

USING EMOTION TO GAIN RAPPORT IN A SPOKEN DIALOG SYSTEM

Jaime C. Acosta and Nigel Ward
Department of Computer Science
University of Texas at El Paso

Although spoken dialog systems are becoming more widespread, their application is today limited largely to domains involving simple information exchange. To enable future applications including customer support, therapy, negotiation and persuasion, new capabilities are needed. One barrier to the creation of such applications is the lack of methods for building rapport between spoken dialog systems and human users; another is the inability to model the emotional and interpersonal aspects of dialog. This research addresses the question of whether a spoken dialog system that recognizes human emotion and reacts with appropriate emotion can successfully gain rapport with humans.

Our specific domain is that of persuasion, as seen in a corpus of 10-20 minute dialogs between a graduate coordinator and undergraduates, in which the coordinator was attempting to persuade them that going to graduate school was an option worth considering. Although much of each dialog was involved in conveying factual information, there was also a heavy use of what appear to rapport-building strategies, and clear variation over the utterances of both coordinator and students in emotional coloring and prosodic properties, including pitch, timing, and volume.

Various models of human dialog, notably Communication Accommodation Theory, describe how prosody and other nonverbal behaviors are used in communication to adjust social distance between interlocutors. For example, when meeting a person who displays sadness in voice, someone wanting to reduce social distance may modify their intonation, speed, and loudness in voice in order to sound emphatic. Implementing such behavior patterns requires the ability to detect emotions from voice, not so much the classic emotions such as sadness, anger, and joy, but the more subtle emotions that are more common in normal spontaneous conversations. For this reason we represent emotions using a dimensional approach, using the three dimensions of activation (active/passive), evaluation (positive/negative), and power (dominant/submissive), each represented by continuous value ranging from -100 to +100.

So far we have used machine learning methods to develop a module able to predict, from the acoustic-prosodic features of each utterance, the emotional coloring of that utterance in terms of the three dimensions. Verification with respect to 962 utterances from the corpus, each annotated by two judges, indicates that the quality of these predictions may be adequate for a rapport-building system. We have also identified a number of correlations between the emotion of the student in one utterance and the emotion of the graduate coordinator's subsequent utterance. We next plan to tune a speech recognizer for this domain, and to develop a method for choosing what lexical content to produce in each situation. We will then integrate these components to produce a dialog system able to interact with students and persuade them to consider planning to go to graduate school. This will be the first dialog system to use emotion in voice to build rapport with users.