Home-made speech synthesis for non-English-speaking patients

H Somers, G Evans, Z Mohammed
School of Informatics
University of Manchester
PO Box 88
Manchester M60 1QD

Key words: natural language processing, speech synthesis, patient-provider interaction

Abstract
This poster concerns the development of computer-based support for health-care seekers with limited English, focusing on speech synthesis and on languages for which such technology has not been developed. Speech synthesis (or text-to-speech – TTS) systems are available only for the world's major languages. In the absence of such technology, we want as a stop-gap solution to use an existing TTS system for a major language to 'fake' TTS for a minor language. We are (at the time of submission) setting up an experiment to evaluate the usability of such a solution for Somali. In the experiment a number (aim: 25) of Somalis with limited or no English will participate in simulated patient–provider interviews. Two approaches of differing complexity will be tested and compared with recordings of human speech. Participants will be played 25 examples of synthetic or recorded human Somali speech and asked to report to a Somali-speaking experimenter what they understand the computer to be saying.

We have been looking at the use of language technology at various points on the pathway to health-care for patients who have little or no English and may also be functionally illiterate[1,2]. In the work presented here we concentrate on the use of speech synthesis (or text-to-speech – TTS) to provide spoken messages in the patient's own language in the context of a patient–provider interview.

TTS systems, which read out ordinary text on the computer, are now fairly widespread and are sufficiently reliable and of a suitable quality

for wide acceptance and use. However, this is only true for the major languages. Currently, development of a TTS system depends on:

- an extensive phonological analysis of the language;
- development of rules to identify how the written form of the language relates to the spoken, and rules to determine the pitch and duration features for stress and intonation; and, depending on the approach taken
- recording of hundreds of individual speech elements or modelling of a similar number of elements using a formant synthesizer.

For many languages, it is likely that economic factors will prevent such work from being undertaken. We want therefore, as an interim solution, to try using an existing major-language TTS system, as is, to 'fake' TTS for a minor language. Our experiment consists in trying to find the most appropriate base language for a given target language, and then evaluating its usability. Our chosen target language is Somali, the scenario that of various patient–provider interactions relating to asthma diagnosis and management.

Most TTS systems consist of two elements: a text-to-phoneme stage, where the basic pronunciation of the text is determined, and a phoneme-to-speech stage, where the actual speech sounds are generated. The first stage involves identifying the phonemes to be uttered, as well as the pitch and duration, in order to produce appropriate intonation and stress (prosody). If we are faking TTS we have to deal first with differences in the letter-to-sound mapping rules between the base and target languages. In the second stage, the actual speech sounds are generated, whether by concatenating prerecorded human speech or by formant synthesis. The main problem for faking it is differences in the set of speech sounds between the TL and the BL. The trick is to choose a BL where this problem is minimized, though there may be a considerable trade-off between finding a BL with a good overlap in the letter-to-sound mapping rules mentioned earlier, versus good coverage of the target phonemes. Taking the above considerations into account, we plan to evaluate two alternative methods of faking Somali TTS. For both methods we first had to identify the most suitable BL for Somali (we are experimenting with German) and then had to map Somali spelling rules onto German.

In the first of the two methods we took an existing readily available German TTS system, and gave it Somali text which had been passed through a simple transliteration program. For example, the Somali text *Ma qufcaysaa?* 'Are you coughing?' is transliterated into 'German' as

<ma kuf-ässa?>. The second method is based on earlier work with this idea[3]: we again transliterate the Somali input, this time identifying the German phonemes to be synthesized. This approach is slightly more complex than the previous one, since *all* the Somali text has to be transliterated rather than just the letters whose pronunciation is different. The treatment of prosody also differs in the two methods.

Our evaluation methodology is based on our previous work evaluating speech translation[4], and focuses on the subject's ability to infer correctly the intended meaning of the utterance. Experiments are being conducted at the time of submission. In simulated patient–provider dialogues, subjects are told to imagine that what is said to them has been translated and will be spoken by the computer. They are first given some training examples of the computer's speech, then are asked to listen to the speech samples, and to say what they understood. Five different scenarios are presented as contextualisation, each consisting of five phrases, giving a total of 25 phrases. The different methods of faking TTS will be mixed together with recordings of human speech. We will report on the results of these experiments, including any general reactions and opinions from the subjects. Pilot experiments suggest that faking it is moderately successful, but various important factors are emerging that determine the best way to implement it.

References

1. Somers H, Lovel H. Computer-based support for patients with limited English. Proc 7th Internat EAMT Workshop "MT and other language technology tools", Budapest 2003, :41–9.
2. Somers H, Lovel H, Johnson M, Mohamed Z. Language technology for patients with limited English. EACH Internat Conf Comm Healthcare, Bruges, 2004, :04-09.
3. Evans G, Polyzoaki K, Blenkhorn P. An approach to producing new languages for talking applications for use by blind people. In: Miesenberger K, Klaus J, Zagler W (Eds), 8th ICCHP, Computers Helping People with Special Needs, Springer Verlag, 2002, :575–82.
4. Somers H, Sugita Y. Evaluating commercial spoken language translation software. Proc 9th Mach Transl Summit, New Orleans, 2003, :370–7.