ACOUSTIC COMMUNICATION:
AN INTERDISCIPLINARY APPROACH

Organized by
Emma Otta & Patrícia Ferreira Monticelli
Universidade de São Paulo
Pró-Reitoria de Pesquisa da USP

DOI: 10.11606/9786587596198

Organizing Committee of the ACOUSTIC COMMUNICATION: AN INTERDISCIPLINARY APPROACH

Profa. Dra. Emma Otta (IPUSP)

Profa. Dra. Patrícia Ferreira Monticelli (FFCLRP)

Prof. Dr. Claudio Possani (IME-USP)

Dra. Tania Kiehl Lucci (IPUSP)

Dr. Ricardo Prist (IPUSP)

Cover Photo: Regina Macedo

Book formatting: Dra. Lilian Cristina Luchesi and Dra Aline Domingues Carneiro Gasco

English Editing Services: Michael Germain and Lisa Burger, MC TRADUÇÕES S/S LTDA

Catalogação na publicação
Serviço de Biblioteca e Documentação
Instituto de Psicologia da Universidade de São Paulo

Acoustic communication: an interdisciplinary approach / Organized por Emma Otta e Patrícia Ferreira Monticelli. -- São Paulo, Instituto de Psicologia da Universidade de São Paulo, 2021.

210 p.

E-book.

ISBN: 978-65-87596-19-8

DOI: 10.11606/9786587596198

1. Animal vocalization 2. Animal communication 3. Ethology I. Title

QL765

Ficha elaborada por: Elaine Cristina Domingues CRB5984/08

Funding Acknowledgements

Part B Human Bioacoustics

Chapter 5

Nonverbal acoustic communication from a psychoethological perspective

Emma Otta[7]

Abstract

The focus of this chapter is on acoustic nonverbal communication and the expression of emotions from a psychoethological perspective. Emotionally modulated speech, indicated by paralinguistic cues – such as speaking rate, tone of voice, and intonation contour – and vocal expressions – such as sobbing, screaming, and laughing – convey information about their sender, whether intentionally or unintentionally. Some paralinguistic cues and vocal cues may be used by the receiver to quickly infer the internal state of the sender, his/her intentions and his/her ensuing behavior, thus influencing the regulation of the interpersonal interaction. In this chapter, I will begin by presenting a definition of emotion and a historical contextualization. The chapter presents evidence of interspecific universals in emotional vocalizations. It distinguishes dimensional and categorical approaches to acoustically transmitted emotions. Evidence for both the universality and the cultural specificity of the detection of emotions from speech and human vocalizations will be presented. Unanswered questions and emerging topics will be pointed out throughout the chapter.

Keywords: Emotions, display, intentionality, paralinguistics, vocalizations.

Defining emotion

I will start by defining emotion from a functional perspective; this perspective reflects the "wisdom of ages" (Lazarus & Lazarus, 1994), as emotion has evolved as an adaptive mechanism through the process of evolution. I will use de Waal's definition, which, in my opinion, is especially appropriate in articulating proximate and distal levels of explanation (de Waal, 2011, p. 194):

[7] Department of Experimental Psychology of the Psychology Institute, University of São Paulo, SP, Brazil. emmaotta@usp.br

> "An emotion is a temporary state brought about by biologically relevant external stimuli, whether aversive or attractive. The emotion is marked by specific changes in the organism's body and mind – brain, hormones, muscles, viscera, heart, etcetera. Which emotion is triggered is often predictable by the situation in which the organism finds itself, and can further be inferred from behavioral changes and evolved communication signals. There exists no one-on-one relation between an emotion and ensuing behavior, however. Emotions combine with individual experience and cognitive assessment of the situation to prepare the organism for an optimal response. (…) Organisms have been selected to enter a particular bodily and mental state under particular circumstances: those who did furthered their interest better than those who did not."

Emotions are triggered by significant life events on the basis of an individual's needs and well-being. According to the component process model (CPM) of emotion (Scherer, 2009), five components are involved: (i) *evaluation of the event* – appraisal of the situation; (ii) *physiological activation* – preparing the body for action; (ii) *expressive movements* – vocal expressions, body postures, and facial expressions; (iii) *sense of purpose* – motivational state directed towards a goal; and (iv) *feelings* – subjective experience (Figure 5.1).

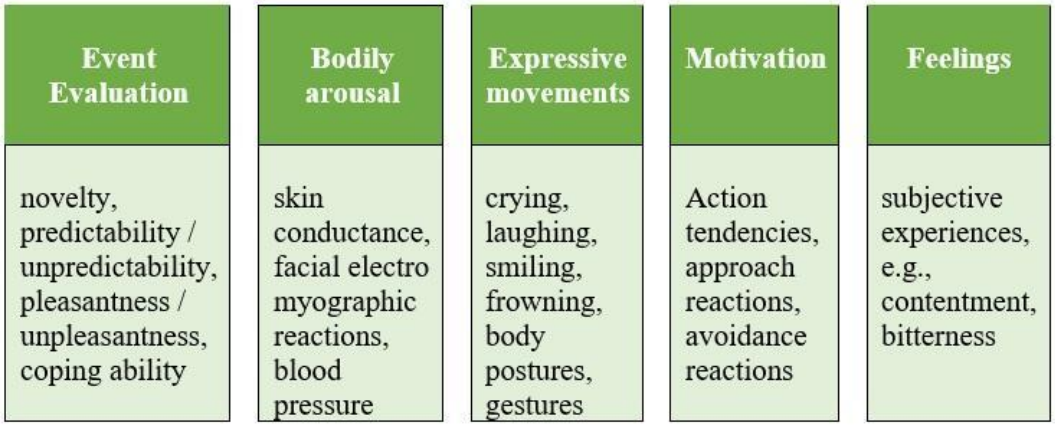

Figure 5.1. The five components of emotion (Source: after Sander et al., 2018).

Historical contextualization

Charles Darwin (1872) wrote about the expression of emotions in his book *The Expression of the Emotions in Man and Animals*. He enunciated the principle of

antithesis according to which certain mental states trigger a strong and involuntary tendency to perform certain movements. The induction of an opposite mental state triggers an equally strong and involuntary tendency to perform antagonistic movements. He illustrated the principle of antithesis with a dog in a threatening body posture, in which the animal appears larger and stronger, and in an appeasing body posture, in which it appears smaller and weaker (Figure 5.2).

> "When a dog approaches a strange dog or man in a savage or hostile frame of mind he walks upright and very stiffly; his head is slightly raised (…); the tail is held erect and quite rigid; the hairs bristle, especially along the neck and back; the pricked ears are directed forward, and the eyes have a fixed stare. (…) These actions follow from the dog's intention to attack his enemy. (…) As he prepares to spring with a savage growl on his enemy, the canine teeth are uncovered. ..." (Darwin, 1872, pp. 50-51)

Darwin then invites us to imagine that the dog suddenly notices that the individual he is approaching is not a stranger but his master. In this case, the dog's behavior changes immediately to the opposite.

> "Instead of walking upright, the body sinks downwards or even crouches, and is thrown into flexuous movements; his tail, instead of being held stiff and upright, is lowered and wagged from side to side; his hair instantly becomes smooth; his ears are depressed and drawn backwards, but not closely to the head; and his lips hang loosely. (…) the eyelids become elongated, and the eyes no longer appear round and staring." (p. 51)

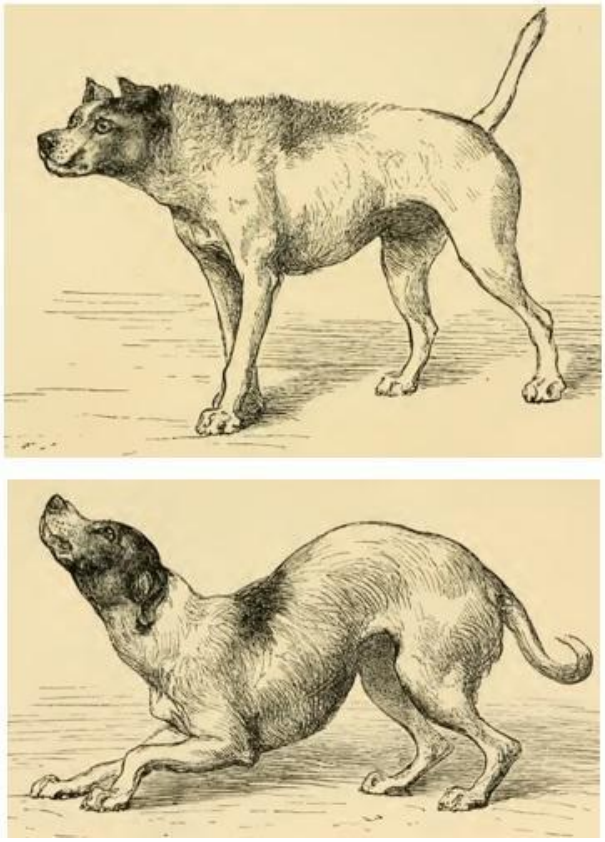

Figure 5.2. Principle of antithesis illustrated by a dog in threatening and appeasing postures (Source: Darwin, 1872, pp. 52-53). http://darwin-online.org.uk/converted/pdf/1897_Expression_F1152.pdf

Approximately 100 years after *The Expression of the Emotions in Man and Animals*, the subject was revisited by Morton (1977, 1983), who compared the physical structure of vocalizations used in the communication of mammals and birds at short distances and the underlying motivation. Animals indicate their mood through vocal sounds. Based on Darwin's antithesis principle, Morton argued that natural selection results in a structural convergence of sounds used in "hostile" and "friendly" contexts. An individual who tries to drive another away uses harsh (broadband), low-frequency sounds. If the receiver does not move away, the interaction may escalate into attack, indicating the sender's aggressive disposition. We can speculate that evolution may have favored roaring in aggressive animals because rough low-frequency vocalizations make the vocalizing individual appear larger and more threatening in the face of a rival (Morton, 1983). In threatening contexts, animals use this type of vocalization, while in friendly contexts, they use relatively tonal, high-frequency sounds. Sound characteristics (harsh quality, tonal quality, and sound frequency [pitch]) interact: 1)

the lower the frequency, the more hostile the sender, and the higher the frequency, the friendlier or more fearful the sender; 2) the greater the harshness, the more hostile the sender, and the more pure and tone-like, the friendlier or more fearful the sender; and 3) a decrease in frequency indicates hostility, and an increase in frequency indicates fear.

In line with Morton (1977, 1983), Ohala (1983) also proposed an innately specified "frequency code", relating the primary meaning of "large vocalizer" and the secondary meanings "dominant, aggressive, and threatening" to low acoustic frequency and the primary meaning of "small vocalizer" and the secondary meanings "subordinate, submissive, non-threatening, and desirous of the receiver's goodwill" to high acoustic frequency. Chuenwattanapranithi et al. (2008) obtained support for the size code hypothesis of emotional speech for two emotions – anger and happiness. Human listeners judged the body size and emotion of the speaker. Thai listeners heard speech sounds produced with a lengthened vocal tract, lowered F0, and roughened voice quality as spoken by an angry individual, and speech sounds produced with a shortened vocal tract, raised F0, and tone-like voice quality as spoken by a happy person.

Inferences about mood in animals can be made based on observations of sender and receiver behavior.

> "This is not to say that animals feel angry when they growl, for we have no idea whether our feelings are the same as the motivation in animals. But by observing what happens when an animal growls (or squeals), we can use words such as "aggressive" or "fearful" to describe whether it will probably attack or flee when it growls or squeals." (Morton, 1983, p. 345)

Returning to Darwin's example to illustrate the principle of antithesis, Morton (1983) added acoustic communication to visual communication.

> "Your pet dog is sleeping on the front porch. As you approach, Fido wakes up and begins barking. The bark means that Fido has perceived something of interest to him but the stimulus is too far away for him to make a "decision." Should he attack or be friendly? When you get closer, or yell his name, he changes from barking to whines, sleeks his fur, and wags his tail at a low angle. On the other hand, if the mailman had elicited the barks, Fido might begin to growl as he approached. It is clear from Fido's actions what moods he exhibited through his vocalizations. (…)

> When a dog growls, it also makes itself visually larger by erecting its fur; when it whines, it sleeks its fur and hunches down to look smaller." (p. 347)

There is redundancy in communication through the auditory and visual channels, and this redundancy may increase the odds of communicating effectively. In humans, the same general relationship exists between the physical structures of sounds and the underlying motivation (hostility or appeasement). One person can say "Go away!" in different ways to another, but when he/she feels truly angry, his/her feelings may be expressed as "growling". Intonation adds information to the content of his/her speech, making it more emphatic. A low or falling voice expresses aggressiveness and assertiveness, while a high or rising voice expresses friendly intentions (Morton, 1983).

Darwin's Musical Protolanguage Hypothesis

Darwin developed a model of language evolution, known as the musical protolanguage hypothesis. Figure 5.3 shows a schematic outline of this hypothesis. In chapter 4 of *The Expression of the Emotions in Man and Animals*, Darwin addressed sound emission and its use as a means of expressing courtship and rivalry. He hypothesized that our ancestors used vocal utterances to express emotions before they had acquired the power to articulate speech in human evolution. In Chapter 3 of *The Descent of Man, and Selection in Relation to Sex*, he proposed a model of language evolution based on a protolanguage that was more musical than linguistic and focused on sexual selection as the underlying mechanism of this evolution.

Darwin (1871, p. 87) stated:

> "I cannot doubt that language owes its origin to the imitation and modification of various natural sounds, the voices of other animals, and man's own instinctive cries, aided by signs and gestures. When we treat of sexual selection we shall see that primeval man, or rather some early progenitor of man, probably first used his voice in producing true musical cadences, that is in singing, as do some of the gibbon-apes at the present day; and we may conclude from a widely spread analogy, that this power would have been especially exerted during the courtship of the sexes, would have expressed various emotions, such as love, jealousy, triumph, and would have served as a challenge to rivals."

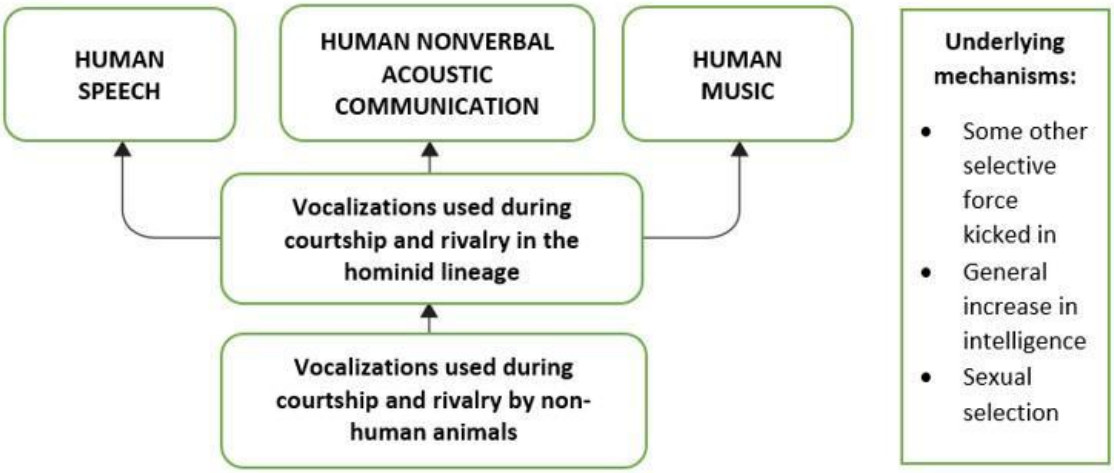

Figure 5.3. Schematic outline of Darwin's musical protolanguage hypothesis (Based on Zimmerman, Leliveld & Schehka, 2013 and Fitch, 2013***).***

Fitch (2010, 2013) summarized and updated this model. The first step was a general increase in intelligence in the hominid lineage. Before vocalizations were used meaningfully, they were used, "so to speak", aesthetically, to fulfil many of the same functions that modern humans use music today (courtship, bonding, territorial advertisement and defense, competitive displays, etc.). (…) at—a later stage (presumably during the evolution of meaningful language) some other selective force kicked in, so that language became equally (or better) expressed in females, and was pushed to develop early" Fitch (2013, p. 494, 497).

In his book *The Singing Neanderthals*, Mithen (2005) also proposed a proto-music/language model. He created the acronym *Hmmmm* to name his model for the early hominid communication system: **h**olistic utterances with their own meaning (not constituted by segmented elements or words), **m**anipulative (with the ability to change the affective states and behavior of others), **m**ultimodal (employing both sounds and movements), **m**usical (rhythmic and melodic), and **m**imetic (uses sound symbolism and gesture).

In search of interspecific universals in emotional vocalizations

Based on Darwin's (1871, 1872) hypothesis that the vocal expression of emotion has ancient roots and Morton's (1977) model on the association of motivation and structural rules of vocalizations, researchers began to search for interspecific

universals in emotional vocalizations. Fillipi et al. (2017) conducted a study in which they asked human participants to evaluate the emotional content of the vocalizations of nine species across three taxonomic classes: Amphibia, Mammalia, and Reptilia (including Aves). They found that speakers of different languages (native speakers of English, German and Mandarin, N=25 in each group) were able to identify increased levels of activation in the vocalizations of all species represented. We conclude from this study that humans are able to identify the emotional content in both conspecific and heterospecific vocalizations, suggesting a biological basis. Basic mechanisms underlying emotion perception in vocalizations may have appeared early during phylogenesis and have been evolutionarily conserved across species.

Cross-species perception of emotion from vocal and visual cues has also been investigated with dogs. In a research study for her master's thesis developed under my supervision as part of the Project Anthrozoo USP at the Postgraduate Program in Experimental Psychology at the University of São Paulo, with the co-supervision of Daniels Mills, Natalia Albuquerque (2013) used an intermodal preferential looking paradigm. Domestic dogs (*Canis familiaris*) were presented with human faces or dog faces with different emotional valences (happy/playful versus angry/aggressive). While the stimuli were projected onto two screens, a single sound was played. The sound could be a dog's bark, a human voice with either positive or negative valence or a control sound (neutral). We found that dogs looked significantly longer at both conspecific and heterospecific faces whose expression corresponded to the valence of the vocalization, leading to the conclusion that dogs have the mental prototypes for positive versus negative categorization of affect and that they are able to integrate acoustic and facial emotional information (Albuquerque et al., 2016). Additionally, the dogs seemed to have a functional understanding of emotional expressions. When they looked at angry human faces, they were more likely to mouth-lick than when they looked at happy human faces (Albuquerque et al., 2018). Mouth-licking may serve as an appeasement signal in dog-human communication (Firnkes, Bartels, Bidoli & Erhard, 2017). It is at the lowest step of Shepherd's ladder of distress signals. Shepherd (2002, 2009) attributes to this behavior the function of a calming signal that defuses conflict and restores harmony in a social interaction. Dogs are the oldest domestic animals, having lived with humans for approximately 10,000 years (Larson, Karlson, & Perri, 2012). It might have been particularly advantageous for them to recognize the emotions of humans and to react with appropriate behaviors during the process of

domestication, in which they may have evolved the ability to read human communication cues.

Dimensional and Categorical Approaches to Acoustically Conveyed Emotions

The chapter proceeds by addressing acoustically transmitted emotions in humans. Emotionally modulated speech and vocal expressions are evaluated in terms of either dimensional or categorical approaches. Self-report instruments linked to these approaches will be presented.

At the beginning of the twentieth century, Wilhelm Wundt (1905), the father of experimental psychology, first proposed a dimensional approach according to which emotions are characterized by their place in a three-dimensional space made up of "arousal-calm", "pleasure-displeasure", and "relaxation-tension". At the end of the twentieth century, Russel (1980) revisited the subject and proposed a circumplex model to express the structure of affect as evaluated by self-report. In this model, affective concepts fell in the circle in the following order: pleasure (0°), excitement (45°), arousal (90°), distress (135°), displeasure (180°), depression (225°), sleepiness (270°), and relaxation (315°). Bradley and Lang (1994) proposed the self-assessment manikin (SAM) as an easy-to-use nonverbal method for evaluating emotional reactions to a wide variety of stimuli, including sounds (e.g., International Affective Digitized Sound [IADS] System; Bradley & Lang, 2007), in terms of pleasure, arousal, and dominance. The pleasure scale shows a smiling figure at one end and a frowning figure at the other. The arousal scale shows a sleepy figure at one end and a wide-eyed figure at the other. The dominance scale shows a small figure at one and a large figure at the other. There was a 9-point rating scale for each dimension: research participants were asked to choose one of five figures in each scale or to place a mark between any two figures. The original paper-and-pencil self-report version evolved into a digital slider version used in smartphones and tablets (Betella & Verschure, 2016). Below each slider are two mirrored isosceles triangles that provide a visual cue for intensity.

Another approach to the study of acoustically conveyed emotions is the categorical emotional approach. At the end of the nineteenth century, Darwin (1872) first proposed a discrete emotions perspective with a main focus on the face, suggesting that facial expressions of emotion are universal. In contrast to Wundt (1905), he considered emotions distinct entities or modules, such as happiness, sadness, fear, anger, disgust, and surprise. Darwin's theory was extended by Ekman

(2009) and Panksepp (1998), assuming that basic emotions have unique characteristics that distinguish them from one another in important ways (behavioral and physiological reactions driven by specific neural reaction systems). These basic emotions are affect programs, phylogenetically evolved adaptation patterns activated by relevant eliciting events. In line with the predictions of the basic emotion theorists, humans assign facial and vocal expressions of emotion to discrete emotion categories with high accuracy. The Product Emotion Measurement Instrument (PrEmo) is an instrument based on the categorical emotional approach (Desmet, 2019). It is a pictorial self-report instrument in which a character expresses 14 different emotions by his/her expressions, body, and voice. Half of the emotions are positive (joy, hope, pride, admiration, satisfaction, fascination and attraction), and half are negative (sadness, fear, shame, contempt, dissatisfaction, boredom and disgust). In addition to basic emotions, this instrument includes social emotions that presuppose the ability to think about emotions and behavior from the point of view of another individual.

Mendl, Burman and Paul (2010) combined dimensional and categorical approaches for the study of emotions. Figure 5.4 schematically represents emotions in a two-dimensional space summarizing a dimensional approach (valence and activation) with an approach of discrete emotions (e.g., happiness, sadness, and fear). The positive emotions are placed in quadrants Q1 and Q2, and the negative emotions are placed in quadrants Q3 and Q4. The Q3-Q1 arrow represents the motivational system of reward acquisition related to increased fitness. The Q4-Q2 arrow represents the motivational system of punishment avoidance triggered by the perception of danger or threat.

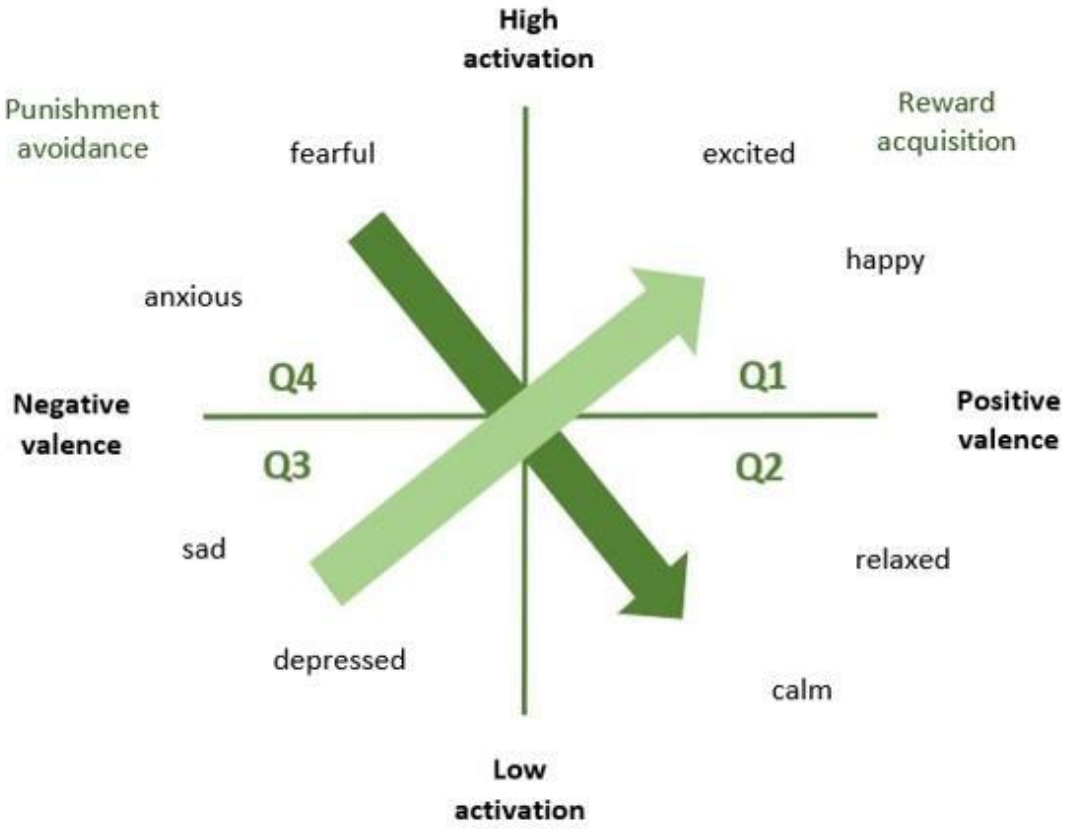

Figure 4. Discrete emotions located in a two-dimensional space of valence and activation (adapted from Mendl et al., 2010).

Stimulus materials

Researchers have developed a variety of stimulus materials that will be illustrated here. Affective prosody recognition tasks are used for research purposes and for clinical practice purposes. Individuals listen to neutral sentences (e.g., "*The girl went to the market"*) spoken in affective tones of voice (happy, sad, angry, etc.) and are asked to name the emotional prosody. The Brazilian version of the Florida Affect Battery (Bowers, Blonder & Heilman, 1999) was adapted from the English original by two researchers from the University of Brasilia (Costa-Vieira & Souza, 2014). In the naming of emotional prosody tasks, there are 20 trials with four repetitions of each of five affects. The Florida Affect Battery also includes tasks in which there is conflict between the semantic content and adequate voice intonation. An example is the sentence "*All the puppies are dead*" said in a happy tone of voice. To correctly evaluate

the affective tone of voice of the speaker, the listener must disregard the content of the message. The Florida Affect Battery also includes tasks in which there is conflict between the semantic context and adequate voice intonation. Additionally, the battery includes cross-modal facial prosody tasks in which individuals are required to match the affect conveyed by a prosodic stimulus with a corresponding facial expression or vice versa.

With the purpose of studying real emotional expressions, Silva, Barbosa and Abelin (2016) selected speech samples of Brazilian Portuguese from the documentary *Jogo de Cena* (*Playing*; Coutinho, 2007). The director had placed an announcement in a newspaper inviting women to tell their stories in his documentary, sharing their joys and sorrows. The film alternates between these women and actresses. There are few studies using real-life voice records, and this study is interesting for this reason.

Further materials used for research focus on vocalizations instead of speech. The Montreal Affective Voices (Belin, Fillion-Bilodeau & Gosselin, 2008) consists of 90 vocalizations corresponding to the emotions of anger, disgust, fear, pain, sadness, surprise, joy and pleasure. Sauter et al. (2010) also created a set of nonverbal vocalizations of negative and positive emotions, such as laughter, an angry growl, retching sounds, screams of fear, moans of sexual pleasure, and sighs of relief.

Detection of emotions from speech and human vocalizations

This section of the chapter will present research findings on the ability of listeners to identify emotions from speakers' voices. Comparisons have been made between accuracy in the recognition of emotion in the face and in the voice. While joy can be almost perfectly identified from facial expressions, listeners have difficulty recognizing this emotion unequivocally in the voice (Scherer, 2003). Anger and sadness were best recognized in the voice, followed by fear, whereas disgust was identified just above chance level.

Comparisons are also made across languages and cultures. In Silva, Barbosa and Abelin (2016), Brazilian and Swedish listeners evaluated authentic speech samples extracted from the Brazilian documentary *Jogo de Cena* and from Swedish television and interview programs. The Swedish and Brazilian listeners evaluated both corpora similarly, leading to the conclusion that the listeners' native language did not influence their perception of the emotions expressed by the speakers.

A meta-analysis conducted by Bąk (2016) showed that research on emotional prosody evaluated from speech is focused on English: 79% of the participants were Germanic language speakers, with a predominance of English; 15% were Japanese speakers; and 6% were speakers of other languages, including Portuguese, Spanish, Arabic, Hindi and Hinba. This meta-analysis showed that humans infer affective states from the emotional prosody of speech in different cultures, although the listeners cannot understand the words and sentences voiced by the speakers (Pell, Monetta, Ekmann & Kotz, 2009). The recognition of basic emotions (e.g., joy, sadness) was superior to chance, but there was an own-language advantage in comparison to foreign languages.

A study on intercultural recognition of basic emotions through nonverbal vocalizations comparing representatives of maximally different populations in terms of language and culture was conducted by Sauter, Eisner, Ekman and Scott (2010). This research team, integrated by a proponent of the "Big Six" basic emotions, compared Himba people from Namibia, who live by herding without contact with Western culture, with English-speaking Europeans. Emotional vocalizations corresponding to the basic emotions of joy, fear, anger, sadness, disgust and surprise communicated the same emotional states regardless of culture, leading to the conclusion that they are universals shared by all humans. In this study, however, the researchers also found a significant interaction between the culture of the listener and the emitter of vocalization, showing that each group performed better in relation to the stimuli produced by members of its own culture. Some social emotions (e.g., pride) were recognized only within culture. Negative emotions were recognized between cultures, but several positive emotions were communicated by culturally specific signals. The researchers concluded that affiliative social signals were shared mainly with in-group members.

In addition to self-report measures, psychophysiological measures have been used to investigate the processing of emotional prosody, including event-related potentials (ERPs), which identify specific brain activity by means of EEG in the presence of speech or vocalization samples. Basic emotions are differentiated by means of the P200 component captured by electrodes located in frontal-central positions. The differentiation between emotional speech and neutral speech occurs in an initial time window of 170 to 230 ms after the onset of stimulus (Paulmann, Bleichner, & Kotz, 2013; Schirmer, Simpson, & Escoffier, 2007). Through ERPs the

differential reactions of humans to fear vocalizations compared to control sounds 150 msec after the onset of the stimulus were demonstrated (Sauter & Eimer, 2010). Rapid detection of affective signals can be important. If conspecific others in the surroundings are frightened, staying alert and preparing for imminent danger may have survival advantages.

Concluding remarks

In this chapter, several topics related to acoustic nonverbal communication and the encoding and decoding of emotions have been addressed from a psychoethological perspective. Although scientists still struggle to define emotions, an operational definition was proposed, taking into account proximate and ultimate levels of causation. Research perspectives were presented based on the categorical approach to emotion, which goes back to Charles Darwin, pioneer in the study of emotions, and the dimensional approach to emotion, which goes back to Wundt, founder of the first psychology laboratory. A review was conducted of research findings of universals in nonhuman emotional vocalizations and cross-cultural recognition of basic emotions among humans. The present chapter shows evidence of interspecific universals and cross-cultural recognition of emotions and of an in-group advantage in the understanding of emotion. A universal acoustic affect program (Sauter et al., 2010) seems to coexist with culturally specific affect programs (Elfenbein & Ambady, 2003).

Acknowledgments

The author acknowledges grant no. 304740/2017-9 of Conselho Nacional de Desenvolvimento Científico e Tecnológico (The Brazilian National Council for Scientific and Technological Development – CNPq) and no. 2014/50282-5 of São Paulo Research Foundation (FAPESP) and Natura Cosméticos S.A. The author would also like to thank Patricia Monticelli for her comments and suggestions on initial versions of the text, which improved its quality.

References

Albuquerque, N. S. (2013). *Reconhecimento de emoções em cães domésticos (Canis familiaris): percepção de pistas faciais e auditivas na comunicação intra e interespecífica*. [Emotion recognition in domestic dogs (*Canis familiaris*):

Perception of facial and auditory cues in intra and interspecific communication] Instituto de Psicologia da USP: Dissertação de Mestrado.

Albuquerque, N., Guo, K., Wilkinson, A., Savalli, C., Otta, E., & Mills, D. (2016). Dogs recognize dog and human emotions. *Biology letters*, *12*(1), 20150883. https://doi.org/10.1098/rsbl.2015.0883

Albuquerque, N., Guo, K., Wilkinson, A., Resende, B. & Mills, D. S. (2018). Mouth-licking by dogs as a response to emotional stimuli. *Behavioural Processes*, *146*, 42-45. http://dx.doi.org/10.1016/j.beproc.2017.11.006

Bąk, H. (2016). The state of emotional prosody research—a meta-analysis. *In Emotional Prosody Processing for Non-Native English Speakers* (pp. 79-115). Springer, Cham.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods*, *40(2)*, 531-539. https://doi.org/10.3758/BRM.40.2.531

Betella, A. & Verschure, P. F. M. J. (2016) The Affective Slider: A Digital Self-Assessment Scale for the Measurement of Human Emotions. *PLoS ONE, 11* (2): e0148037. https://doi.org/10.1371/journal.pone.0148037

Bowers, D., Blonder, L. X., & Heilman, K. M. (1999). Florida affect battery, a manual. Centre for Neuropsychological Studies. *Cognitive Science Laboratory, University of Florida*, 3-19.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49-59. https://doi.org/10.1016/0005-7916(94)90063-9

Costa-Vieira, H. A., & Souza, W. C. D. (2014). O reconhecimento de expressões faciais e prosódia emocional: Investigação preliminar em uma amostra brasileira jovem. *Estudos de Psicologia (Natal)*, *19*(2), 119-127. http://dx.doi.org/10.1590/S1413-294X2014000200004

Chuenwattanapranithi, S., Xu, Y., Thipakorn, B., & Maneewongvatana, S. (2008). Encoding emotions in speech with the size code. *Phonetica*, *65*(4), 210-230. https://doi.org/10.1159/000192793

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*, 1st ed.; London, UK: John Murray.

Darwin, C. (1965). *The expression of the emotions in man and animals.* Chicago: University of Chicago Press. (Original work published 1872)

Desmet, P. M. A. (2019). PrEmo card set: Male version. Delft, Delft University of Technology. ISBN: 978-94-6384-076-7.

De Waal, F. B. (2011). What is an animal emotion? *Annals of the New York Academy of Sciences*, *1224*(1), 191-206. https://doi.org/10.1111/j.1749-6632.2010.05912.x

Ekman, P. (2009). Darwin's contributions to our understanding of emotional expressions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3449-3451. https://doi.org/10.1098/rstb.2009.0189

Elfenbein, H. A., & Ambady, N. (2003). Universals and cultural differences in recognizing emotions. *Current Directions in Psychological Science*, *12*(5), 159-164. https://doi.org/10.1111/1467-8721.01252

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., ... & Güntürkün, O. (2017). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: Evidence for acoustic universals. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1859), 20170990. https://doi.org/10.1098/rspb.2017.0990

Firnkes, A., Bartels, A., Bidoli, E., & Erhard, M. (2017). Appeasement signals used by dogs during dog–human communication. *Journal of Veterinary Behavior*, *19*, 35-44. http://dx.doi.org/10.1016/j.jveb.2016.12.012

Fitch, W. T. (2010). *The evolution of language*. Cambridge: Cambridge University Press.

Fitch, W. T. (2013). Musical Protolanguage: Darwin's Theory of Language Evolution Revisited. In Bolhuis, J.J. & Everaert, M., (Eds.) *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain* (pp. 489–503). Cambridge, MA: The MIT Press.

Larson, G., Karlson, E. K., and Perri, A (2012) Rethinking dog domestication by integrating genetics, archeology, and biogeography. *PNAS, 109*(23), 8878-8883. https://doi.org/10.1073/pnas.1203005109

Lazarus, R. & B. Lazarus. 1994. *Passion and Reason*. New York: Oxford University Press.

Mendl, M., Burman, O. H., & Paul, E. S. (2010). An integrative and functional framework for the study of animal emotion and mood. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1696), 2895-2904. https://doi.org/10.1098/rspb.2010.0303

Mithen, S. (2005). *The Singing Neanderthals: The Origins of Music, Language, Mind, and Body*. London: Weidenfeld & Nicholson.

Morton, E. S. (1977) On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist, 111* (981), 855– 869. https://doi.org/10.1086/283219

Morton, E. S. (1983). Animal communication: What do Animals say? *The American Biology Teacher*, *45*(6), 343-348. https://doi.org/10.2307/4447717

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F_0 of voice. *Phonetica*, *41*(1), 1-16. https://doi.org/10.1159/000261706

Panksepp, J. (1998). *Affective neuroscience: The foundations of human and animal emotions*. London, England: Oxford University Press.

Paulmann, S., Bleichner, M., & Kotz, S. A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology*, *4*, 345. https://dx.doi.org/10.3389%2Ffpsyg.2013.00345

Pell, M. D., Monetta, L., Paulmann, S. & Kotz, S. A. (2009) Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, *33*, 107–120. https://doi.org/10.1007/s10919-008-0065-7

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161-1178. https://doi.org/10.1037/h0077714

Sander, D., Grandjean, D., & Scherer, K. R. (2018). An appraisal-driven componential approach to the emotional brain. *Emotion Review*, *10*(3), 219-231. https://doi.org/10.1177/1754073918765653

Sauter, D. A., & Eimer, M. (2010). Rapid detection of emotion from human vocalizations. *Journal of Cognitive Neuroscience*, *22*(3), 474-481. https://doi.org/10.1162/jocn.2009.21215

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, *107*(6), 2408-2412. https://doi.org/10.1073/pnas.0908239106

Schirmer, A., Simpson, E., & Escoffier, N. (2007). Listen up! Processing of intensity change differs for vocal and nonvocal sounds. *Brain Research*, *1176*, 103–112. https://doi.org/10.1016/j.brainres.2007.08.008

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1-2), 227-256. https://doi.org/10.1016/S0167-6393(02)00084-5

Scherer, K. R. (2009) The dynamic architecture of emotion: Evidence for the component process model. *Cognition and Emotion*, *23*(7), 1307-1351. https://doi.org/10.1080/02699930902928969

Shepherd K. (2002). Development of behavior, social behavior and communication in dogs. In Horwitz D, Mills D, and Heath S. (Eds.). BSAVA *Manual of Canine and Feline Behaviour Medicine (pp. 8–20)*. Gloucester, UK: British Small Animal Veterinary Association.

Shepherd K. (2009). Behavioural medicine as an integral part of veterinary practice. In Horwitz D. and Mills D. (Eds.) BSAVA Manual of Canine and Feline Behaviour (pp 10–23), 2nd. ed.https://doi.org/10.22233/9781905319879.2

Silva, W. D., Barbosa, P. A., & Abelin, Å. (2016). Cross-cultural and cross-linguistic perception of authentic emotions through speech: An acoustic-phonetic study with Brazilian and Swedish listeners. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, *32*(2), 449-480. https://doi.org/10.1590/0102-445003263701432483

Wundt, Wilheim (1905). *Grundriss der Psychologie*. [*Outlines of psychology*]. Leipzig: Wilheim Engelmann.

Zimmermann, E., Leliveld, L., & Schehka, S. (2013). Toward the evolutionary roots of affective prosody in human acoustic communication: a comparative approach to mammalian voices. Eckart Altenmuller, Sabine Schmidt, Elke Zimmermann & Eckart Altenmüller (Eds.) *Evolution of emotional communication: from sounds in nonhuman mammals to speech and music in man* (pp. 116-132). Oxford: Oxford University Press.

Peer Commentary

By Sylvia Corte

Emma Otta is a Professor at the Department of Experimental Psychology, Institute of Psychology (IPUSP), of the University of São Paulo, São Paulo, SP, Brazil. I will comment on some of the aspects discussed by Professor Otta on a topic sometimes forgotten but with profound implications for acoustic communication and its emotional components, acoustic nonverbal communication. She approaches the topic from a psychoethological perspective, explaining emotions from a functional perspective and giving evidence of interspecific universals in the interpretation of emotions, for instance, the sounds emitted during laughter and crying in human babies. Otta refers to some research that showed that babies laughed in some situations that also triggered crying. However, this is a developmental process, where laughing substitutes crying. In addition, the stimuli that triggered laughing became less physical, such as tickling, and more cognitive during development. From an evolutionary perspective, there is a discussion about the contexts in which smiling and laughing appear. For example, smiling occurs in more conflictive situations and laughing in rougher and tumble play situations. Finally, we could study the acoustic manifestations of crying, laughing, and smiling from a longitudinal perspective in children.

Otta commented about the Twin Panel's interest in studying these manifestations in their participants. They are now studying acoustic communication in twins (see Claudio Possani chapter). They recorded twins' voices individually in the lab before COVID-19 and compiled the voices of one hundred pairs, which are currently being decoded using the PRAAT software. The Panel now is also studying children. It will be very interesting to study crying in twins and the development of their nonverbal vocalizations (crying and laughing) and language.

I wanted to inquire about the famous "Neanderthal flute," discovered in a cave in Slovenia, to be exposed to another of the topics exposed during the talk. One study dismissed the artifact as nothing more than a bone that had been chewed on by hyenas, while others argue it was a musical instrument. Could early humans, like the Neanderthal, have such sophisticated ways of communication?

Otta addresses archaeological evidence of wind instruments from the Upper Paleolithic (between 12,000 and 50,000 years ago) in association with modern humans (*Homo sapiens*). Neanderthals, the closest relatives of modern humans, became extinct around 40,000 years ago. A juvenile bear femur with two complete holes from the Middle Paleolithic was found in a cave in Slovenia in 1995. This piece of bone divides opinions. Some experts believe it is a flute, the oldest musical instrument, while others believe it is only a chewed carnivore bone, a pseudo-artifact. Computed microtomography studies show that a Neanderthal-made artifact cannot be ruled out. Although we lack a thorough understanding of Neanderthal behavior, some anthropologists believe they may have been intelligent, self-aware individuals using a primatological analogy. Chimpanzees, for example, drum on hollow trees and have a preference for some music over silence. One could surmise that Neanderthals may also have expressed something similar. Perhaps the Neanderthal flute was a flute and not a chewed carnivore bone.

Next, Professor Otta added some comments regarding another complex issue: how and why human language skills differ from our hominid ancestors and other living hominid species. How language evolved has been debated since Darwin. There are perceptual and cognitive abilities underlying language comprehension and production shared with other perceptual and cognitive processes present in nonhuman animals. However, there are grammar and syntax components that differentiate human language from animal communication systems. Human language refers to external things in the world and objects and events distant in place and time employing arbitrary symbols based on rules for combining these elements. According to Herbert Terrace in his article Noam Chomsky, all animals, ourselves included, communicate, but only humans use language as we do. Nonverbal communication appears earlier in nonhuman development and is also present in nonverbal individuals. Very interesting research had been carried out on children with an underdeveloped brain system. In other words, they have no cortex and communicate nonverbally. The researchers discuss their expectations that these individuals will not communicate. Nevertheless, they have a very rich communication system, characterized by smiling and gazing, for example. So, the role of nonverbal and verbal communication in humans can be studied separately.

Non-verbal and verbal communication in humans is a case of multimodal communication that can be exploited in some situations. We can study how they are

related and how we can learn from this. We can examine this multimodal combination in recent research (Albuquerque, 2013) on dogs: does a dog understand that an angry vocalization accompanies an angry face? The situation is more complex when we are dealing with humans. Professor Otta considered that both animals and humans make sense of the variety of sights, sounds, and affective states. They need to coordinate this input and make associations between one sense (e.g., sight) and another (e.g., sound). In the case of dogs, their behavior suggested that they could relate what they saw, heard, and felt. For example, they licked the nose more frequently when observing a negative facial expression than with a positive valence, especially in response to humans. It was an appeasement gesture, and researchers can hypothesize that it was a reaction to a threat exhibited in a prone conflict situation.

Dogs are very sensitive to human behavior and friendly or threatening communication signals. It is the result of the domestication process that probably involved relatively intense selection for tameness. However, nonverbal and verbal communication is a case of multimodal communication, and the combination of these various dimensions is a very interesting subject.