

# Non-verbal expressions of emotion – acoustics, valence and cross cultural factors

*Sophie Scott and Disa Sauter*

Institute of Cognitive Neuroscience  
University College London  
sophie.scott@ucl.ac.uk

## Abstract

This presentation will address aspects of the expression of emotion in non-verbal vocal behaviour, specifically attempting to determine the roles of both positive and negative emotions, their acoustic bases, and the extent to which these are recognized in non-Western cultures.

## 1. Introduction

Most work on the basic emotions has focused on facial expressions of emotion, with the successful delineation of their universal recognition, their basis in facial musculature, their perceptual features and their neural basis. Work from our lab has been attempting to extend this approach to the expression of emotion in the human voice, with a particular emphasis on non-verbal expressions of emotion – i.e. those not expressed as part of the speech signal (such as emotionally inflected speech). Thus our focus has been on ‘spontaneous’ emotional vocalisations such as laughs or screams ([1], see also [2] and [3] for related approaches to vocal emotions).

In this talk we present three different studies, each aimed at outlining different aspects of the processing of emotional vocalisations. In the first study we determined the ability of subjects to accurately categorise exemplars of emotional vocalisations, and used Principal Components Analysis to identify candidate dimensions underlying the subjects’ ratings of the same stimuli. In the second study we addressed the acoustic factors affecting the perception of these emotional vocalisations, by using a recognition study in which we selectively distorted aspects of the acoustics, and a separate acoustic analysis in which we related the previous ratings data to acoustic properties of the stimuli. Finally we tested the processing of these stimuli by a non-Western group, the Himba of Namibia.

## 2. Stimuli generation

First, a set of stimuli were generated by four different speakers, all native English speakers, two men and two women. These were recorded in an anechoic chamber. The speakers were given emotional scenarios, and asked to produce examples of appropriate vocalisations – importantly, these were not imitated or mimicked, and the speakers were instructed not to produce verbal items (such as ‘yuck’ or ‘hooray’). The ten emotional categories used were: amusement, triumph, relief, contentment (positive [4]), anger, disgust, sadness, fear (negative) and surprise. These stimuli were then piloted to select a set of sounds for each emotional condition which were roughly matched for performance, with 16 exemplars in each emotional condition.

## Study one – categorization and ratings of emotional stimuli

Twenty British English speakers were completed a study in which they categorized the emotional sounds according to the ten different emotional labels. Another twenty participants rated all the stimuli according to all the different emotional labels, and for arousal and valence, using seven point Lickert scales. The results showed that the emotional sounds could be accurately classified, with all categories being recognized at rates significantly above chance. There were also some systematic confusions, with contentment (which had the lowest recognition score) being confused with sensual pleasure, and vice versa. In the ratings task, the stimuli in each emotional category were rated most highly on that emotional label, although when rated for ‘contentment’, both sensual pleasure and triumph were stimuli were rated as highly as the contentment stimuli themselves. A PCA analysis of the ratings stimuli revealed two underlying dimensions, accounting for 53% and 16% of the overall variance, and which correlated with valence and arousal ratings respectively.

## Study two – acoustic factors

First, seventeen participants performed a categorization task on a sub-set of 5 of the stimuli from each emotional condition, randomly presented with transformed versions of the same stimuli, manipulated to either preserve pitch profile, some spectral detail, or the amplitude envelope only. Each manipulation had an effect on the recognition rate, relative to performance with the original stimuli, indicating that fine spectral detail is important in the accurate classification of emotional stimuli. The stimuli with preserved pitch were somewhat better recognized, suggesting that there is also a role for pitch variation in the perception of these emotional vocalisations. There was also some interaction with emotional class, and the amusement stimuli were well recognized across all transformation conditions, evidence that amplitude envelope alone is sufficient to recognize laughter.

In the acoustic analysis, a range of acoustic measures were taken from all sixteen stimuli for each emotional class; the measures encompassed factors associated with the amplitude envelope (e.g. duration, loudness), pitch (e.g. maximum, mean) and spectrum (e.g. spectral centre of gravity). These values were then used in a multiple regression with the ratings of the stimuli from study one, to identify with multiple regression which acoustic properties predict the participants’ ratings scores. The revealed that the ratings of each emotion were predicted by a combination of acoustic factors, and that no two emotional ratings were predicted by the same combination, with the exception of contentment and sensual pleasure. Furthermore, every emotional class was predicted by a combination of acoustic factors from two of the broad categories of envelope, pitch and spectral cues, except for disgust, which involved all three. Finally, surprise was the

only emotion not to involve spectral cues, further emphasizing the importance of spectral information in the identification of these stimuli.

- **Study three – recognition of vocal expressions in non-Western participants.**

In the final part of this paper, we tested a group of non-Western participants with little exposure to Western culture on their perception of these stimuli. Disa Sauter traveled to Namibia and tested members of the Himba tribe, a semi-nomadic group who herd cattle in the Namibian desert. The Himba were tested with the ten best recognized stimuli for nine emotional categories (as it seems to be ambiguous, contentment was excluded), and the tasks used included selecting a sound to match an emotional story, and same/different decisions on the emotional class of the stimuli. On the story task, the Himba accurately matched triumph, amusement, disgust, fear and relief stimuli to stories, and there was some suggestion of a sex difference, with the female participants correctly matching the sensual pleasure stimuli over twice as often as the male participants. In the same different task, performance improved and signal detection analysis showed a significant performance on all the emotions except for surprise. The Himba were also asked to generate emotional stimuli to the nine emotional classes, and these were tested with Western participants in an emotion categorisation task: all nine emotional classes were recognized at rate significantly above chance. This is thus the first study showing that recognition of emotional vocalizations is possible both from western to non-western, and vice versa. Further work with the Himba will unpack the areas where there seems to be disparity – for example, the Himba do not seem to associate ‘anger’ with a *noise*, and when asked to make an angry noise, typically start speaking in an angry fashion.

### 3. Discussion

‘Basic emotions’ are considered to be universal, to have unique expressions, to be innate, automatic and rapid, and to have distinct neural substrates. In this paper, we have attempted to address some of these issues with respect to non-verbal vocal expressions of emotion. We have identified that anger, triumph, amusement, disgust, fear, relief, sensual pleasure, sadness and surprise have different expressions, that these expressions are underpinned by different acoustic factors, and that there is some evidence that they can be recognized in cross cultural settings. In terms of the acoustic factors, we have shown that as well as pitch spectral and envelope cues are vital in the vocal expressions of emotion, and further work will be able to establish the extent to which this is closely related to the acoustics of emotional expression in speech [5] This is the first systematic attempt to establish the properties of non-verbal vocal expressions of emotion, and further studies will be able to distinguish whether, for example, the positive emotions which are considered to be principally conveyed vocally [4] do have correlates in other emotional channels, and what factors account for apparent cross-cultural differences.

### 4. References

[1] Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., & Johnson, M. (1997). Impaired auditory

recognition of fear and anger following bilateral amygdala lesions. *Nature*, 385, 254-257.

- [2] Schröder, M. (2003). Experimental study of affect bursts. *Speech Communication*, 40, 99-116
- [3] 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
- [4] Ekman, P. (1992a) An Argument for Basic Emotions. *Cognition and Emotion*, 6, 169-200.
- [5] Murray, I. R & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America*, 93, 1097-1108