REPRESENTATION OF PROSODIC AND EMOTIONAL FEATURES IN A SPOKEN LANGUAGE DATABASE

Peter Greasley, Jane Setter†*, Mitch Waterman*, Carol Sherrard*, Peter Roach†, Simon Arnfield† and David Horton**

**Department of Psychology, University of Leeds, Leeds, LS2 9JT, UK*

†Department of Linguistic Science, University of Reading, Reading, RG6 2AA, UK

ABSTRACT

This article reports on research in progress on the Emotion in Speech project, funded by the UK ESRC and the Ministry of Defence (project No. R000235285). The ToBI transcription system is used to represent prosodic information. Additional layers have been created to code emotional speech according to the emotional lexicon, affective valence and cognitive appraisals. The aim is to produce a database of fully labelled emotional speech.

INTRODUCTION: THE PROJECT

The study of prosodic function in relation to grammar and syntax is a well-developed field, enjoying formalisms to represent linguistic information which are widely agreed. The prosodic characteristics of *emotional* speech are less tractable, and there is little agreement on the best method to analyse emotional speech or, indeed, how best to represent its specific features.

The Emotion In Speech project will transcribe 20 hours of naturally occurring emotional speech, eventually to be available as a CD-ROM database. The transcription system used to represent prosody is ToBI, devised by Silverman et. al. [1]. As ToBI is a multi-tiered system, we have created additional tiers to incorporate emotional information. The first four layers represent tones (tier one), words (tier two), break indices (tier three) and miscellaneous items, e.g., laughter (tier four). The fourth tier can be used to represent a broader range of prosodic and paralinguistic features than has been used in ToBI so far. Our coding system will be based on the work of Crystal [2]. The remaining four layers contain the emotional labels devised for the project. The aim is to collate multiple characterisations of the emotional response, i.e., prosodic features, judged emotion, lexical valency, appraisal category and cognitive antecedents; analysis should be consistent across levels. The resulting multi-layered database will allow more formal description of emotional speech, and improved understanding of the relationship between situation and communication.

EMOTION IN SPEECH

Due to the practical difficulties in obtaining records of naturally occurring emotional expression, the majority of data on emotion and its vocal correlates comes from studies which have either used actors [3], or have manipulated speech in some way using a computer [4, 5]. While using actors to simulate emotion may have methodological advantages, this approach does raise questions as to the ecological validity of the data. Actor portrayals of emotions may reproduce stereotypes [6] which stress the obvious vocal cues but "miss more subtle cues that further differentiate discrete emotions in natural expression" [7]. Williams & Stevens [8] compared contrived and natural speech and found the overall range of F_0 to be wider in contrived speech.

This approach also suffers from a simplistic and ambiguous labelling method, e.g., 'say x as if you are feeling angry/sad/happy etc.' which 1) neglects individual differences in the interpretation of verbal labels of emotion (the 'basic emotions' like *anger, joy* or *fear* can take different forms depending on the situation [7]), and 2) assumes that emotions are experienced as discrete psychological states whereas, in reality, they predominantly occur in complex blends [9].

It is important, then, that research on the vocal expression of emotion uses samples of speech which are as close to natural speech as possible. In our research, the primary source of data is television and radio (e.g., documentary programmes, talk/debate shows and sports commentary), though we are also seeking other sources. It is also clear that the coding will require a more sophisticated model to represent the varieties of emotional expression.

CODING EMOTIONAL SPEECH

In the psychological literature it is possible to discern three distinct approaches to the categorisation of emotion: 1) emotion labels (the affective lexicon), e.g., anger, rage, grief; 2) abstract dimensions of affect: i) valency (pleasant/unpleasant feelings); ii) potency (strong/weak feelings); iii) level of activity or arousal; 3) cognitive appraisals/antecedents which produce the emotional experience. Each of these approaches is utilised in our coding system.

The first layer (tier five in the modified ToBI system) is being used to record the judgments of listeners. Judges will be given a list of emotion labels from which to select those most appropriate and accurate in describing the emotion(s) expressed. (A number of recent publications provide taxonomies of the emotion lexicon [10]).

The second layer (tier six) is being used to code lexical valence. Osgood, Saporta & Nunnally [11] use a technique referred to as 'evaluative assertion analysis' which involves rating particular words or 'common meaning phrases' according to their valency and intensity. Extremes of valency should increase with emotionality of the source. More recently, Anderson & McMaster [12] and Bestgen [13] have coded the lexical valency of words and phrasal units to determine the emotional tone of literary documents. The occurrence of lexically valenced words, along with their frequency and intensity, provides a measure of the emotional force of the content.

Emotion labels can be notoriously imprecise in representing emotional states. It has been argued (Scherer [7]) that they are "rather unsuitable for the scientific description of affective states," and that "future work should use a conceptual system to describe emotional states and their antecedents that is more systematic than the natural language taxonomy of emotions ... the state referred to should be specified in terms of its underlying process" (p.146). In this respect, Ortony et. al. [14] provide a model of emotional experience based upon valenced reactions to events (pleased/displeased), actions of agents (approve/disapprove), or objects (like/dislike). Whether or not the valenced reaction is actually experienced as an emotion depends upon how intense the reactions are.

This model forms the basis of our third and fourth coding layers (tiers seven and eight). For the third layer we are using this model to identify the general reference category of cognitive appraisals, based on reactions to events, agents or objects. For example, if an event has occurred which the individual is displeased about, we have 'distress emotions'; if the individual disapproves of the actions of another person, we have 'reproach emotions'.

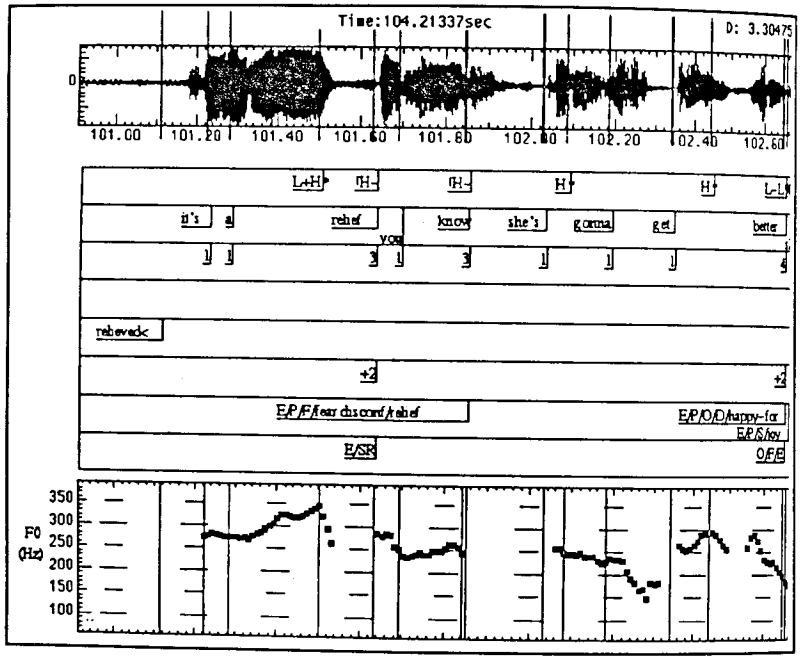

Figure1. Example of the coding system used in the Emotion in Speech project.

The fourth layer (tier eight) is adapted from Waterman's detailed qualitative analysis of interview data [15, 16], based on the theoretical position of Ortony et. al. [14]. From self-reports of emotional responses to emotionally loaded musical extracts, Waterman has devised a classification system to code cognitive antecedents of emotional responses. It can be applied to appraisal statements (i.e., that imply some reference to internalised representations) that need not necessarily be emotional. The purpose of the system is to inform deduction about the type of emotion being experienced by the speaker, and it provides greater detail than the layer above (the general reference category). In practice, application of the coding scheme is exhaustive for all possible appraisal possibilities in a statement. The system does not indicate actual emotion type, only the possible emotions, given the statement under examination.

An example of the coding system is provided in Figure 1. In this example we can see how the four coding tiers combine with the ToBI system to provide a multi-layered representation of emotion in speech. The first coding layer (tier five) provides emotion labels from the affective lexicon provided by judges who have listened to the speech. On the second layer we have two positive word valences: *relief, get better.* The third layer represents the speaker's cognitive appraisals (reactions that elicit the emotional expression), inferred from the speech content. In this case we have the *potential* emotional complex of: 'joy emotions', i.e., pleased about an event (lexical tokens include: delighted, glad, happy, pleased); 'happy-for emotions', i.e., pleased about an event desirable for someone else; and 'relief emotions', i.e., pleased about the disconfirmation of the prospect of an undesirable event. Finally, the fourth layer records every evidenced appraisal contained within the speech content (and thus provides a more detailed and comprehensive record of appraisals than the third, which simply targets particular valenced reactions within the text).

The coding system will, then, provide us with accurate emotion labels, the valenced direction of the emotion along with its intensity, and a sophisticated record of the possible emotions being experienced by the speaker, inferred from the appraisals evident within the verbal content.

REFERENCES

[1] Silverman, L., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992) ToBI: A standard for labelling English prosody, *Proceedings of the International Conference in Speech and Language Processing*, Alberta.

[2] Crystal, D (1969) *Prosodic Systems and Intonation in English*, Cambridge: Cambridge University Press.

[3] Scherer, K. R., Banse, R., Wallbott, H. G. & Goldbeck, T.(1991) Vocal cues in emotion ecoding and decoding, *Motivation & Emotion*, 15(2), 123-148.

[4] Uldall, E. (1961) Dimensions of meaning in intonation, in Bolinger, D. (ed) (1973) *Intonation*, Harmondsworth: Pengin, 250-259.

[5] Lieberman, P. & Michaels, S. B. (1962) Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech, *Journal of the Acoustical Society of America* 34(7), 922-927.

[6] Kramer, E. (1963) Judgment of personal characteristics and emotions from nonverbal properties of speech, *Psychological Bulletin* 60(4), 408-420.

[7] Scherer, K. R. (1986) Vocal affect expression: A review and a model for future research, *Psychological Bulletin*, 99, 143-165.

[8] Williams, C. E. & Stevens, K. N. (1972) Emotions and speech: some acoustical correlates, *Journal of the Acoustical Society of America* 52, 1238-1250.

[9] Ellsworth, P. C. & Smith, C. A. (1988) From appraisal to emotion: Differences among unpleasant feelings, *Motivation and Emotion*, 12(3), 271-302.

[10] Shaver, P. , Schwartz, J., Kirson, D. & O'Connor, C. (1987) Emotion knowledge: further exploration of a prototype approach, *Journal of Personality and Social Psychology*, 52, 1061-1086.

[11] Osgood, C., Saporta, S., & Nunnally, J. (1956) Evaluative Assertion Analysis, *Litera*, 3, 47-102.

[12] Anderson, C. W., & McMaster, G. E. (1982) Computer assisted modeling of affective tone in written documents, *Computers and the Humanities*, 16, 1-9.

[13] Bestgen, Y. (1994) Can emotional valence in stories be determined from words? *Cognition and Emotion*, 8(1), 21-36.

[14] Ortony, A., Clore, G. & Collins, A. (1988) *The Cognitive Structure of Emotions*, Cambridge: Cambridge University Press.

[15] Waterman, M. G. (1992) Emotion in Music: Towards a new methodology for the investigation of appreciation. *International Journal of Psychology*, 27 (3/4) 189.

[16] Waterman, M. G. (1994) Emotion in Music: Implicit and explicit effects in Listeners and Performers. In I. Deliege (Ed.) *Proceedings of the 3rd International Conference for Music Perception and Cognition.* Liege: ESCOM Publications.