

Emotional Speech in Human–Computer Communication

Maciej Karpiński (Adam Mickiewicz University, Poznań)

The mechanisms of emotional behavior in humans have been systematically examined since the advent of contemporary psychology [1], [2]. Presently, emotions are studied not only in the context of basic survival, but also in social interactions and as a supplement to human intellect [3], [4]. Since emotions strongly affect human communicational and linguistic behavior, emotionality may shortly become a vital part of many speech-based human-computer communication systems [5], [6], [7]. While many meaningful and revealing studies have been carried out in the fields of emotional expression in spoken language and many relevant features of emotional have been determined, more research is still urgently needed. This especially applies to the studies of naturally occurring speech (as opposed to “laboratory” emotional speech, produced consciously by professional speakers), complex (mixed) emotions, and the implementation of emotional speech engines in machines and environments [8], [9], [10].

A substantial part of the present knowledge about emotional speech comes from corpus-based studies. The design and preparation of emotional speech corpora and databases is a demanding task [8], [10], [11]. In the corpora of naturally (spontaneously) occurring speech, usually only a small proportion of utterances can be clearly classified as expressing certain emotions. Moreover, emotional labeling itself poses serious problems, because “pure” emotions are rarely met while their mixes are often quite complex in description [12], [13]. Emotional categories and possible hierarchies of emotions still remain controversial [3].

While lexical and syntactic properties of emotional speech are equally important, this paper is focused on its acoustic-phonetic features which are discussed on the basis of a number of contemporary studies. Special attention is paid to the suprasegmental component, with intonation as an extremely rich information source.

Pitch parameters seem to be relatively easy to track instrumentally with existing phonetic software. Pitch range, average pitch level, as well as the character of pitch changes in time may be important cues to the emotional content of an utterance [14], [15], [16]. However, the final shape of intonational contour is determined by many factors related to the utterance itself, speech situation, and the speaker. In emotional speech, normal influences of these factors may be

disturbed, leading to general comprehension problems. Loudness and tempo (especially their changes) may also provide valuable information about emotions conveyed in speech signal (e.g., [12]). Speech rhythm (and its disfluencies) may also prove revealing in terms of emotional information. Finally, voice quality (e.g., harshness, breathiness, laryngealization, brilliance) [17] and segmental phenomena [18] are also relevant components of emotional speech to be included in its general model. Most of contemporary emotional speech synthesizers make use of these parameters [17], [18], [19]. Obviously, emotional speech recognition poses much more problems [20], [21].

Naturally occurring emotional speech results from the action of certain underlying human emotional mechanisms [24]. Accordingly, emotional robots and virtual agents should be provided with software that would enable them to simulate emotional behavior and, consequently, to produce emotional speech in a contextually relevant and communicationally meaningful way. However, providing a machine with such abilities may mean giving it consciousness and a question arises whether we really need or want it [25].

Keywords:

Emotional speech, intonation, human-computer communication

Selected references

- [1] James, W. 1884. What is an emotion? *Mind*, vol. 9, pp. 188 – 205.
- [2] Darwin, C. 1872. *The expression of emotions in man and animals*. New York: D. Appleton and Company.
- [3] Cornelius, R. R. 2000. Theoretical approaches to emotions. *ISCA Workshop on Emotions in Speech, Belfast 2000*.
- [4] Ekman, P., Davidson, R. J. 1994. *The Nature of Emotions: Fundamental Questions*. New York: Oxford University Press.
- [5] Cañamero, D. 1999. What Emotions are Needed in HCI? [In:] H.-J. Bullinger, J. Ziegler (Eds.) *Human-Computer Interaction: Ergonomics and User-Interfaces*, vol. 1, Mahwah, NJ: Lawrence Erlbaum Associates, pp. 838 – 842.
- [6] Bates, J. 1994. The role of emotions in believable agents. *Communications of the ACM*, vol. 37, no. 7. pp 122 – 125.
- [7] Dautenhahn, K., Bond, A., Canamero, L. D., B. Edmonds, B. (Eds.) 2002. *Socially Intelligent*

Agents: Creating Relationships with Computers and Robots. Norwell, MA: Kluwer Academic Publishers.

- [8] Campbell, N. 2000. Databases of Emotional Speech. *ISCA Workshop on Emotions in Speech, Belfast 2000*.
- [9] Picard, R. W. 1995. Affective computing. *MIT Media Lab Perceptual Computing Section Tech. Rep.*, No. 321, 1995.
- [10] Karpiński, M. (with W. Jassem and J. Kleśta) 2002. Polish Intonational Database: Project Report. (available in Polish from the project team: maciej.karpinski@amu.edu.pl).
- [11] Tsan-Long, P. 2004. The Construction and Testing of a Mandarin Emotional Speech Database. *Proceedings of ROCLING04*.
- [12] Douglas-Cowie, E., Cowie, R., Schroeder, M. 2003. The description of naturally occurring emotional speech. *Proceedings of the 15th ICPhS, Barcelona*.
- [13] Roach, P. 2000. Techniques for the Phonetic Description of Emotional Speech. *Proceedings of ISCA Workshop on Emotions in Speech, Belfast 2000*.
- [14] Paeschke, A., Sendlmeier, W. F. 2000. Prosodic characteristics of emotional speech: Measurements of fundamental frequency movements. *ITRW on Speech and Emotion, Newcastle 2000*.
- [15] Paeschke, A., Kienast, M. & Sendlmeier, W. F. (1999): F0-contours in emotional speech. *Proceedings ICPhS 99*, San Francisco, Vol. 2, pp. 929 – 932.
- [16] Karpiński, M. 2001. The prosodic expression of surprise and astonishment in jokes: A listening task. [In:] St. Puppel, G. Demenko (Eds.) *Prosody 2000*. Poznań: Faculty of Modern Languages and Literature, UAM.
- [17] Johnstone, T, Scherer, K. R. 1999. The effects voice quality. *The Proceedings of ICPhS99*, pp. 2029 – 2032.
- [18] Kienast, M., Paeschke, A. & Sendlmeier, W. F. 1999. Articulatory reduction in emotional speech. *Proceedings Eurospeech 99*, Budapest, Vol. 1, pp. 117 – 120.
- [19] Iida, A., Campbell, N., Yasumura, M. 1998. Design and Evaluation of Synthesised Speech with Emotion. *Journal of Information Processing Society of Japan*, vol. 40, pp. 479 – 486.
- [20] Hofer, G. O. 2004. Emotional Speech Synthesis. Master of Science School of Informatics, University of Edinburgh.
- [21] Murray, I. R. & Arnott, J.L. 1993. Towards a simulation of emotion in synthetic speech: a

review of the literature on human vocal emotion. *JASA* 93 (2), p. 1097–1108

- [22] Cowie, R. et al. 2001. Emotion recognition in human–computer interaction. *IEEE Signal Processing Magazine*, vol. 18, pp. 32 – 80.
- [23] Kwon, O.-W., Chan, K., Hao, J., Lee, T.-W. 2003. Emotion Recognition by Speech Signals. *Eurospeech 2003*, Geneva.
- [24] Berckmoes, C., Vingerhoets, G. 2004. Neural Foundations of Emotional Speech Processing. *Current Directions in Psychological Science*, vol. 13, no. 5, pp. 182 – 185.
- [25] Ball, G., Breese, J. 2001. Emotion and personality in a conversational agent. [In:] Cassell *et al.* (Eds.): *Embodied conversational agents*. Cambridge, MA: MIT Press.