Identifying 'stance taking' cues to enable sophisticated voice recognition

October 28 2014

In the future, computers may be capable of talking to us during meetings just like a remote teleconference participant. But to help move this science-fiction-sounding goal a step closer to reality, it's first necessary to teach computers to recognize not only the words we use but also the myriad meanings, subtleties and attitudes they can convey.

During the 168th Meeting of the Acoustical Society of America (ASA), to be held October 27-31, 2014, at the Indianapolis Marriott Downtown Hotel, Valerie Freeman, a Ph.D. candidate in the Department of Linguistics at the University of Washington (UW), and colleagues will describe their National Science Foundation-sponsored work for the Automatic Tagging and Recognition of Stance (ATAROS) project. The project's goal is to train computers to recognize the various stances, opinions and attitudes that can be revealed by human speech.

"What is it about the way we talk that makes our attitude clear while speaking the words, but not necessarily when we type the same thing? How do people manage to send different messages while using the same words? These are the types of questions the ATAROS project seeks to answer," explained Freeman.

Identifying cues to "stance taking" in audio recordings of people talking is a good place to start searching for answers, according to Freeman and the principal investigators on the project, including Professors Gina-Anne Levow and Richard Wright in the Department of Linguistics, and Professor Mari Ostendorf in the Department of Electrical Engineering.

"In our recordings of pairs of people working together to complete different tasks, we've found they tend to talk faster, louder and with more exaggerated pitches when expressing strong opinions as opposed to weak opinions," Freeman said.

Not too surprising? Maybe not in terms of heated arguments, but the researchers found the same patterns within ordinary conversations, too. "People talk faster and say more at once when they're working on more engaging tasks such as balancing an imaginary budget as opposed to arranging items within an imaginary store," Freeman noted.

The researchers' also noticed that people also appear to be less fluent in the engaging tasks—displaying more false starts, cut-off words, "ums" and repetitions.

Further, it appears that "men might do this more than women—regardless of whether they're talking to another man or a woman." Freeman places a heavy emphasis on the word "might," because to date they've only explored this particular lack of fluency with 24 people.

So far, for the entire project, the researchers have worked with and recorded a total of 68 people of varying ages and backgrounds, all from the Pacific Northwest.

"We plan to continue to analyze these conversations for subtler cues and more complex patterns—variations in pronunciations when comparing positive and negative opinions, men vs. women, and older vs. younger people," said Freeman. "In the future, we hope to record people from other locations to see whether different regions have different ways of expressing the same opinions."

The lessons learned from this work should help enable sophisticated

speech recognition systems of the future. "Think of all of the amazing things the computer on Star Trek can do," Freeman said. "To reach that level of sophistication, we need computers to understand all the subtle parts of a message—not just the words involved. Projects like ATAROS are working to help computers learn how to figure out what people really mean when they speak, so that in the future computers will be capable of responding in a much more 'human-like' manner."

More information: Presentation #2pSC18, "Phonetic correlates of stance-taking," by Valerie Freeman, Richard Wright, Gina-Anne Levow, Yi Luan, Julian Chan, Trang Tran, Victoria Zayats, Maria Antoniak and Mari Ostendorf will be shown during a poster session on Tuesday, October 28, 2014, from 1:00 to 5:00 PM in Marriott 5. The abstract can be found by searching for the presentation number here: https://asa2014fall.abstractcentral.com/planner.jsp

Provided by Acoustical Society of America