



ELSEVIER

Available at  
[www.ComputerScienceWeb.com](http://www.ComputerScienceWeb.com)  
POWERED BY SCIENCE @ DIRECT®

**SPEECH**  
COMMUNICATION

Speech Communication 40 (2003) 1–3

[www.elsevier.com/locate/specom](http://www.elsevier.com/locate/specom)

Editorial

## Speech and Emotion

To the non-expert, the face and the voice stand out as the two chief instruments for communicating human emotion. Research on emotion and the face is highly developed. Research on speech and emotion is much less so. However, there are signs that the balance is changing. This special issue reflects that trend, and aims to encourage it.

A key factor in the situation is that research on speech and emotion stands between territories that have traditionally been cultivated by distinct communities. Each has some of the equipment and know-how necessary to make progress in the territory that lies between them, but not all. Of course speech is a highly technical topic. Equally, though, the literature on emotion is notoriously large and tricky, with a long history of conflicting theories.

As a result, the natural pattern of development for research on speech and emotion is a joint enterprise involving several communities: some primarily concerned with speech, others primarily with emotion. That kind of enterprise is notoriously fragile. It has to survive the tendency for experts in one area to react defensively when people who are not experts in their particular disciplines encroach into neighbouring territory.

The aim of this special issue is to promote constructive interaction among the communities involved in the emerging field. Its roots lie in a workshop sponsored by the International Speech Communication Association, whose proceedings are available via the web ([www.qub.ac.uk/en/isca/index.htm](http://www.qub.ac.uk/en/isca/index.htm)).

The papers represent the major themes that emerged in the workshop. First come two papers dealing with preliminaries that need to be addressed, and where detached analysis is clearly

needed. Perhaps the single most contentious theme at the workshop was defining and describing emotion as it relates to speech. Cowie and Cornelius (both psychologists) offer an integrative overview of the issues surrounding that theme. Conversely, the most generally agreed theme was the need to develop databases capable of underpinning speech research. Douglas-Cowie, Campbell, Cowie and Roach (with backgrounds spanning linguistic fieldwork, phonetics and speech-processing) review the collections underpinning contemporary research, and offer critical evaluations of the more ambitious approach that is beginning to emerge.

It is a feature of the field that attention is being paid to an increasing range of variables. The paper by Alter, Rank et al. reflects the fact that traditional phonetic measurement, with due care over potentially confounding linguistic factors, is still indispensable; but it also raises the possibility of linking that kind of measurement to techniques from the rapidly developing field of brain scanning. Menezes, Pardo et al. provide fine-grained analysis of the changes in timing associated with irritation, using acoustic analysis in conjunction with direct measurements of relevant articulators. The paper by Aubergé and Cathiard reflects the fact that the expression of emotion is multimodal, with the face and the voice playing roles that are neither independent nor redundant. The paper by Schröder highlights the role of utterances that intrude into the speech stream, in contrast to the more usual emphasis on variation in the way the speech stream is realized.

Classically, emotion recognition has been concerned with classifying examples of the archetypal

emotions that psychologists would call basic (or, mainly in older literature, primary). That approach is reflected as part of a wider picture in the review papers which end the collection. However two experimental reports also reflect a different emphasis which is increasingly prominent. Batliner, Fischer et al. provide concrete evidence for a view that several groups have come to hold for less direct reasons. It is that research on archetypal emotions does not easily transfer to the manifestations of emotion that occur in real interactions. Hence, if the field intends to have real applications, it has to deal directly with the emotional and emotion-related states that are likely to arise in real applications. One of those is stress. Fernandez and Picard from the MIT group, who have been prominent in the development of 'affective computing', report techniques for eliciting speech that shows various levels of stress, and for recognizing it.

Synthesis is an area where there clearly are applications for systems oriented towards the traditional emotion archetypes. Iida, Campbell et al. describe one example. Their synthesis system is designed to provide vocal expression for users with disabilities that prevent them from speaking themselves. Its acceptability is materially enhanced by allowing users to choose one of several emotional states the speech should convey. Gobl and Ní Chasaide report progress on an alternative, and less familiar approach. Their topic is one of the classic issues for synthesis, the relationship between voice quality and emotion. But instead of setting out to model predefined emotional states, they measure the emotional impact of synthesizing the cardinal voice qualities described by Laver. The implication of their findings is that questions about emotion become unavoidable once synthesis begins to vary voice quality: the variations will convey impressions of particular emotional balances, whether or not that is the intended goal.

The collection ends with two wide ranging reviews. Ten Bosch considers emotion recognition from a technical point of view, as a potential extension to existing work on recovering linguistic content from speech signals. His conclusions are skeptical: the benefits of attempting to incorporate

emotion into speech recognition paradigms remain to be demonstrated. For those who are convinced that the benefits are real, the paper stands as a reminder of the arguments that still need to be made.

It is fitting that the last word belongs to Klaus Scherer. His work has dominated research on speech and emotion for several decades, and his paper draws together the work done by himself, his colleagues, and the many groups influenced by him. The structure is provided by a model that he has advocated for many years, the lens model proposed by Egon Brunswik in the mid twentieth century. The model provides a unifying framework for discussing issues ranging from the intrinsic nature of emotion, to techniques for eliciting emotion; the factors that impinge on the expression of emotion; the statistical issues raised by emotion databases; and the methods appropriate for automatic recognition. To the best of our knowledge, no other source provides as comprehensive an introduction to the field.

The papers in this volume represent a sample of the topics and the discussions that took place at the ISCA Speech and Emotion Workshop at Newcastle. We hope that they will lead to further investigations of the areas where emotion and speech overlap, and that the Newcastle workshop will be remembered as the first in a continuing series of meetings dedicated to combining knowledge about the two subjects.

#### Acknowledgements

We thank the following colleagues who reviewed the different papers: Veronique Aubergé, Jo-Anne Bachorowski, Alan Black, Tim Bunnell, Rolf Carlson, Soojin Chung, Randy Cornelius, Grzegorz Dogil, Anders Eriksson, Christer Gobl, Daniel Hirst, Wolfgang Hess, David House, Jill House, Peter Howell, Adam Janin, Arvid Kappas, Inger Karlsson, Bernd Moebius, Alex Monaghan, Sylvie Mozziconacci, Ailbhe Ní Chasaide, Elmar Nöth, Roz Picard, Peter Regel-Brietzmann, Willibald Ruch, Klaus Scherer, Elizabeth Shriberg, Quentin Summerfield, Mark Tatham, Louis ten Bosch, Jacqueline Vaissiere, Raymond Veldhuis,

Anne Wichmann. We also thank the anonymous colleagues who reviewed papers authored by the editorial team.

Ellen Douglas-Cowie  
Roddy Cowie  
Nick Campbell