



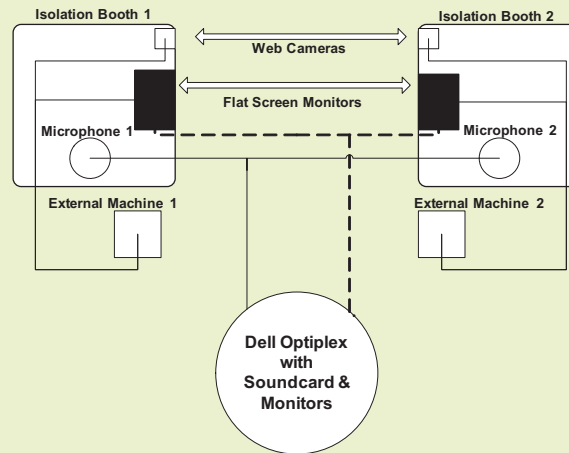
The Use of Task Based Mood-Induction Procedures to Generate High Quality Emotional Assets.



B. Vaughan , C. Cullen, , S. Kousidis , Wang Yi , C. McDonnell and D. Campbell
Cognition Speech & Audio Lab, DMC, DIT, Aungier Street, Dublin 2, Ireland
Email: brian.vaughan@dit.ie

ABSTRACT

Detecting emotion in speech is important in advancing human-computer interaction, especially in the area of speech synthesis [1]. This poster details experimental procedures based on Mood Induction Procedure 4 [2], using performance related tasks to engender natural emotional responses in participants. These tasks are aided or hindered by the researcher to illicit the desired emotional response. These responses will then be recorded and their emotional content graded to form the basis of an emotional speech corpus. This corpus will then be used to develop a rule-set for basic emotional dimensions in speech.



METHODS

In order to illicit natural emotional responses, the researcher aids or hinders the participants in their co-operative task. This is achieved by varying the instructions given or the materials or time available. The use of sound isolation booths serves to prevent outside distractions from affecting the participants as well as allowing the dialogue to be recorded as two separate high quality audio streams, specifically at a 24 bit/192 KHz [7] professional resolution. These high quality audio assets obtained in a properly isolated acoustic environment will allow a robust and effective analysis to be performed with reliable acoustic results. The assets obtained will subsequently be graded according to their emotional content, and then assessed so that they can be segregated into relevant emotional dimensions.

INTRODUCTION

Research has shown that there is potential for grading the basic emotional content of natural speech using the dimensions of activation and evaluation [3]. There is considerable debate concerning the methods used for obtaining emotional content, and distinction is made between natural [4] and acted speech [5], mainly as regards to their level of authenticity. In this research, Mood Induction Procedures (MIP's) [2] are used to stimulate emotions in controlled situations. In these experiments, participants sit in isolation booths and must co-operate to complete timed tasks [6].

STATUS OF WORK

Various experimental designs were considered with three final designs being devised. The three are based on: Tetris, Mia/Kuriki (a dice game) and the construction of a Lego structure (first devised by Kehrin [9]). These three are currently being tested to ascertain which one experiment or combination of experiments are the most successful in engendering emotional states in participants.

OBJECTIVES

- To obtain audio recordings at a professional quality [7]
- To record realistic natural emotional responses
- To define the emotional dimensions of the recordings [8]
- To analyse the recordings to develop a rule set for basic emotional dimensions in speech.

REFERENCES

- Schroeder, M., Speech and Emotion Research. An Overview of Research Frameworks and a Dimensional Approach to Emotional Speech Synthesis, in Faculty of Philosophy. 2004, Universit'at des Saarlandes. p. 288.
- Gerrards-Hesse, A., K. Spies, and F.W. Hesse, Experimental inductions of emotional states and their effectiveness: A review. *British Journal of Psychology*, 1994. 85(1): p. 55-78.
- Cowie, R., et al., Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 2001. 18(1): p. 32-80.
- Campbell, N. Databases of emotional speech. in *ISCA Workshop on Speech and Emotion*. 2000. Northern Ireland.
- Roach, P., et al., Transcription of Prosodic and Paralinguistic Features of Emotional Speech. *Journal of the International Phonetic Association*, 1998(28): p. 83-94.
- Kehrein, R. The prosody of authentic emotions. in *Speech Prosody*. 2002. Aix-en-Provence, France.
- Watkinson, J., *Art of Digital Audio*, ed. M. Burlington. Vol. Third Edition. 2000: Focal Press.
- Cowie, R., et al., 'FEELTRACE': An instrument for recording perceived emotion in real time. in *ISCA Workshop on Speech and Emotion*. 2000. Northern Ireland.
- Kehrein, R. The prosody of authentic emotions, in *Speech Prosody*. 2002. Aix-en-Provence, France.

RESULTS

There are a number of methods [3] that can be used to evaluate the emotional dimensions of the recorded audio, but listening tests are considered to be the most statistically robust. The listening tests will be carried out by participants who did not take part in the co-operative task based experiment. The assets will be graded using the FeelTrace [8] tool in order to specify the emotional dimension of each clip.

CONCLUSION

Using the FeelTrace tool to analyse and grade the emotional content of the audio clips allows for a justifiable set of results that can then be used to specify the emotional dimension of the clips in the corpus. These emotionally graded assets can then be analysed further in order to create a rule set for basic emotional dimensions in speech.