

Automatic classification of emotional states: purpose, possibilities, prospects

Anton Batliner

University of Erlangen-Nuremberg,
Lehrstuhl für Mustererkennung / Chair for Pattern Recognition,
Martensstr.3, 91058 Erlangen, F.R.G,
batliner@informatik.uni-erlangen.de

Abstract

If we look back at the development of automatic speech processing during the last decades, we see that it started with - seemingly - simple tasks like pinning down what people read aloud; little by little, the tasks got (even) more complicated: speaker-dependent dictation systems, speaker-independent recognition, spontaneous speech, language understanding, dialogue processing. In the last decade, an even more ambitious task has been addressed, namely the paralinguistic aspects of human-human and human-machine conversation: the recognition and generation of emotional behaviour. Here, we want to tell apart two different application types: the first type is 'information' dialogues that can be observed, e.g., in conversations with automatic dialogue systems (call-center applications). Normally, the user displays only a rather restricted repertoire of emotional behaviour - for instance, she is getting angry if the system constantly misunderstands. The second type is 'emotional' dialogues in the realm of edutainment or entertainment; here, emotional behaviour can be an integral part of the dialogues, i.e., we can observe a mutual re-inforcement of emotions by repeated exchange.

After presenting these two types of applications, we first want to address some specific problems as, e.g., finding the right 'reference', i.e., the classes that have to be classified, and we want to discuss the differences between basic research dealing with elicited, pre-defined emotions on the one hand, and application-oriented research on the other hand. In the second part, we present the necessary 'ingredients' for successful classification: databases, extraction of acoustic and other features, automatic classifiers, aso. In a third part, we want to exemplify all these topics with own studies, one of them dealing with finding emotional behaviour in 'information' dialogues, the other dealing with finding such behaviour in sort of 'emotional' dialogues, namely human-robot communication. In the last part, we summarize the findings and put them into a broader perspective.

1. References

- A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Nöth. 2000. Desperately Seeking Emotions: Actors, Wizards, and Human Beings. In (Cowie et al., 2000), pages 195–200.
- A. Batliner, C. Hacker, S. Steidl, E. Nöth, and J. Haas. 2003. User States, User Strategies, and System Performance: How to Match the One with the Other. In *Proc. of an ISCA tutorial and research workshop on error handling in spoken dialogue systems*, pages 5–10, Chateau d'Oex, August. ISCA.
- A. Batliner, V. Zeissler, C. Frank, J. Adelhardt, R. P. Shi, and E. Nöth. 2003. We are not amused - but how do you know? User states in a multi-modal dialogue system. In *Proc. European Conf. on Speech Communication and Technology*, pages 733–736, Geneva, Switzerland, September.
- A. Batliner, K. Fischer, R. Huber, J. Spilker, and E. Nöth. 2004. How to Find Trouble in Communication. *Speech Communication*, 40:117–143.
- A. Batliner, C. Hacker, S. Steidl, E. Nöth, S. D'Arcy, M. Russell, and M. Wong. 2004. "You stupid tin box" - children interacting with the AIBO robot: A cross-linguistic emotional speech corpus. In *Proceedings of the 4th International Conference of Language Resources and Evaluation LREC 2004*, Lisbon.
- R. R. Cornelius. 2000. Theoretical Approaches to Emotion. In (Cowie et al., 2000), pages 3–10.
- R. Cowie and R.C. Cornelius. 2004. Describing the emotional states that are expressed in speech. *Speech Communication*, 40:5–32.
- R. Cowie, E. Douglas-Cowie, and M. Schröder, eds. 2000. *Proc. of the ISCA Workshop on Speech and Emotion: A Conceptual Framework for Research*, Newcastle, Northern Ireland, September.
- C.M. Lee, S. Narayanan, and R. Pieraccini. 2001. Recognition of Negative Emotions from the Speech Signal. In *Proceedings of the Automatic Speech Recognition and Understanding Workshop (ASRU'01)*.