Klaus R. Scherer

Swiss Center for Affective Sciences
University of Geneva

Corpus design for studying the expression of emotion in speech

In this presentation I will address the hotly debated issue of emotional speech sampling and corpora in research on emotional expression in psychology, linguistics, neuroscience, and affective computing. I argue that emotions are rare and fleeting events that are difficult to capture in a purely spontaneous fashion, especially as they are likely to be constantly manipulated for the purpose of self-regulation or social constraints. With the help of the new TEEP model (Tripartite Emotion Expression and Perception model) I propose an account of the fundamental mechanisms underlying emotional expression and perception and present a theoretical analysis that distinguishes between push (physiologically driven) and pull (social regulation and strategic intention) factors. Based on the model and the major determinants and mechanism under investigation, I propose that emotion sampling design and the use of specific corpora should be adapted to the specific goals pursued by a researcher. Is the work oriented toward an understanding of the production mechanisms underlying emotional speech (Speaker perspective, symptom function) or does it concern the nature of the recognition or attribution side (Listener perspective, appeal function)? Or are both perspectives to be employed in the aim to understand transmission and communication issues. Is the approach oriented toward an examination of non-linguistic or extralinguistic cues, or are linguistic and paralinguistic features of importance? To what extent does the symbolic aspect (iconic, indexical, referential) of the speech feature studied play a role?

I suggest that corpora with unobtrusive recordings of real-life expressions, laboratory induction, recording of expressive behavior from media shows, and explicit actor portrayals all have their place in emotion expression research. These vary continuously on several dimensions rather than representing completely different classes of expressions. I argue that an analysis of pure push factors, spontaneous unregulated expressions, is unrealistic in practice and probably of little interest, given the scarcity of such pure expressions in social life. Rather, the focus of current research should instead be directed toward the explicit study of pull effects, taking into account a variety of potential pull factors such as strategic communication intentions of the speaker, socio-normative and situational constraints, cultural and group expectations, media-specific concerns, and interactional or group dynamics. This, again, requires the choice of or the recording of appropriate corpora or data bases.

The search for vocal markers of emotion in speech has been hampered by the difficulty of obtaining access to speech samples that represent authentic expressions of the speaker's felt emotions. The recent trend to privilege real-life, naturalistic speech tokens, often obtained by convenience sampling, encounters two major problems: (1) the assumption that speech recorded in the field or from the media is a direct, uncontrolled expression of the speaker's "true" feeling state is unrealistic, given the widespread use of expression control due to display rules and strategic concerns; (2) the use of convenience samples, often of rare events, can engender the neglect of minimal requirements for experimental control of important determinants. Conversely, the use of performance induction of affect allows systematic control of influence factors and shows that even subtle variations of task characteristics and appraisal can produce major paralinguistic differences. The disadvantage of this type of elicitation is that the emotional effects are often weak and vary greatly over individuals. The use of actor portrayals is often criticized because of lack of authenticity. However, given that, as remarked by Erving Goffman, "we all play emotion theatre all the time" (pull effects!), this is not

much different in many real-life settings. The use of sophisticated techniques of "enacting emotions" using Stanislaski or method acting techniques can be an appropriate methodological choice for certain research aims, given the possibility of manipulating and standardizing pull effects. The central role of actor portrayals clearly lies in the empirical and experimental study of the shared code of emotional signaling and the examination of cue utilization in emotion perception and inference. This is particularly important if different languages and cultures are to be compared systematically.

In order to make the right choice for one's object of study, one needs to (1) be able to distinguish what is true and what is false or what is real and what is artificial or faked; (2) obtain access to the true and valid expressions; (3) ascertain that all requirements for experimental control in scientific research are fulfilled, a precondition for valid inference and evidence; and (4) make sure the speech material is appropriate for the question under investigation. In the presentation some general issues concerning the expression of emotion in the voice that are of relevance for systematic research in this area will be addressed: (1) the conditions that need to be attained to safeguard the minimal requirements for obtaining scientifically valid evidence, based on the established canon of systematic experimental research; (2) the paradigm of performance induction of affect; and (3) the difference between experimental mood induction, a psychological method to induce real emotions in laboratory settings, and the enacting of emotions through memory recall and mental simulation by professional actors. Examples from our past research will be provided throughout the presentation.