

Universal vocal signals of emotion

Disa A Sauter (disa.sauter@mpi.nl)

Department of Psychology, University College London, Gower Street, London
WC1E 6BT, United Kingdom.

Comparative Cognitive Anthropology, Max Planck Institute of Psycholinguistics, PO Box 310,
6500 AH Nijmegen, The Netherlands

Frank Eisner (f.eisner@mpi.nl)

Institute of Cognitive Neuroscience, University College London, 17 Queen Square,
WC1N 3AR, United Kingdom

Max Planck Institute of Psycholinguistics, PO Box 310,
6500 AH Nijmegen, The Netherlands

Paul Ekman

Human Interaction Laboratory, University of California San Francisco,
401 Parnassus Avenue, San Francisco CA 94143 USA

Sophie K Scott (sophie.scott@ucl.ac.uk)

Institute of Cognitive Neuroscience, University College London, 17 Queen Square,
WC1N 3AR, United Kingdom

Abstract

Emotional signals allow for the sharing of important information with conspecifics, for example to warn them of danger. Humans use a range of different cues to communicate to others how they feel, including facial, vocal, and gestural signals. Although much is known about facial expressions of emotion, less research has focused on affect in the voice. We compare British listeners to individuals from remote Namibian villages who have had no exposure to Western culture, and examine recognition of non-verbal emotional vocalizations, such as screams and laughs. We show that a number of emotions can be universally recognized from non-verbal vocal signals. In addition we demonstrate the specificity of this pattern, with a set of additional emotions only recognized within, but not across these cultural groups. Our findings indicate that a small set of primarily negative emotions have evolved signals across several modalities, while most positive emotions are communicated with culture-specific signals.

Keywords: emotion; voice; happiness; affective signals.

Background

Some human traits are shared by all human beings, despite differences in language, culture and ecology. These psychological universals tell us what features of the human mind are part of our shared biological heritage and which are products of culture and language. Because all humans share the vast majority of their genetic makeup with all other humans, there is great similarity in the physical features that are typical for our species, while minor characteristics vary between individuals. Similarly to physical features, many aspects of the human psychology are shared. For example, all human societies have complex systems of communication to

convey their thoughts, feelings, and intentions to those around them (Hauser, Chomsky & Fitch, 2002). But although there are some commonalities between different communicative systems, speakers of different languages cannot understand each other. Or rather, they cannot understand either other's words and sentences. Other aspects of communicative systems do not rely on common lexical codes and may be shared across linguistic and cultural borders. One candidate of communicative systems for constituting a psychological universal is emotional signals.

Humans use a range of signals to communicate emotions, including vocalizations, facial expressions, and postural cues. Auditory signals allows for affective communication when the recipient cannot see the sender, for example, across a distance or at night. In addition, young infants are sensitive to vocal cues from the very beginning of life, when their visual system is still relatively immature (Mehler, Bertoncini & Barrier, 1978).

Vocal signals of emotions can occur overlaid on speech in the form of affective prosody (Scherer, Banse & Wallbott, 2001). However, humans also make use of a range of non-verbal vocalizations to communicate how they feel, such as screams and laughs. In this study we investigate whether these kinds of non-verbal emotional vocalizations communicate the same affective states regardless of the listener's culture. Currently, the only available cross-cultural data comes from studies of emotional prosody in speech (e.g., Scherer et al., 2001). This work has indicated that listeners can infer some affective states from emotionally inflected speech across cultural boundaries. However, emotional information overlaid on speech (or nonsense-speech) is restricted by the segmental and supra-segmental structure of speech. In contrast, non-

verbal vocalizations are relatively “pure” expressions of emotions. Furthermore, no study to date has investigated emotion recognition from the voice in a population that has had no contact with other cultural groups, through media or personal contact.

We examined the universality of vocal signals of emotions using the two-culture approach, in which participants from two populations that are maximally different in terms of language and culture are compared (Norenzayan & Heine, 2005). The claim of universality is strengthened to the extent that the same phenomenon is found in both groups. This approach has previously been used in work demonstrating the universality of emotional facial expressions of the emotions happiness, anger, fear, sadness, disgust, and surprise (Ekman, Sorenson & Friesen, 1969), a result that has now been extensively replicated using different sets of facial signals (Elfenbein & Ambady, 2002).

In order to investigate whether emotional vocalizations universally communicate affective states, we compared British English speakers with the Himba, a semi-nomadic, pastoral people living in the Kaokoland region in Northern Namibia. The Himba live completely traditional lives, with no electricity, water, formal education, or any contact with Western culture or people.

Methods

Stimuli

The stimuli were taken from a previously validated set of non-verbal vocalisations of negative and positive emotions (Sauter, Calder, Eisner & Scott, 2009; Sauter & Scott, 2007). The stimulus set was comprised of ten tokens of each of nine emotions: achievement, amusement, anger, disgust, fear, sensual pleasure, relief, sadness, and surprise, based on demonstrations that all of these categories can be reliably recognized from non-verbal vocalizations by Western listeners. The sounds were produced by two male and two female native English speakers and the stimulus set was matched for peak amplitude. Further information about the stimuli is available in Sauter et al. (2009).

Participants

The Western sample consisted of 25 native British English speakers (10 male, 15 female; mean age 28.7 years). Twenty-nine participants (13 male, 16 female) from Himba settlements in Northern Namibia comprised the non-Western sample. The Himba do not have a system for measuring age, but no children or very old adults were included in the study.

Design and Procedure

In this study, we used an adapted version of a task used in previous cross-cultural research on recognition of emotional facial expressions (Ekman et al., 1969). In the original task, a participant heard a story about a person feeling in a particular way and was then asked to choose which of three emotional facial expressions fit with the story. This task is suitable for

use with a pre-literate population, as it requires no ability to read – unlike the forced-choice format that is common in emotion perception studies. Furthermore, the current task is particularly well suited to cross-cultural research, as it does not rely on the translation of precise emotion terms. The original task included three response alternatives on each trial, with all three stimuli presented simultaneously. However, as sounds necessarily extend over time, the response alternatives in the current task had to be presented sequentially. Thus, participants were required to remember the other response alternative(s) whilst listening to the current response option. In order to avoid overloading the participants’ working memory, the number of response alternatives in the current study was reduced to two.

The Western participants were tested in the presence of an experimenter; the Himba participants were tested in the presence of two experimenters and one translator. For each emotion, the participant was told a short emotion story describing a scenario that would elicit that emotion. After each story, the participant was asked how the person was feeling to ensure that they had understood the story. If necessary, participants could hear the story again. The emotion stories used with the Himba participants were developed together with a local person with extensive knowledge of the culture of the Himba people, who also acted as a translator during testing. The emotion stories used with the Western participants were matched as closely as possible to the Himba stories, but adapted to be easily understood by Western participants. The stories were played from recordings, spoken in a neutral tone of voice by a male native speaker each language (the Himba local language Otji-Herero and English). Once they had understood the story, the participant was played two sounds. The stimuli were produced by the experimenter pressing two computer mice in turn, each playing one of the sounds. The participant was asked which one was the kind of sound that the person in the story would make. They were allowed to hear the stimuli as many times as they need to make a decision. Participants indicated their choice on each trial by pointing to the computer mouse that had produced the sound appropriate for the story, and the experimenter inputted their response into the computer. Throughout testing, the experimenters and the translator were naïve to which response was correct and which stimulus the participant was hearing. Speaker gender was constant within any trial, with participants hearing two male and female trials for each emotion. Thus, all participants completed four trials for each of the nine emotions, resulting in a total of 36 trials. The target stimulus was of the same emotion as the story, and the distractor was varied in terms of both valence and difficulty, such that for any emotion participants heard four types of distractors: maximally and minimally easy of the same valence, and maximally and minimally easy of the opposite valence, based on confusion data from a previous study (Sauter et al., 2009). Which mouse was correct on any trial, as well as the order of stories, stimulus gender, distractor type, and whether target was first or second, was randomized for each participant.

Results

As expected, listeners from the British sample matched the sounds to the story at a level that significantly exceeded chance ($\chi^2 = 271.82$, $df = 1$, $p < 0.0001$), and they performed better than would be expected by chance for each of the emotion categories ($\chi^2 = 96.04$ (achievement), 88.36 (amusement), 81.00 (anger), 96.04 (disgust), 96.04 (fear), 51.84 (sensual pleasure), 67.24 (relief), 81.00 (sadness), and 70.56 (surprise), all $df = 1$, all $p < 0.001$, Bonferroni corrected; See Figure 1). This replicates previous findings that have demonstrated good recognition of a range of emotions from non-verbal vocal cues both within (Sauter et al., 2009) and between (Sauter & Scott, 2007) Western cultures.

The Himba listeners also matched the sounds to the stories at a level that was significantly higher than would be expected by chance ($\chi^2 = 271.82$, $df = 1$, $p < 0.0001$). For individual emotions however, they performed at better-than-chance levels only for a sub-set of the emotions ($\chi^2 = 49.79$ (amusement), 8.83 (anger), 27.03 (disgust), 18.24 (fear), 9.96 (sadness), and 25.14 (surprise), all $df = 1$, all $p < 0.05$, Bonferroni corrected; See Figure 1).

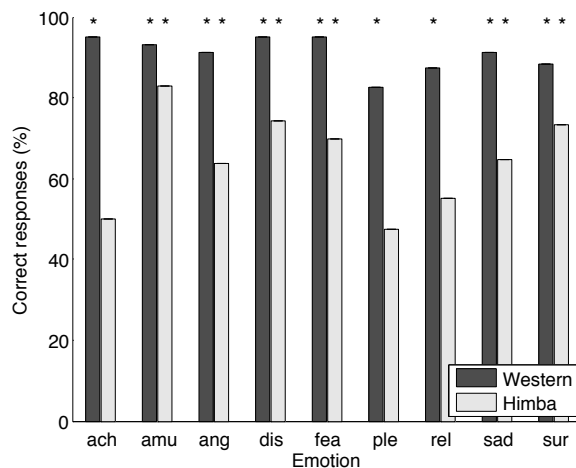


Figure 1: Recognition performance (%) for Western (dark bars; $n = 25$) and Himba participants (light bars; $n = 29$) for each emotion category. Stars indicate significantly better than chance performance (50%) for a group for a particular emotion category.

The emotions that were reliably identified by this non-Western sample comprise the set of emotions commonly referred to as the Basic Emotions (Ekman, 1992). These emotions are thought to constitute evolved functions that are shared between all human beings, both in terms of phenomenology and communicative signals. Notably, these emotions have also been shown to have universally recognizable facial expressions (Ekman et al., 1969; Elfenbein & Ambady, 2002). In contrast, vocalizations of several other emotions (achievement/triumph, relief, and sensual pleasure) were not recognized by the Himba participants, although non-verbal vocalizations of these emotions have previously been

shown to be reliably identified by several groups of Western listeners (Sauter & Scott, 2007). This pattern demonstrates that there are universally recognizable vocal signals for communicating the Basic Emotions, but that this does not extend to all affective states, including ones that can be identified by listeners from closely related cultures.

Discussion

Our results show that emotional vocal cues communicate affective states across cultural boundaries. The Basic Emotions anger, fear, disgust, happiness, sadness, and surprise, were reliably identified from vocalizations (See Figure 1). This indicates that some affective states are communicated with vocal signals that are broadly consistent across human societies, and do not require that the producer and listener share language or culture. This is consistent with research in the domain of visual affective signals. Facial expressions of the Basic Emotions are recognized across a wide range of cultures and correspond to consistent constellations of facial muscle movements (Ekman, 1999). These facial configurations produce alterations in sensory processing suggesting that they likely evolved to aid in the preparation for action to particularly important types of situations (Susskind et al., 2008). Furthermore, despite the considerable variation in human facial musculature, the facial muscles that are essential to produce the Basic Emotions are constant across individuals, suggesting that specific facial muscle structures have likely been selected to allow individuals to produce universally recognizable emotional expressions (Waller, Cray & Burrows). The consistency of emotional signals across cultures supports the notion of universal affect programs, that is, evolved systems that regulate the communication of emotions, which take the form of universal signals (Ekman, 1992). These signals are rooted in ancestral primate communicative displays. In particular, facial expressions produced by human and chimpanzees have substantial similarities (Parr, Waller & Heintz, 2008). Although a number of primate species produce affective vocalizations (Seyfarth & Cheney, 2003), the extent to which these parallel human vocal signals is as yet unknown. The data from the current study suggests that vocal signals of emotion are, like facial expressions, biologically driven communicative displays, that may be shared with non-human primates.

In humans, the basic emotional systems are modulated by cultural norms which dictate which affective signals should be emphasized, masked, or hidden (Matsumoto, Yoo, Hirayama & Petrova, 2005). In addition, culture introduces subtle adjustments of the universal programs, producing differences in the appearance of emotional expression across cultures (Elfenbein & Ambady, 2002). These cultural variations, acquired through social learning, result in the finding that emotional signals tend to be recognized most accurately when the producer and perceiver are from the same culture. The current study could provide a preliminary

demonstration of this pattern in the context of non-verbal vocalizations of emotion. It suggests that these signals are also modulated by culture-specific variation in a similar way to emotion facial expressions and affective speech prosody. Which cultural aspects are most important for modulating these different types of affective communication will be a question for future studies.

Some affective states are communicated using signals that are not shared across cultures, but specific to a particular group or region. In our study, vocalizations intended to communicate a number of positive emotions were not reliably identified by the Himba listeners. Why might this be? One possibility is that this is due to the function of positive emotions. It is well known that the communication of positive affect facilitates social cohesion with group members (Shiota, Campos, Keltner & Hertenstein, 2004). Such affiliative behaviors may be restricted to in-group members with whom social connections are built and maintained. However, it may not be desirable to share such signals with individuals who are not members of one's own cultural group. An exception may be self-enhancing displays of positive affect. Recent research has shown that postural expressions of pride are universally recognized (Tracy & Robbins, 2008). However, pride signals high social status in the sender rather than group affiliation, differentiating it from many other positive emotions.

In the current study, one type of positive vocalization was reliably recognized by both groups of participants: Listeners agreed, regardless of culture, that sounds of laughter communicated amusement, exemplified as the feeling of being tickled. Tickling triggers laugh-like vocalizations in non-human primates (Vettin & Todt, 2005) as well as other mammals (Knutson, Burgdorf & Panksepp, 2002), suggesting that it is a social behavior with deep evolutionary roots. Laughter is thought to have originated as a part of playful communication between young infants and mothers, and also occurs most commonly in both children and non-human primates in response to physical play (Provine, 2000). Our results support the idea that the sound of laughter is universally associated with being tickled, as participants from both groups of listeners selected the amused sounds to go with the ticking scenario. Indeed, given the well-established coherence between expressive and experiential systems of emotions (Rosenberg & Ekman, 1995), our data suggest that laughter universally reflects the feeling of enjoyment of physical play. Future work should address whether this universality extends to conceptual representations in semantic systems as well, which has been explored in the context of primarily negative emotions (Boster, 2005; Kimball Romney, Moore & Rusch, 1997).

In this study we show that a number of emotions are universally recognized from vocal signals, which are perceived as communicating specific affective states. The emotions found to be recognized from vocal signals correspond to those universally inferred from facial

expressions of emotions (Ekman et al., 1969). This supports theories proposing that these emotions are psychological universals and constitute a set of basic, evolved functions, that are shared by all humans. In addition we demonstrate that several positive emotions are recognized within, but not across cultural groups, which may suggest that affiliative social signals are shared only with in-group members.

Acknowledgments

This research was funded by a grant from the University of London Central Research fund to D. A. S., a contribution to travel costs from the University College London Department of Psychology, and a Wellcome Trust fellowship to S. K. S. The authors would like to thank David Matsumoto for useful discussions.

References

- Boster, J. S. (2005). Emotion categories across languages. In H. Cohen & C. Lefebvre (Eds.), *Handbook of Categorization in Cognitive Science*. Amsterdam: Elsevier.
- Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion*, 6, 169–200.
- Ekman, P. (1999). Facial expressions. In T. Dalgleish & T. Power, (Eds.), *The Handbook of Cognition and Emotion*. Sussex, UK: John Wiley & Sons.
- Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science*, 164, 86–88.
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128, 203–235.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298, 1569–1579.
- Kimball Romney, A., Moore, C. C. & Rusch, C. D. (1997). *Proceedings of the National Academy of Sciences of the USA*, 94, 5489–5494.
- Knutson, B., Burgdorf, J., & Panksepp, J. (2002). Ultrasonic vocalizations as indices of affective states in rats. *Psychological Bulletin*, 128, 961–977.
- Matsumoto, D., Yoo, S. H., Hirayama, S., & Petrova, G. (2005). Development and validation of a measure of display rule knowledge: the display rule assessment inventory. *Emotion*, 5, 23–40.
- Mehler, J., Bertoncini, J., & Barriere, M. (1978). Infant recognition of mother's voice. *Perception*, 7, 491–497.
- Norenzayan, A., & Heine, S. J. (2005). Psychological Universals: What are they and how can we know? *Psychological Bulletin*, 131, 763–784.
- Parr, L. A., Waller, B. M., & Heintz, M. (2008). Facial expression categorization by chimpanzees using standardized stimuli. *Emotion*, 8, 216–231.
- Provine, R. R. (2000). *Laughter: A Scientific Investigation*. London, UK: Faber and Faber.
- Rosenberg, E. L., & Ekman, P. (1994). Coherence between

- expressive and experiential systems in emotion. *Cognition and Emotion*, 8, 201–229.
- Sauter, D. A., Calder, A. J., Eisner, F., & Scott, S. K. (2009). Perceptual cues in non-verbal vocal expressions of emotion. *Manuscript under review*.
- Sauter, D. A. & Scott, S. K. (2007). More than one kind of happiness: Can we recognize vocal expressions of different positive states? *Motivation and Emotion*, 31, 192-199.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross Cultural Psychology*, 32, 76–92.
- Seyfarth, R. M. and Cheney, D. L. (2003). Meaning and emotion in animal vocalizations. *Annals of the New York Academy of Science*, 1000, 32–55.
- Shiota, M. N., Campos, B., Keltner, D., & Hertenstein, M. J. (2004). Positive emotion and the regulation of interpersonal relationships. In P. Philippot & R. S. Feldman (Eds.), *The regulation of emotion*. Mahwah, NJ: Erlbaum.
- Susskind, J. M., Lee, D. H., Cusi, A., Feiman, R., Grabski, W., and Anderson, A. K. (2008). Expressing fear enhances sensory acquisition. *Nature Neuroscience*, 11, 843–850.
- Tracy, J. L., & Robins, R. W. (2008). The nonverbal expression of pride: Evidence for cross-cultural recognition. *Journal of Personality and Social Psychology*, 94, 516-530.
- Vettin, J., & Todt, D. (2005). Human laughter, social play, and play vocalizations of non-human primates: An evolutionary approach. *Behaviour*, 142, 217–240.
- Waller, B. M., Cray, J. J., & Burrows, A. M. (2008). Selection for universal facial emotion. *Emotion*, 8, 435–439.