The Role of Prosody in Affective Speech

Bearbeitet von
Sylvie Hancil

1. Auflage 2009. Taschenbuch. 403 S. Paperback
ISBN 978 3 03911 696 6
Format (B x L): 15 x 22 cm
Gewicht: 590 g

schnell und portofrei erhältlich bei

beck-shop.de
At the centre of the relationship between ‘langue’, defined as “the underlying system shared by all speakers of a given language” (Culioli 1995: 4), and ‘langage’, defined as “speech activity as well as an evolving mass of individual speech events” (Culioli 1995: 4), lies the speaker, as was already acknowledged by Lyons (1982: 102), who associates the role of the speaker with the origin of enunciation. Here the speaking subject, to paraphrase Benveniste (1966), is understood as the one that is at the origin of enunciation.

In everyday conversational speech, the speaker will go through different types of affective states, all of which will reflect the interaction between his inner state and his environment in discourse. The complex manifestation of the various subjective states of a speaker, including attitudes and emotions, in discourse defines what will be the focus of our attention in this volume, namely affective speech, which cannot be reduced to a purely physiological phenomenon. Despite being a still new, emerging field, affective speech has been the object of significant, fruitful research over the last few years, especially with the official launching by Roddy Cowie of the HUMAINE Association at the end of the last plenary session organized with much success by Laurence Devillers and Jean-Claude Martin in June 2007 at the CNRS, Paris. The goal of this volume is to present a collection of review papers and research reports illustrating state-of-the-art research on prosody and affective speech, especially emotions.

The first part of this volume focuses on the sociolinguistic variables that are likely to influence the manifestation and the interpretation of affective speech in prosody. The first three analyses discuss the impact of culture on the manifestation of affective speech from various points of view. On the basis of ethology, which mainly explains behaviour in terms of phylogenetic adaptations, John J. Ohala, provides a critical analysis of the various theoretical accounts underlining the connections between sounds and meanings and applies these principles to another level, involving the cross-cultural use of other emotional signals, namely kinesic elements such as eyes and eyebrows in nonverbal communication, where a gaze into the eyes of another individual can be interpreted as a threat or not. The novelty of his argumentation, which is borne out by biological and ethological evidence, contributes to shed more light on the richness of emotional communication. Takaaki Shochi, Albert Rilliard, Véronique Aubergé and Donna Erickson emphasize the interrelation of language with culture in the expression and perception of affects. More especially, their study focuses on the perception of the attitudes of twelve Japanese by groups of American English and French-speaking listeners unfamiliar with Japanese language and culture. The experiments show the cross-cultural manifestations and interpretations of affective speech that different cultures might have in common, along with the differences that can be a source of misunderstanding. Sophie de Abreu, Catherine Mathon and Alessandra Mosca offer a comparison of prosodic features of anger in three Romance languages, namely French, Italian and Portuguese. They also show that in the expression of this affective burst, the pragmatic context plays a key role in the strategic use of the acoustic parameters of fundamental frequency, intensity and speech rate, whether they are taken separately or together.

The next two contributions take into consideration the age criterion in their research, one being cross-generational, the other centering around senior citizens. Relying on the Code of Effort elaborated by Carlos Gussenhoven, Geneviève Caelen-Haumont’s study explores in detail a specific emotional state, which reflects the personal involvement of the speaker, in a corpus of spontaneous conversations involving four female speakers from four different generations in the same family in the Bordeaux region. She analyzes the internal structure of the melodic prominences, coined melisms, along with the group duration, pauses and speech flow rates. The statistical findings clearly demonstrate that there is an underlying structure behind the apparent chaos of the F0 values and that melodic subjectivity is not manifested by the speaker at random but under specific laws. Likewise, the analysis also takes into account a comparison between melisms and sequences of joy. The second contribution interested in the role of age is by Brigitte Zellner-Keller, who examines the relation of speech with aging, a still unexplored domain. The age factor affects various aspects of speech among senior citizens. The decrease of complexity in syntactic structures is correlated with a slowdown process in articulation and tempo, a decrease in fundamental frequency and intensity, an alteration of voice quality, along with an increasing difficulty in grasping emotional prosody.

The volume includes two neurobiological studies that underline the importance of sex differences in the expression of emotional prosody. The first one is by D. Erik Everhart, Amy J. Shiply...
and Heath A. Demaree. The article demonstrates that even though the right-hemisphere hypothesis of emotion processing is well supported in data, the results need to be refined since recent research indicates variation within this model as a function of sex. The differences observed are explained in relation to arguments elaborated in social cognitive neuroscience, a promising, emerging field. The second study pointing to the relevance of sex differences in the neurophysiological response to emotional expressions is by Annett Schirmer and Qingyang Li. It investigates the processing of attended and unattended vocal emotional expressions in men and women by measuring event-related brain potentials. The results show that unattended speech is more likely to attract attention when it is produced with emotional as compared to neutral prosody. Moreover, in women emotional vocalizations seem to capture attention more efficiently than neutral vocalizations, whereas in men attention capture seems largely unaffected by speaker emotion.

The second part of the volume pertains to the way emotion recognition is implemented in synthesis systems and how machine applications can contribute to a better description of emotion(s). Roddy Cowie and Ellen Douglas-Cowie focus on what they call ‘conversational emotion’, a still highly unstudied emotion encountered in everyday conversation, which is characterized as moderate. Relying on the ‘Sensitive Artificial Listener’ scenario to arouse a whole spectrum of emotions, they recorded a corpus of this type of emotion within two European projects, ERMIS and HUMAINE, and in this volume they present an analysis of the underlying speech patterns, a research that has not been carried out yet. In order to analyze the mass of information in a coherent and efficient way, they offer a stimulating model, which enlightens and challenges the complexity of the relationships between speech variables and perceived affect, especially the significant correlations between speech and emotion ratings for each rater and tuner length and the relationship between pitch range at the end of a tune and net affect, without neglecting the gender and rater effects, and the relationship to machine recognition and psychological theory. Petri Laukka, Nicolas Audibert and Véronique Aubergé’s work can be situated in continuity with the Cowies’ proposal of integrating the various theories of cognitive psychology in their interpretation of natural emotional speech, since Laukka et al. use work on categorization in their research. They examine what determines typicality, namely graded structure, in vocal expressions of emotion. The tests conducted show that the typicality of vocal expressions is in most cases influenced by their similarity to ideal category members, and not by the frequency of instantiation or similarity to central tendency. The findings suggest that instead of classifying prototypical vocal expressions as common taxonomic categories, it would be more appropriate to characterize them as goal-derived categories. The cognitive model under examination could well be of interest to make useful comparisons between prototypical and spontaneous expressions of emotional speech.

The next two studies focus on advanced algorithms and statistical models with a view to enhancing emotion recognition by machines. In order to predict the relevant variations of prosodic parameters in relation to expressivity, defined as an information level of communication, Grégory Beller uses a Bayesian model, which can infer the variations corresponding to a new utterance, whose context is known. This statistical context-dependent method presents the advantage of inferring new cases, which are absent from the learning experience and are necessary for the transformation of new utterances, while maximising the likelihood of observed data. By taking into consideration recent results on acoustical transformations, this model turns out to be a good intermediary between data and rules, as it can be a useful tool for artistic creation, and it contributes to render human-machine interfaces more credible. Björn Schuller, Martin Wöllmer, Florian Eyben, Xiaodi Wang, and Gerhard Rigoll center their research on how to identify the most adequate emotion-signalling features. They investigate feature effects on multiple, partly public databases, using layer-wise optimized Support-Vector-Tree constructions, together with an attribute evaluator based on Correlation-based Feature Subset-selection (CFS). The results show that spectral features display the best information about emotion within speech, whether it be in the recognition performance or in the results of correlation-based feature subset-selection. The findings can be used to design machines more efficiently due to lower extraction effort and more compact feature space representation.

Carlos Busso, Murtaza Bulut, Sungbok Lee and Shrikanth Narayanan criticize the standard approach in current emotion recognition and analysis systems, which analyzes high-level information in terms of various statistics of prosodic and spectral features. Although powerful, such methods do
not directly take into account the interactions between different parameters and therefore are not sufficient to represent the complex variations in acoustic parameters that occur during expression of emotions. Their chapter presents results of a detailed analysis on F_0 contour, energy and duration characteristics. The purpose is to identify how these parameters are manipulated in expressive natural speech and to show how they can be used in the implementation of emotion recognition and synthesis systems. The results present a useful guide for the design of human-machine interaction interfaces that better incorporate expressive content.

The last two contributions are concerned with the way synthesis systems can contribute to the improvement of human-machine interfaces. Olivier Piot’s chapter studies the prosodic expression of two attitudinal parameters, one being ignorance of the speaker/hearer, the other being the desire to know/inform. Relying on Klatt-synthesized stimuli in standard French, the experiments make sure that speech rate, accent loudness and F_0-range are cross-variated. In both questions and assertions, the value of the F_0 has a highly positive influence on the two attitudinal evaluations tested. A higher speech rate iconically reflects a more intense desire to know; in the same way, a lower speech rate iconically shows greater hearer’s ignorance. A higher loudness on a stressed syllable emphasizes both cognitive parameters in assertions but has no impact on questions. The results are also discussed in relation to the main theories of affective speech. Ioulia Grichkovtsova, Michel Morel and Anne Lacheret conduct three perception tests in French and in English with the speech synthesizer Kali, relying on the gating and transplantation paradigms, which help identify the role played by voice quality and prosodic pattern in the studied affective states. Their main goal is to develop prototypical constructions for vocal emotions in both languages with this synthesizer. The findings are discussed in relation to the methodo-logical problems encountered and to the practical applications to be developed.

In conclusion, by thoroughly discussing the interaction of empirical data, methodological discussions and theoretical issues, this volume shows how fine-grained analyses of prosody in everyday emotion-coloured conversations can contribute to a better understanding of affective speech.