Expression corpus for experimental research on emotion perception. *Emotion*. Advance online publication. doi: 10.1037/a0025827

Boersma, P. and Weenink, D.J.M. (1996). *Praat, a System for Doing Phonetics by Computer, Version 3.4.* Amsterdam: Institute of Phonetic Sciences of the University of Amsterdam.

Bonanno, G.A. and Keltner, D. (2004). The coherence of emotion systems: Comparing "on-line" measures of appraisal and facial expressions, and self-report. *Cognition & Emotion*, **18**, 431–4.

Brown, W.M. and Moore, C. (2002). Smile asymmetries and reputation as reliable indicators of likelihood to cooperate: an evolutionary analysis, in S.P. Shohov (ed.) *Advances in Psychology Research*, pp. 59–78. Huntington, NY: Nova Science Publishers.

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments.* Berkeley, CA: University of California Press.

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

Dael, N., Mortillaro, M., and Scherer, K.R. (2011). Emotion expression in body action and posture. *Emotion*. Advance online publication. doi: 10.1037/a0025737

Darwin, C. (1872/1998). *The Expression of the Emotions in Man and Animals* (3rd ed.). London: HarperCollins.

Davis, P.J., Zhang, S.P., Winkworth, A., and Bandler, R. (1996). Neural control of vocalization: respiratory and emotional influences. *Journal of Voice*, **10**, 23–38.

Dawkins, R. and Krebs, J. (1978). Animal signals: information or manipulation? in J.R. Krebs and N.B. Davies (ed.) *Behavioural Ecology: An Evolutionary Approach*, pp. 282–309. Oxford: Blackwell Scientific Publications.

DeJong, R.N. (1979). *The Neurologic Examination.* Hagerstown, MD: Harper and Row.

Delplanque, S., Grandjean, D., Chrea, C., *et al.* (2009). Sequential unfolding of novelty and pleasantness appraisals of odors: Evidence from facial electromyography and autonomic reactions. *Emotion*, **9**, 316–28.

Désiré, L., Veissier, I., Després, G., and Boissy, A. (2004). On the way to assess emotions in animals: do lambs *(Ovis aries)* evaluate an event through its suddenness, novelty, or unpredictability? *Journal of Comparative Psychology*, **118**, 363–74.

Ekman, P. (1979). About brows: Emotional and conversational signals, in M.V. Cranach, K. Foppa , W. Lepenies, and D. Ploog (ed.) *Human Ethology*, pp. 169–202. Cambridge: Cambridge University Press.

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

Ekman, P. and Friesen, W.V. (1972). Hand movements. *Journal of Communication*, **22**, 353–74.

Ekman, P. and Friesen, W.V. (1978). *The Facial Action Coding System.* Palo Alto, CA: Consulting Psychologist Press.

Ekman, P. and Friesen, W.V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, **6**, 238–52.

Ellgring, H. and Scherer, K.R. (1996). Vocal indicators of mood change in depression. *Journal of Nonverbal Behavior*, **20**, 83–110.

Flecha-Garcia, M.L. (2010). Eyebrow raises in dialogue and their relation to discourse structure, utterance function and pitch accents in English. *Speech Communication*, **52**, 542–54.

Fridlund, A.J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology*, **60**, 229–40.

Frijda, N.H. (2007a). What emotions might be? Comments on the comments. *Social Science Information*, **46**, 433–43.

Frijda, N.H. (2007b). *The Laws of Emotion.* Mahwah, NJ: Lawrence Erlbaum Associates.

Frijda, N.H. and Scherer, K.R. (2009). Emotion definitions (psychological perspectives), in D. Sander and K.R. Scherer (ed.) *The Oxford Companion to Emotion and Affective Sciences*, pp. 142–4. New York, NY: Oxford University Press.

Gellhorn, E. (1964). Motion and emotion: The role of proprioception in the physiology and pathology of the emotions. *Psychological Review*, **71**, 457–72.

Goschke, T. (2003). Voluntary action and cognitive control from a cognitive neuroscience perspective, in S. Maasen , W. Prinz and G. Roth (ed.) *Voluntary Action: Brains, Minds, and Sociality*, pp. 49–85. New York, NY: Oxford University Press.

Goudbeek, M. and Scherer, K.R. (2010). Beyond arousal: Valence and potency/control in the vocal expression of emotion. *Journal of the Acoustical Society of America*, **128**, 1332–6.

Grandjean, D. and Scherer, K.R. (2008). Unpacking the cognitive architecture of emotion processes. *Emotion*, **8**, 341–51.

Guilford, T. and Dawkins, M.S. (1991). Receiver psychology and the evolution of animal signals. *Animal Behaviour*, **42**, 1–14.

Halgren, E. and Marinkovic, K. (1995). Neurophysiological networks integrating human emotions, in M.S. Gazzaniga (ed.) *The Cognitive Neurosciences*, pp. 1137–51. Cambridge, MA: MIT Press.

Hamilton, W.D. and Zuk, M. (1982). Heritable true fitness and bright birds: A role for parasites? *Science*, **218**, 384–7.

Hammond, K.R. and Stewart, T.R. (2001). *The essential Brunswik: Beginnings, Explications, Applications.* New York, NY: Oxford University Press.

Hess, U., Banse, R., and Kappas, A. (1995). The intensity of facial expression is determined by underlying affective state and social situation. *Journal of Personality and Social Psychology*, **69**, 280–8.

Hess, U., Kappas, A., and Scherer, K.R. (1988). Multichannel communication of emotion: Synthetic signal production, in K.R. Scherer (ed.) *Facets of Emotion: Recent Research*, pp. 161–82. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hess, U. and Kleck, R.E. (1990). Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology*, **20**, 369–85.

House, D., Beskow, J., and Granström, B. (2001). Timing and interaction of visual cues for prominence in audiovisual speech perception. *Proceedings of Eurospeech 2001*, 387–90.

Hutton, R. (1972). Neurosciences: mechanisms of motor control in R.N. Singer (ed.) *The Psychomotor Domain: Movement Behaviour*, pp. 349–84. Philadelphia, PA: Lea & Ferbiger.

Izard, C.E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, **2**, 260–80.

Johnstone, T., Van Reekum, C.M., and Scherer, K.R. (2001). Vocal expression correlates of appraisal processes, in K.R. **Scherer,** A. Schorr and T. Johnstone (ed.) *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 271–84. New York, NY: Oxford University Press.

Juslin, P.N. and Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, **129**, 770–814.

Juslin, P.N. and Scherer, K.R. (2005). Vocal expression of affect, in J.A. Harrigan, R. Rosenthal and K.R. Scherer (ed.) *The New Handbook of Methods in Nonverbal Behavior Research*, pp. 65–135. New York, NY: Oxford University Press.

Kaiser, S. and Wehrle, T. (2001). Facial expressions as indicators of appraisal processes, in K.R. Scherer, A. Schorr and T. Johnstone (ed.) *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 285–300. New York, NY: Oxford University Press.

Krumhuber, E.G. and Scherer, K.R. (2011). Affect burst: Dynamic patterns of facial expression. *Emotion*, **11**, 825–841.

Leventhal, H. and Scherer, K.R. (1987). The relationship of emotion to cognition: A functional approach to a semantic controversy. *Cognition & Emotion*, **1**, 3–28.

Lipps, T. (1907). *Psychologische Untersuchungen.* Leipzig: Engelmann.

Maynard-Smith, J. and Harper, D.G. (1995). Animal signals: models and terminology. *Journal of Theoretical Biology*, **177**, 305–11.

Maynard-Smith, J. and Harper, D.G. (2003). *Animal Signals.* New York, NY: Oxford University Press.

Mehu, M., Grammer, K. and Dunbar, R.I.M. (2007). Smiles when sharing. *Evolution and Human Behavior*, **28**, 415–22.

Mehu, M., Mortillaro, M., Bänziger, T, and Scherer, K.R. (2012). Reliable facial muscles activation enhances recognisability and credibility of emotional expression. Emotion, 12(4), 701–715.

Mortillaro, M., Mehu, M., and Scherer, K.R. (2011). Subtly different positive emotions can be distinguished by their facial expressions. Social Psychological and Personality Science, 2, 262–271.

Nesse, R.M. (2009). Evolution of emotions, in D. Sander and K.R. Scherer (ed.) *The Oxford Companion to Emotion and the Affective Sciences*, pp. 159–64. Oxford: Oxford University Press.

Niedenthal, P.M. (2007). Embodying emotion. *Science*, **316**, 1002–5.

Owren, M.J. and Bachorowski, J.A. (2003). Reconsidering the evolution of nonlinguistic communication: the case of laughter. *Journal of Nonverbal Behavior*, **27**, 183–200.

Owren, M.J. and Rendall, D. (1997). An affect-conditioning model of nonhuman primate vocal signaling, in D.H. Owings, M.D. Beecher, and N.S. Thompson (ed.) *Perspectives in Ethology Vol. 12*, pp. 299–346. New York, NY: Plenum Press.

Partan, S.R. and Marler, P. (1999). Communication goes multimodal. *Science*, **283**, 1272–3.

Partan, S.R. and Marler, P. (2005). Issues in the classification of multimodal communication signals. *The American Naturalist*, **166**, 231–45.

Pell, M.D. (2005). Prosody-face interactions in emotional processing as revealed by the facial affect decision task. *Journal of Nonverbal Behavior*, **29**, 193–215.

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.

Rinn, W.E. (1991). Neuropsychology of facial expression, in R.S. Feldman and B. Rimé (ed.) *Fundamentals of Nonverbal Behavior*, pp. 3–30. Cambridge: Cambridge University Press.

Roberts, J.A., Taylor, P.W. and **Uetz, G.W.** (2007). Consequences of complex signaling: predator detection of multimodal cues. *Behavioural Ecology*, **18**, 236–40.

Roseman, I.J. and Smith, C.A. (2001). Appraisal theory: Overview, assumptions, varieties, controversies, in K.R. Scherer, A. Schorr and T. Johnstone (ed.) *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 3–19. New York, NY: Oxford University Press.

Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour*, **58**, 921–31.

Russell, J.A. and Fernandez-Dols, J.M. (1997). What does a facial expression mean? in J.A. Russell and J.M. Fernandez-Dols (ed.) *The Psychology of Facial Expression*, pp. 3–30. Cambridge: Cambridge University Press.

Scherer, K.R. (1985). Vocal affect signalling: A comparative approach, in J. Rosenblatt , C. Beer , M.C. Busnel, and P.J.B. Slater (ed.) *Advances in the Study of Behavior*, pp. 189–244. New York, NY: Academic Press.

Scherer, K.R. (1992). What does facial expression express? in K.T. Strongman (ed.) *International Review of Studies on Emotion Vol. 2*, pp. 139–65. Oxford: John Wiley and Sons.

Scherer, K.R. (1994a). Toward a concept of "modal emotions," in P. Ekman and R.J. Davidson (ed.) *The Nature of Emotion: Fundamental Questions*, pp. 25–31. New York, NY: Oxford University Press.

Scherer, K.R. (1994b). Affect bursts, in S.H.M. van Goozen and N.E. Van de Poll (ed.) *Emotions: Essays on Emotion Theory*, pp. 161–93. Hillsdale, NJ: Lawrence Erlbaum Associates.

Scherer, K.R. (2001). Appraisal considered as a process of multilevel sequential checking, in K.R. Scherer, A. Schorr and T. Johnstone (ed.) *Appraisal Processes in Emotion: Theory, Methods, Research*, pp. 92–120. New York, NY: Oxford University Press.

Scherer, K.R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, **40**, 227–56.

Scherer, K.R. (2005). What are emotions? And how can they be measured? *Social Science Information*, **44**, 693–727.

Scherer, K.R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, **23**, 1307–51.

Scherer, K.R. and Bänziger, T. (2010). On the use of actor portrayals in research on emotional expression, in K.R. Scherer, T. Bänziger, and E.B. Roesch (ed.) *Blueprint for Affective Computing: A Sourcebook*, pp. 166–76. Oxford: Oxford University Press.

Scherer, K.R. and Ellgring, H. (2007a). Are facial expressions of emotion produced by categorical affect programs or dynamically driven by appraisal? *Emotion*, **7**, 113–30.

Scherer, K.R. and Ellgring, H. (2007b). Multimodal expression of emotion: affect programs or componential appraisal patterns? *Emotion*, **7**, 158–71.

Scherer, K.R. and Grandjean, D. (2008). Facial expressions allow inference of both emotions and their components. *Cognition & Emotion*, **22**, 789–801.

Scherer, K.R., Zentner, M.R., and Stern, D. (2004). Beyond surprise: The puzzle of infants' expressive reactions to expectancy violation. *Emotion*, **4**, 389–402.

Smith, C.A. and Scott, H.S. (1997). A Componential Approach to the meaning of facial expressions, in J.A. Russell and J.M. Fernandez-Dols (ed.) *The Psychology of Facial Expression*, pp. 229–54. New York, NY: Cambridge University Press.

Tinbergen, N. (1951). *The Study of Instinct.* Oxford: Clarendon.

Tinbergen, N. (1952). The curious behavior of the stickleback. *Scientific American*, **187**, 22–6.

Tracy, J.L. and Robins, R.W. (2004). Show your pride: evidence for discrete emotion expression. *Psychological Science*, **15**, 194–7.

van den Stock, J., Righart, R., and de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion*, **7**, 487–94.

van Reekum, C.M., Johnstone, T., Banse, R., Etter, A., Wehrle, T., and Scherer, K.R. (2004). Psychophysiological responses to appraisal dimensions in a computer game. *Cognition & Emotion*, **18**, 663–88.

Wallbott, H.G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, **28**, 879–96.

Wehrle, T., Kaiser, S., Schmidt, S., and Scherer, K.R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology*, **78**, 105–19.

Zahavi, A. (1975). Mate selection: selection for a handicap. *Journal of Theoretical Biology*, **53**, 205–14.

Zahavi, A. and Zahavi, A. (1997). *The Handicap Principle: A Missing Piece of Darwin's Puzzle.* Oxford: Oxford University Press.