

Faking it: Synthetic text-to-speech synthesis for under-resourced languages – Experimental design

Harold Somers

School of Informatics

University of Manchester

Manchester, UK

Harold.Somers@manchester.ac.uk

Abstract

Speech synthesis or text-to-speech (TTS) systems are currently available for a number of the world's major languages, but for thousands of the world's 'minor' languages no such technology is available. While awaiting the development of such technology, we would like to try the stop-gap solution of using an existing TTS system for a major language (the base language) to 'fake' TTS for a minor language (the target language). This paper describes the design for an experiment which involves finding a suitable base language for the Australian Aboriginal language Pitjantjajara as a target language, and evaluating its usability in the real-life situation of providing language technology support for speakers of the target language whose understanding of the local majority language is limited, for example in the scenario of going to the doctor.

1. Introduction

Speech synthesis systems, in particular text-to-speech (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. For example, Microsoft's *Agent* includes American and British English, Dutch, French, German, Italian, Japanese, Korean, Portuguese, Russian and Spanish. Scansoft's *RealSpeak* provides for all of the above, plus Basque, Cantonese, Mandarin, Danish, two varieties of Dutch, Australian and Indian English, Canadian French, Norwegian, Polish, and two varieties of Portuguese. The list is impressive, but there are still thousands of languages not covered.

Our interest is in providing language technology-based support for speakers of 'minor' languages¹ when they find themselves in situations

where their lack of ability in another language is a significant disadvantage: we have been focusing on the case of newly arrived immigrants seeking healthcare (a visit to the doctor), but the possibilities are almost endless. This is the scenario envisaged by the CANES framework,² as described by Somers and Lovel (2003) and Somers et al. (2004). This envisages software support for getting general information about healthcare problems, arranging an appointment at the clinic or hospital, understanding information leaflets and instructions regarding treatment and drugs, and of course face-to-face meetings with healthcare providers, notably GPs and nurses. Other projects in this field have focused on spoken language translation (SLT) of the doctor-patient interview (Rayner et al., 2003; Narayanan et al., 2004). We have recognized that, while face-to-face dialogues have an important role in the pathway to healthcare, other means of communication play an equally important role, some of them text-based. In all these cases, we see TTS as an essential technology, particularly for users who not only may have limited or no English, but also whose reading ability in their own language may be poor, whether due to low literacy, dyslexia or visual impairment.

A long-term solution is of course to develop TTS tools for more languages, but this is by no means trivial. Currently, development of a TTS system depends on an extensive phonological analysis of the language to identify the individual speech sounds (phonemes) and their variants (allophones); development of text-to-phoneme rules to identify how the orthography of the language relates to the phonology, and rules to determine the pitch and duration features (prosody); and, depending on the approach taken, recording of human speech and extracting

'major' language, as (extensionally) defined above. This is not to be confused with a 'minority' language, since the list of minor languages includes some of the most widely spoken languages in the world (e.g. lists of languages with the most speakers include Bengali, Hindi/Urdu and Arabic in the top 5 or 6).

² Computer Assistance for Non-English Speakers.

¹ A 'minor' language is any language which is not a

hundreds of individual speech elements (diphones, triphones or demisyllables) or modelling a similar number of elements using a formant synthesizer.

While waiting for this work to be done, in the meantime we want to try using an existing major-language TTS system, as is, to fake TTS for a minor language, in this case, the Australian Aboriginal language Pitjantjatjara.³ We are undertaking a similar experiment with Somali (Somers et al., 2006)

The idea has been briefly explored by Evans et al. (2002), who have dubbed the process ‘gibbering’,⁴ whereby speech synthesizers for new languages that are suitable for use with a screen reader are produced with a minimum amount of development time, and can be made available at no cost to the user. They suggest that

... the minimum requirement is that the speech has to be consistent and understandable, but does not necessarily have to be the especially natural sounding or indeed linguistically accurate. The key requirement is that the speech synthesiser speaks the language consistently and can be fully understood by a speaker of the language. (p. 576)

2. Text-to-speech synthesis

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. It is beyond the scope of this paper to describe detail how TTS works, but we do need to explain how the basic design of TTS systems relates to faking it.

2.1. Text to phonemes

The first stage involves identifying the phonemes to be uttered, but also the pitch and duration, in order to produce appropriate prosody (intonation and stress). This is generally done on the basis of letter-to-sound mapping rules, together with a dictionary where any irregular cases are made explicit. As well as cases of anomalous spellings,⁵

³ This work was initiated while the author was on study leave at the Centre for Language Technology, Macquarie University, Sydney.

⁴ To ‘gibber’ is to talk foolishly, or in a way that cannot be understood, hence ‘gibberish’. In our case, the input to the TTS system is gibberish from the system’s point of view, but, we hope, meaningful to the listener.

⁵ It is interesting that most laypersons usually assume that if a word is not pronounced as it is spelt it is because the pronunciation is anomalous; however, since writing is a representation of speech and not vice versa, it is of course the written form that is ‘wrong’. Whichever view is taken, it is clear that we must have a list of words for which the ‘rules’ do not apply.

the dictionary must ‘spell out’ abbreviations, numbers, symbols (e.g. &)⁶ and so on. The text-to-phoneme module must also contain rules that indicate unusual readings for sequences of symbols, e.g. \$5 is pronounced <five dollars>, not <dollar five>. In addition, most languages have homographs, the pronunciation of which may need more or less sophisticated syntactic analysis to determine. This analysis may also contribute information about prosody.

The problems with this module for faking it arise first from differences in the letter-to-sound mapping rules between the language for which the TTS system was designed (henceforth, following Evans et al. (2002), the ‘base language’, BL) and the language we are trying to produce (the ‘target language’, TL). For example, while a *j* is pronounced [dʒ] in English, in Spanish it is [x]. Further, some words in the TL may be written the same way as words in the BL, but pronounced differently, e.g. *train* in English [tɹeɪn] and French [tʀɛ̃]. The rules for reading symbols are generally language-specific, as are any rules relating to prosody.

It may be of course that the TL does not use the same writing system as the BL, or does not have a writing system at all.

In all the above cases, one thing we can do is to try to rewrite the TL word so that it follows the rules of the BL, for example rewriting French *train* as <tran> (though see next section on phoneme-to-speech).

Some TTS synthesizers accept as input streams of phonemes instead of plain text. Depending on the software, these may be in a kind of transcription (e.g. <dh-ah-s> for *thus*), or IPA symbols may be used. In their approach, Evans et al. (2002) go somewhat further:

The rules [for text-to-phoneme translation] are contained in a text-based table and applied by a generic piece of software that is capable of applying any appropriately specified rule set. Thus, to construct a new language the text to phoneme rules for the target language need to be developed and encoded in a table. (p. 578)

In fact, this approach is a significant alternative to our method, with some advantages and disadvantages. They have to define the mapping for all the phonemes, whereas we only define the ones that are different. But they

⁶ In this paper the following conventions are used: text strings and graphemes are shown *in italics*; strings to be read out by the TTS system are shown <in angle brackets>; as is customary, IPA phonetic symbols are shown in square brackets [ð□s], while slant brackets // are used to indicate phonemes.

potentially have more control over intonation, and can define a dictionary of special pronunciations, e.g. for the numbers and symbols.

2.2. Phonemes to speech

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 that the set of speech sounds for the TL will almost certainly not be the same as for the BL, and even if they are similar, it is likely that the rules for allophonic variation will be different. Some phonemes from the TL will simply be missing; others may be similar to phonemes in the BL, but may differ in the realization of different allophones.

The trick is obviously 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 in the previous section, versus good coverage of the target phonemes and allophones.

One of the goals in the research described here is to try to evaluate which of a number of BLs is most suitable for a given TL, and to identify which factors make identifying the best BL easier.

3. Faking it for Pitjantjatjara

Pitjantjatjara is an Australian Aboriginal language, spoken by the Anangu people, best known to tourists as the traditional owners of the land which includes Uluru, formerly known as Ayers Rock, in Northern Territory. Pitjantjatjara has about 1,300 speakers, but is one of a group of mutually intelligible dialects which form the Western Desert language group which, with around 4,000 speakers is one of the three or four ‘largest’ Aboriginal languages. According to Eckert and Hudson (1991), Pitjantjatjara has been written in the Roman alphabet since the 1940s, and the orthography has been more or less standardized since the 1979 meeting of Pitjantjatjara literates, and confirmed by the publication in 1987 of a Pitjantjatjara–English dictionary (Institute for Aboriginal Development, Alice Springs).

3.1. Pitjantjatjara phonology

Like many Aboriginal languages, Pitjantjatjara has a relatively simple set of consonant phonemes (Table 1). Five places of articulation are used: bilabial, alveolar, retroflex, palatal and velar, with a single plosive and nasal phoneme at each place. There are three lateral phonemes, and three approximants, plus a trill or flap alveolar /r/ sound. There are no fricatives. As in other Aboriginal languages, there is no phonemic distinction between voiced and voiceless consonants.

	bilabial	alveolar	retroflex	palatal	velar
plosive	[p] <i>p</i>	[t] <i>t</i>	[ɽ] <i>ɽ</i>	[c] <i>tj</i>	[k] <i>k</i>
nasal	[m] <i>m</i>	[n] <i>n</i>	[ɳ] <i>ɳ</i>	[ɲ] <i>ny</i>	[ŋ] <i>ng</i>
trill/flap		[r] <i>r</i>			
lateral		[l] <i>l</i>	[ɭ] <i>ɭ</i>	[ʎ] <i>ly</i>	
approximant	[w] <i>w</i>		[ɻ] <i>ɻ</i>	[j] <i>y</i>	

Table 1: Pitjantjatjara consonants: phonetic symbol in [] and standard orthography in *italics*.

Pitjantjatjara has six vowel phonemes, long and short [i], [u] and [a]. In the orthography, long vowels are doubled, *ii*, *uu*, *aa*. Syllables are highly constrained: all syllables are (C)V(C): there are no consonant clusters except across syllable boundaries. Most word-final syllables are CV. Word stress in Pitjantjatjara is quite regular, always on the first syllable, with subsequent syllables receiving equal prominence.

3.2. Choosing a BL for faking Pitjantjatjara

Perhaps the most difficult part of the experiment is to choose the best BL for our fake Pitjantjatjara TTS system. Nowadays, as mentioned in the first paragraph, developers of TTS systems are marketing more and more versions, often of surprisingly good quality. Practically speaking, we cannot test all of them on native speakers, so we need to have some criteria to enable us to narrow down the field. Three factors, possibly conflicting, will guide our decision: phoneme sets, comparative prosody, and orthographic mapping.

3.2.1. Phoneme sets

Perhaps the most obvious requirement is to find a BL which as nearly as possible covers the same set of phonemes as the TL. There are two sides to this problem: (a) when the TL phoneme has no equivalent in the BL, or (b) when the ‘equivalent’ phoneme is significantly different.

The first case may not be as important as it may seem: Pitjantjatjara for example has retroflex and palatal consonants which we may not find in any of the candidate BLs. But we can find a way around this: *sequences* of BL phonemes may sound sufficiently like the target phonemes, e.g. <rd> for [ɽ]. Alternatively, if the BL is forced to conflate a phonemic distinction, this may result only in the synthetic speech ‘having an accent’, rather than rendering it incomprehensible.

Possibly more damaging will be the second scenario. Thinking again of the ‘r’-like sounds in Pitjantjatjara [r ɻ] plus the retroflex sounds [ɽ ɳ ɭ], BLs such as French and Portuguese which have an /r/ phoneme realised as a uvular fricative are probably not going to be suitable. Vowel

phonemes could be a significant problem for other applications, though fortunately for us the Pitjantjatjara vowel system is about as simple as any found in all the languages of the world.

3.2.2. Prosody

Prosody plays a big part in how realistic a TTS system is judged to be, and developers of TTS systems work very hard to get this right. Even when given ‘gibberish’ to read, good TTS systems will do so with the appropriate intonation and accent, which can be very distinctive. The typical prosody for Pitjantjatjara is rather flat, so any BL with a wide-ranging prosody, such as Swedish and Danish, may well be ruled out. The stress patterns of Pitjantjatjara are also very simple, with word-initial stress, and evenly stressed syllables. This makes languages like English, which almost never has word-initial stress on long words, quite unsuitable. Intuitively, of the BLs available, Basque turns out to be a better bet, with its flat intonation and syllable-timed rhythm, despite the fact that Basque words are said to be stressed word-finally.

3.2.3. Text-to-phoneme mapping

Against the phoneme and prosody match is the question of the built-in text-to-phoneme mapping rules. Again there are two sides to this: How are the individual letters in the BL pronounced? And what is the likelihood of TL *words* being the same as BL words, but with a different pronunciation?

Since Pitjantjatjara orthography is based on English, there is in general a good text-to-phoneme mapping between English and Pitjantjatjara. The writing system is also largely ‘phonetic’, by which we mean that there is a simple mapping between letters and phonemes. Thinking of other BL–TL pairings, typical ‘problem letters’ are *c, g, j, v, w, x, y* and *z*, and all vowels, plus digraphs. This problem is minimised for Pitjantjatjara, since only 12 consonant letters and 3 vowel letters are used. The writing system has 4 digraphs (*tj, ny, ly, ng*), the 4 underlined letters representing retroflex sounds (*t*, *n*, *l*, *r*), plus the three long vowel digraphs. None of the digraphs found in English (such as *ch, sh, th*) appear in (native) Pitjantjatjara words.

Bearing in mind the comments relating to prosody, if we choose Basque as the BL, we have to consider *its* orthography. Fortunately, Basque too has only recently had its orthography standardized (in 1964), and so we find that the letter-to-phoneme mappings are quite straightforward. Table 2 shows the mapping between Pitjantjatjara phonemes and Basque spelling. Note that the letter *y* has to be avoided (it only occurs in

<i>p</i> <p>	<i>t</i> <t>	<u><i>t</i></u> <rt>	<i>tj</i> <tt>	<i>k</i> <k>
<i>m</i> <m>	<i>n</i> <n>	<u><i>n</i></u> <rn>	<i>ny</i> <ñ>	<i>ng</i> <ng>
	<i>r</i> <rr>			
	<i>l</i>	<u><i>l</i></u> <rl>	<i>ly</i> <ll>	
<i>w</i> <w>		<u><i>r</i></u> <r>	<i>y</i> <i>	

Table 2: Mapping from Pitjantjatjara orthography in *italics* onto Basque ‘letters’ in <angle brackets>.

borrowings in Basque, when it has its Spanish value [i]). The palatal sounds are well catered for in Basque: the palatal stop, written <tt>, is used in diminutives and childish forms, while Spanish <ñ> and <ll> are needed for Spanish borrowings, and so give us palatal *ny* and *ly*.

3.3. Implementation

All the transliterations identified in Table 2 can be applied to Pitjantjatjara text by a simple string-substitution program. Examples (1) and (2) show Pitjantjatjara texts in standard orthography (a) and as they are input to the Basque TTS (b).

- (1) a. *Tjitji nyanga kati atjupitlakutu.*
b. ttitti ñanga kati attupitlakutu.
‘Take this child to the hospital.’
- (2) a. *Yaaltjitu arana tjikinma?*
b. iaalttitu arana tkinma?
‘How many times should I drink it?’

4. Evaluation

Evans et al. (2002) describe a modest evaluation of a prototype Greek synthesizer using both English and Spanish as the BL with just three native Greek speakers who were also fluent English speakers. A first evaluation involved a variation of the Modified Rhyme Test (House et al. 1965, Miner and Danhauer, 1976, Logan et al. 1989, Goldstein 1995) in which subjects must match from a list of five options the word which they think they have heard. The subjects underwent 15 tests, with 5 words in each test. Evans et al. report that “with the Spanish synthesiser ... there were a total of 7 errors in 45 trials”, a 97% success rate.⁷ With the English synthesizer there were 10 errors (96%). A second evaluation involved the use of nonsense words, with much lower recognition rates (50% for both systems). In a third evaluation, the subjects were tested with a number of complete sentences, including “common simple sentences and a small number of ‘tongue-twisters’”. Correct identification in this case was 100%, though it is unclear whether subjects had to identify what they heard,

⁷ It is not explicitly stated, but Evans (personal communication) confirms that the 45 trials were 15 tests × 3 subjects. Since each test had 5 items, the total number of items is 225.

or choose from a number of alternatives.

Evans et al. readily admit the shortcomings in this evaluation, mentioning the small sample size, and the fact that all the subjects were fluent English speakers, long-term UK residents, therefore familiar with English phonemes and prosody. In addition, they note that the tests did not reflect the actual intended use of the software, in their case as a screen reader.

Our evaluation attempts to bypass these shortcomings: we aim to recruit sufficient subjects to make the results statistically significant; English (or Basque) language ability will not play a role in the experiment; and experience/exposure to English is more or less constant for all Pitjantjatjara speakers; and, most important of all, we attempt to simulate better the situation in which the software might be used.

4.1. A realistic scenario

Our methodology is based on the method used by Somers and Sugita (2003) in their evaluation of SLT software. SLT research, almost without exception, focuses on one particular type of application, namely task-oriented co-operative dialogues, for example, scheduling a meeting, or arranging travel. Accordingly, Somers and Sugita chose to evaluate the SLT software by translating phrases taken from a tourist's phrasebook. Importantly, they were interested in "the subject's ability to infer correctly the *intended* meaning of the utterance" (emphasis added) rather than the grammar or style of the translation. In a similar manner, we are interested above all in whether the faked output is intelligible, with little interest in naturalness and phonetic accuracy, unless it impinges on intelligibility, in our healthcare scenario.

In our experiment, participants will be told (in their own language by a native speaker experimenter) to imagine that they have gone to the doctor's office with some specific medical problem, say, respiratory difficulties, and that whatever the doctor says is going to be translated and 'spoken' by the computer. They will then be asked to listen to the synthesised speech, and to tell the experimenter what they understood. Because the syntax etc. of the 'translation' is not an issue, it will be acceptable if they simply repeat verbatim what they have heard. The experimenter will make a judgment about whether they have understood, and will ask clarificational questions if necessary. Sessions will be recorded to enable judgments to be corroborated.

As in the Somers and Sugita (2003) experiment, five different sub-scenarios will be presented, each with a contextualisation (e.g. the

doctor asks you about your symptoms, the nurse will explain the treatment, the receptionist wants to make a follow-up appointment), with five phrases in each, giving a total of 25 phrases (see Appendix). Sessions should last about 15 minutes. Phrases have been specifically tailored so as to contain essential information which must be repeated by the participant if the experimenter is to judge the phrase to have been understood, e.g. *Take two puffs every four hours*. At the end of each session, the experimenter will elicit any general reactions and opinion from the participants in an informal interview.

If enough subjects can be found, a variant of the experiment will use human recordings (rather than synthetic speech) for a selection of the phrases, to provide a base-line control. Alternatively, it would be interesting to contrast the use of Basque as the BL, with preprocessing to adapt the orthography, against the cheaper but 'dumber' implementation using the free English-based TTS software typically available on most computers (e.g. using the voices of Microsoft Sam and Microsoft Mary).

In our experiments with Somali (Somers et al., 2006), we have tested several different implementations, combining the approach described here with the 'gibbering' approach of Evans et al. (2002), as mentioned above. We have told the subjects that the computer would have several different 'voices', and fitted the scenario around this, with each section associated with a different speaker: a receptionist, the doctor, a pharmacist, and so on. Obviously we need to sample each 'voice' with each section, so the experimental set-up is a Latin square of voices \times sections, which determines the minimum number of participants needed. As each different voice is introduced, a sample is introduced and played: for example, the experimenter will say "The doctor says: *I want to get some background details*", and then plays the sample.

5. Pilot studies

At the time of writing we are still hoping to set up an experiment with Pitjantjatjara speakers via a number of agencies. During pilot experiments with Somali speakers (Somers et al., 2006), a number of interesting elements have led us to adjust the experimental design.

The first of these concerns the translations. It will not come as a surprise to find that Pitjantjatjara does not have its own words for some vocabulary items such as *pollution*, *medical history*. There are also some constructions which are hard to render in Pitjantjatjara, for example comparatives (e.g. *Is it worse at night?*). It might

even turn out that asthma is not an appropriate testbed for that particular locale. The significance of these is that, as we quickly discovered with the Somalis, loanwords were difficult to understand: subjects who otherwise were obtaining good scores for comprehension were baffled by Somali words such as *dhamb* ‘damp’, *boolushan* ‘pollution’ and *hayla* ‘inhaler’.

A second idea to emerge from the pilot experiments was to adjust our scoring system. We have found that subjects would like to hear each sample two or three times, and they typically ‘get’ more of the message on each hearing. So we are devising a scoring method which takes this into account, and quantifies not just whether the subjects could understand the synthetic speech, but how many repetitions they need (up to, say, three). In a practical situation it is not unreasonable to ask to hear something again.

Finally, we have noticed that subjects quickly get used to the voice, and seem to have a better level of recognition towards the end of the test. This is something we should try to measure, although the experimental design with multiple ‘voices’ may mask this effect considerably.

We look forward to reporting results of the experiment in due course.

Acknowledgements

Our thanks go to Andrew Longmire at the Department of Environment and Heritage’s Cultural Centre, Uluru–Kata Tjuta National Park, Yulara NT, and to Bill Edwards, of the Unaipon School, University of South Australia, Adelaide, for their interest in the experiment, and, we hope eventually, for their assistance in conducting it.

Appendix – Rubric and test items to be translated into Pitjantjatjara

[The text in italic is spoken by a human – the experimenter – in the participant’s own language – text in bold represents the sample in each case]

You are going to hear some sentences spoken by the computer. I want you to tell me in your own words what you understand the computer to be saying. You don’t have to answer the questions. I can play them up to three times each. Don’t worry if you don’t really understand it – it’s not you who is being tested!

The computer has different voices. Each time I will play you an example and tell you what it is saying.

Please imagine that you want to see the doctor because you have difficulty with breathing.

1. The receptionist says: *First we have to make an appointment.*

- a. What is your name?
- b. Have you been to this clinic before?
- c. Can you come next Thursday?
- d. How about ten o’clock?
- e. Do you want an appointment card?

2. The doctor’s assistant says: *I am going to ask you some questions about your circumstances.*

- a. Do you smoke, or have you ever smoked?
- b. Do you have any pets?
- c. Do you take regular exercise?
- d. Does anyone in your family have asthma?
- e. Is your home dry?

3. The doctor says: *I want to get some background details.*

- a. When did you first have difficulty with breathing?
- b. Is your breathing worse at any particular time of day?
- c. Does it prevent you from sleeping?
- d. Is it worse after strenuous activity?
- e. Does it depend on the weather?

4. The nurse says: *I will make some suggestions.*

- a. Your asthma may be triggered by the weather.
- b. Check your bathroom for mold.
- c. Use the brown inhaler if you have a mild attack.
- d. Use it once a day for a week after an attack.
- e. Take two puffs every four hours.

5. The pharmacist says: *Your prescription is ready.*

- a. Are you taking any other medication?
- b. What is your name and address?
- c. Take this three times a day.
- d. Are you entitled to a free prescription?
- e. That will be six dollars fifty please.

References

- Eckert, P. and J. Hudson (1991) *Wangka Wiru: A Handbook for the Pitjantjatjara Language Learner*. Underdale, SA: University of South Australia.
- Evans, G., K. Polyzoaki and P. Blenkhorn (2002) An approach to producing new languages for talking applications for use by blind people. In K. Miesenberger, J. Klaus and W. Zagler (eds) *8th ICCHP, Computers Helping People with Special Needs*, (LNCS 2398), Berlin: Springer Verlag, pp. 575–582.
- Goldstein, M. (1995) Classification of methods used for assessment of text-to-speech systems

- according to the demands placed on the listener. *Speech Communication* **16**:225–244.
- House, A.S., C.E. Williams, M.H.L. Hecker and K.D. Kryter (1965) Articulation-testing methods: Consonant differentiation with a closed-response set. *Journal of the Acoustic Society of America* **37**:158–166.
- Logan, J., B. Greene and D. Pisoni (1989) Segmental intelligibility of synthetic speech produced by rule. *Journal of the Acoustical Society of America* **86**:566–581.
- Miner, R. and J.L. Danhauer (1976) Modified rhyme test and synthetic sentence identification test scores of normal and hearing-impaired subjects listening in multitalker noise. *Journal of the American Audiological Society* **2**:61–67.
- Narayanan, S., S. Ananthakrishnan, R. Belvin, E. Ettelaie, S. Gandhe, S. Ganjavi, P.G. Georgiou, C.M. Hein, S. Kadambe, K. Knight, D. Marcu, H.E. Neely, N. Srinivasamurthy, D. Traum, and D. Wang (2004) The Transonics spoken dialogue translator: An aid for English-Persian doctor-patient interviews. In: T. Bickmore (ed.), *Dialogue Systems for Health Communication: Papers from the AAAI Fall Symposium*, Menlo Park, CA: The AAAI Press, pp. 97–103.
- Rayner, M., P. Bouillon, V. Van Dalsem, H. Isahara, K. Kanzaki and B.A. Hockey (2003) A limited-domain English to Japanese medical speech translator built using REGULUS 2. *Companion Volume, 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, 137–140.
- Somers, H., G. Evans and Z. Mohamed (2006) Developing speech synthesis for under-resourced languages by ‘faking it’: An experiment with Somali. Paper submitted to *LREC 2006* (Genoa).
- Somers, H. and H. Lovel (2003) Computer-based support for patients with limited English. *Association for Computational Linguistics EACL 2003, 10th Conference of The European Chapter, Proceedings of the 7th International EAMT Workshop on MT and other language technology tools*, Budapest, pp. 41–49.
- Somers, H., H. Lovel, M. Johnson and Z. Mohamed (2004) Language technology for patients with limited English. *EACH International Conference on Communication in Healthcare*, Bruges, P04.09.
- Somers, H. and Y. Sugita (2003) Evaluating commercial spoken language translation software. *MT Summit IX: Proceedings of the Ninth Machine Translation Summit*, New Orleans, pp. 370–377.