Oxford Companion to the Affective Sciences
Oxford University Press

Type of Entry: L (1233 words)

Contact:

Martijn Goudbeek
Department of Psychology
University of Geneva
40, Boulevard du Pont d'Arve
1205 Geneva
Switzerland
Email: goudbeek@pse.unige.ch

Expression of Emotion – Vocal Expression

Martijn Goudbeek and Klaus R. Scherer

Swiss Center for Affective Sciences
University of Geneva

Speakers consistently modulate their speech, intentionally or unintentionally, to express different emotions and listeners are sensitive to these modifications and use vocal cues consistently to infer the emotional experience or message intention of the speaker. Emotional vocalizations range from nonverbal interjections or affect bursts <cross-ref affect bursts> and laughter <cross ref laughter> to linguistically complex spoken utterances with important suprasegmental and prosodic features such as pausing, rhythm, and intonation. This reflects the fact that vocalization is an evolutionary old emotion expression and signaling device, used both in animal and human communication, as well as the carrier of human speech. In consequence, the effects of linguistically determined speech motor commands and physiologically driven emotional voice modulations are inextricably intertwined in speech production. The emotional and attitudinal aspects of speaking are semiotically designated as

pragmatics as compared to syntax and semantics. Emotional communication is an inherently multimodal phenomenon <cross ref multimodal expression> and emotional vocalizations are not independent of the other modalities (de Gelder & Vroomen, 2000).

Encoding, transmission and decoding

The study of vocal emotion expression involves the entire chain from from voice production and encoding in the speaker via acoustic wave transmission to the listener's perception and inference (decoding). .

Encoding

There are several ways to study the encoding process (Scherer, 2003): spontaneous portrayals, laboratory induced emotions and acted emotional expressions. Spontaneous portrayals have been the favored approach in speech technology; journalists covering disasters, call center conversations, sporting events, and reality television provide a rich source of spontaneous emotional material. This approach, however, has serious drawbacks: most media appearances are most likely not completely spontaneous and determining which emotion is displayed is a difficult task. In many cases it is impossible to obtain systematic sampling of different expressions by the same speakers, which is absolutely essential given the importance of interindividual differences (Scherer, 2003).

Inducing emotional expression by means of experimental manipulations has been the favored approach of laboratory psychologists because of the degree of control it affords (Johnstone, van Reekum, Hird, Kirsner, Scherer, 2001). As with the previous approach, there are serious

drawbacks: the ethical considerations in inducing emotions like fear or sadness in naive participants and, by consequence, the fact that, generally, only rather weak affective responses can be elicited.

To avoid the disadvantages of these two methods, actor portrayals of emotional vocalizations are often used to obtain purer and more intense expressions as well as to systematically sample different emotions for the same speaker, generally using standard utterances (to control for semantic and phonetic content). Common approaches to provoke relatively realistic and authentic expressions by actors are the Stanislawski technique (method acting; recalling past emotional experiences from memory) and the Velten technique (repeating a highly emotionally charged sentence several times to induce consistent changes in emotional state). A common critique of acted emotional expressions is that they might result in more stereotypical portrayals. While most likely true, this is also an interesting source of information as stereotypical expressions can shed light on the evolutionary and societal origins of vocal emotional expression (Scherer, 2003).

Encoding studies pursue two major aims: 1) producing systematic stimuli for listener judgment studies (see below), and 2) to allow systematic acoustic analyses (generally using digital parameter extraction procedures from the recorded speech waves) of emotion-specific voice and speech feature profiles. The relevant acoustic properties for emotional speech can be grouped into four categories <cross ref voice production > : time related (utterance duration, speech rate), intensity related (loudness , dynamic range), fundamental frequency related (mean pitch, frequency perturbations) and complex time-frequency-energy related measures (voice quality and timbre, mainly visible in the speech spectrum). Research over the past decades has revealed a number of fairly reliable configurations of acoustic cues related to specific emotional expressions. Table 1 contains a compilation of acoustic patterning over

emotions based on the empirical data reported in several studies (see also Juslin & Scherer, 2005).

----------Insert Table 1 around here----------

Transmission

This part of the process concerns the transmission of the speech waves from the speaker's mouth to the listener's ear through the medium of wave propagation in the air. Transmission properties are important as ambient noise and amplitude and frequency composition of the speech signal can strongly effect the efficiency of the communication in the vocal/audio channel. However, this is rarely studied although one might expect that emotions can be differentially affected (e.g., low, soft sadness vocalizations traveling less well that loud, strident fear utterances).

Decoding

Listener decoding studies are used to examine the auditory discriminability of vocally expressed emotions and individual differences in recognition ability in the vocal modality. The major dependent variable, recognition accuracy, measures how well emotions can be identified from vocal cues alone in categorical forced choice or free response paradigms. Recent reviews of the literature (Scherer, 2003, Juslin & Laukka, 2003) indicate that the average percentage correctly classified based on vocal cues alone lies around 60%. While far from perfect and lower than that found in recognizing facial expression <cross ref Matsumoto entry >, this recognition accuracy is far above the a priori chance level and equals the

maximum accuracy of sophisticated statistical classification methods (Banse & Scherer, 1996). This indicates that quite specific patterns of acoustic parameters are used to infer different emotional states.

There are consistent differences in vocal recognition accuracy between emotions, with sadness and anger being recognized best, followed by fear. Disgust, joy, and despair are recognized least well vocally, although these emotions are often better recognized when nonverbal affect bursts rather than standard speech are used <cross ref affect bursts>. In addition to simple accuracy percentages, the reporting of confusion matrices is considered state of the art as these provide information on the decoding process and the cues that are used by the listeners. For example, most confusions occur within emotion families (anger is confused more readily with fear than with sadness), suggesting that similar cues are involved for the former pair. These differences can be explained by evolutionary considerations linked to transmission. Vocal information is well suited for the communication over larger distances, allowing to warn all conspecifics (fear) or threaten foes (anger) in hearing range. In contrast, disgust (provoked by unsavoury food) needs only be communicated over a short distance to those likely to want to share the food and is efficiently communicated by facial expression visible to interaction partners at close range.

As with the recognition of emotions in the facial modality, vocal emotion expressions are accurately recognized across cultures (Johnstone & Scherer, 2000; Elfenbein & Ambady, 2002), although there may be an in-group advantage: recognition of vocal emotion is easier in one's own language and culture (Scherer, Banse & Walbott, 2001).

In some cases, the Brunswikian lens model <cross ref lens model>, an integrative model of the complete vocal communication process, is used to study the relative weight of particular acoustic cues for the listener inferences (Scherer, 2003). Several methods can be used: 1)

correlating subjective ratings of unaltered portrayals with the extracted acoustic parameters, 2) using cue *masking* (or subtractive synthesis) to remove or mask specific cues from the original speech material (e.g., low-pass filtering, randomized splicing, backward replay, pitch inversion). Subsequent listener judgments of the resulting masked speech samples then reveal the relative contribution of different cues, and 3) cue *manipulation*, using rule or copy (re-) synthesis, to artificially create or modify specific acoustic parameters and measure their impact on decoding.

References

Banse, R., Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. J. Pers. Soc. Psychol. 70(3), 614-636.

Elfenbein, H., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: a meta-analysis. Psych Bull., 128, 203-235.

de Gelder, B., Vroomen, J. (2000). Bimodal emotion perception: Integration across separate modalities, cross-modal perceptual grouping or perception of multimodal events? Cognition Emotion 14, 321-324.

Johnstone, T. & Scherer, K. (2000). Vocal communciation of emotion. In M. Lewis & J. Haviland-Jones (Eds.), Handbook of Emotions, Second Edition (pp. 220-235). New York: Guilford Press.

Johnstone T, van Reekum C.M., Hird K, Kirsner K, Scherer K.R. (2005). Affective speech elicited with a computer game.. Emotion, 513-518.

Juslin, P.N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? Psychol. Bull, 129, 770-814.

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

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

State of the art readings

Juslin, P.N. and Scherer, K.R. (2005). Vocal expression of affect. In J. A. Harrigan, R. Rosenthal, & K. Scherer, (Eds.). The New Handbook of Methods in Nonverbal Behavior Research (pp. 65-135). Oxford University Press, Oxford, UK.

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

Scherer, K.R., Johnstone, T., Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer, H. Goldsmith (Eds.). Handbook of the Affective Sciences (pp. 433-456). New York and Oxford: Oxford University Press.

Tables

Table 1

Compilation of the review of empirical data on acoustic patterning of basic emotions (based on Johnstone & Scherer, 2000).

	Stress	Anger/Rage	Fear/ Panic	Sad- ness	Joy/ Elation	Boredom
Intensity	▲	▲	▲	▼	▲	
F0 floor/mean	▲	▲	▲	▼	▲	
F0 variability		▲		▼	▲	▼
F0 range		▲	▲ (▼)	▼	▲	▼
Sentence contours		▼		▼		
High frequency energy		▲	▲	▼	(▲)	
Speech and articulation rate		▲	▲	▼	(▲)	▼