

EMOTIONAL SPEECH DATABASE:

This database is intended for studies in dimensional affect recognition in speech.

The database consists of 59 clips in total from 11 movies with durations between 5 and 25 seconds. The clips include audio and text in the form of subtitles. The audio clips are resampled at 16 kHz. Since the loudness of movies may differ and the loudness is one of the important acoustics features, the long-term loudness (Please note that this does not really effect the instantaneous loudness) of the clips has been normalized using Replay Gain. Replay Gain is an open standard loudness calculation algorithm [1] in which the main idea is to calculate the gain needed on an audio file to match the perceived loudness level of a reference audio file.

The clips are rated in valence and arousal dimensions. Considering a format where the rating 1 represents 'not happy' and 9 'represents' 'happy', if the subjects treat 'not happy' as 'sad', it is bipolar. However, if the aim is a 'unipolar' format where 'not happy' means 'neutral', then this confusion might bring considerable error. To avoid these kinds of errors and to coincide with the ANEW work in [2] whose results are used in the semantics part of our work, we use the same response format and similar instructions they used. The bipolar scale with ratings between 1 and 9 for both dimensions are used with SAM figures [3]. The subjects are asked to rate the emotion expressed in the clips, not the emotion induces at themselves.

A Java applet has been designed for the experiments to get the emotional ratings. In this applet, there are 3 experiments to be carried out by the users and we ask them to fill in some personal information for statistical purposes and a questionnaire at the end to get some feedback. In the first experiment, the clips include just text, in the second just audio and in the third both text and audio. The order of the clips are toggled before each experiment. The experiments take around 1 hour in total. The applet has been made available online [4] thus the users did the experiments wherever they wanted. However, they were asked to do it somewhere not very disturbing and without giving long breaks (the time track of the users have been taken thorough the applet and been checked). We recruited 13 people (7 female, 6 male), between ages 19 and 28, speaking English fluently. They all claimed that the instructions were clear and they were confident rating.

The Figure 1 shows the arithmetic mean values of the ratings of all clips for the three experiments.

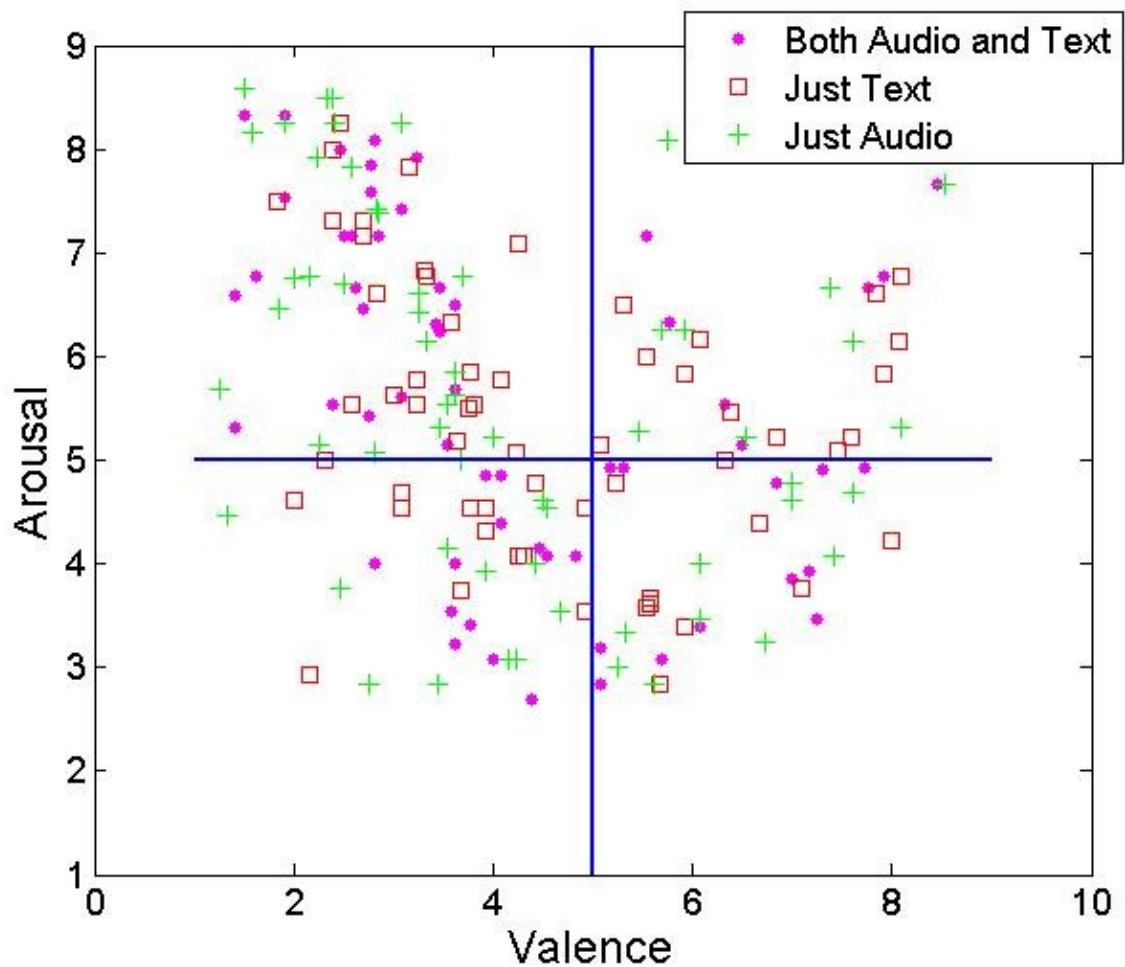


Figure 1

References:

- 1) D. Robinson, "Replay gain," 2001, <http://www.replaygain.org/>.
- 2) M. Bradley and P. Lang, "Affective norms for english words (anew): Instruction manual and affective ratings," *University of Florida: The Center for Research in Psychophysiology*, 1999.
- 3) M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.
- 4) S. G. Karadogan, "Emotional data set creation," 2011, <http://www.student.dtu.dk/~seka/>.