Can a spoken dialog system be used as a tool to study convergence?

José Lopes[1,2]*, Andrew Fandrianto*[3]*, Maxine Eskenazi*[3] *and Isabel Trancoso*[1,2]
[1]*Instituto Superior Técnico, Lisboa, Portugal*
[2]*INESC-ID Lisboa, Portugal*
[3]*Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, USA*
jose.david.lopes@l2f.inesc-id.pt

Abstract

The finding that people entrain to one another in a conversation Brennan (1996) has fostered much interest in this phenomenon within a variety of research communities, such as psychology. Members of the automatic speech processing community have viewed it as a potential functionality that, if present in human-machine interaction, could be capitalized upon to improve system performance Lopes et al (2011). Beyond the benefits to SDS research, we argue here that automated systems can, in turn, benefit research in other areas. We believe that, for the study of entrainment, SDS can provide platforms on which to run studies, offering more control over conditions in some ways that do human-human studies. We use the term entrainment here, from Brennan and others. This term may represent the action of one of the speakers. Assuming both speakers entrain, there should be convergence.

The literature does show that humans can be made to change their speech patterns to imitate the output of a spoken dialog system (SDS). Stoyanchev and Stent (2009) used a set of dialogs to study entrainment using two verbs and two prepositions as primes. They confirmed that callers can adapt their choice of terms to the terms used by the automated system. In Parent and Eskenazi (2010) the system primes were directly manipulated in the Let's Go spoken dialog system (real, not paid callers, Raux et al (2005)) and observed caller adaptation over time. The authors found that users do adapt and are more likely to do so in the first few turns following the first appearance of the prime. The same was done in European Portuguese with the Noctívago spoken dialog system Lopes et al (2011) with the result that the enlisted callers entrained to all of the primes that were proposed by the system.

Despite confirming the presence of entrainment, not all proposed primes were copied. Looking in more detail on the lexical and prosodic levels, both Lopes et al (2011) and Parent and Eskenazi (2010) found differences in how often words were copied. Less frequent words were copied less frequently if they were new primes (and the system was already offering a very frequent prime, for example, “help” > “assistance”) and that conversely, if an infrequent word had been used and was replaced with a more frequent prime, the latter was easily copied (“start a new query” to “start a new request”). Lopes et al (2011) observed in Noctívago that if a very frequent and contextually appropriate word had been used (like "agora", now) it would continue to be used whether the system still used it in its prompts or not. But the primes proposed here, for example “imediatamente” (immediately) and “neste momento” (right now), are all much longer and not necessarily more natural than “agora”. Neither study found any influence of the part of speech on the likelihood to be copied. Both studies confirmed that continued exposure to the primes increases the likelihood of their uptake.

The individual choices may not, for some words, follow the lexical frequency in the language. This can be due to individual preference, local uses, professional uses or a myriad of other reasons. In a relatively short dialog, like

the examples presented here, it would be difficult to adapt to these individual differences. If a dialog system was to be used by the same person over a longer period of time, this would be possible. And choices that may be made due to avoidance of difficult phonetic clusters (as in foreign words) can be dealt with automatically.

Entrainment on the prosodic level was further analyzed. In an attempt to get callers to stop shouting or hyperarticulating, the system spoke more softly or more slowly, respectively. It was observed that callers more frequently copied the first condition than the second. In this case, the system was adjusted to speak precisely 25% faster (measured in syllables per second) or 25% softer. This type of precise control would be difficult to obtain in human-human studies even if one speaker was instructed to speak 25% softer. Given the training and tuning, a limited domain speech synthesizer can vary elements like speaking rate, pitch variability and contour, rhythm, and intensity with great precision.

Interestingly, to our knowledge there has not yet been a study that compared entraining to an SDS and entraining to another human on a similar task with similar constraints. This could help further our understanding of the differences in the two conditions. Armed with this knowledge, some studies could be carried out where the appearance of the prime is tightly controlled using an SDS and have some way to relate the findings above described to what humans might do when speaking to one another.

There are several other benefits to the use of an SDS in this area. Running studies on an SDS with real users reduces long term cost and increases scalability. While one laboratory study may painstakingly find 50 participants, running the study over a week or two, an SDS with real users, as in the case of the Let's Go platform, can get over 500 in the same time frame. These callers receive no remuneration other than getting bus scheduling information, thus additionally reducing costs.

We have seen that spoken dialog systems can offer controlled conditions for studies on how humans copy speech. We believe that these types of platforms should be considered as one of many tools that those who study entrainment can use.

Bibliography

Brennan, Susan E (1996). *Lexical entrainment in spontaneous dialog*. Proceedings International Symposium on Spoken Dialog.

Lopes, José, Eskenazi, Maxine and Trancoso, Isabel (2011). *Towards Choosing Better Primes for Spoken Dialog Systems*. Proceedings ASRU 2011.

Parent, Gabriel and Eskenazi, Maxine (2010). *Lexical Entrainment of Real Users in the Let's Go Spoken Dialog System*. Proceedings Interspeech 2010.

Raux, A., Langner, B. Bohus, D. Black, A W, Eskenazi, M. (2005). *Let's Go Public! Taking a Spoken Dialog System to the Real World*. Proceedings Interspeech 2005.

Stenchikova, S and Stent, A (2007). *Measuring adaptation between dialogs*. Proceedings SIGdial 2007.