CONSIDERATIONS IN THE REALISATION OF A TEXT-TO-SPEECH SYNTHESIS SYSTEM FOR PITJANTJATJARA LANGUAGE

K.S. Ananthakrishnan*#, Callan Hanley#, John Asenstorfer*, Bill Cowley*, Bill Edwards**

* Institute for Telecommunication Research, University of South Australia
** Faculty of Aboriginal and Islander studies, University of South Australia

ABSTRACT- Pitjantjatjara is one of the most widely used of the Western Desert Australian Aboriginal languages. We intend to exploit the information technology to facilitate flexible learning of this language by students, children and adults. Another objective is to preserve the heritage of Aboriginal languages for future generations. To achieve these goals, a project has been undertaken to study the feasibility of realising a Text-to-Speech Synthesis system for Pitjantjatjara language, where the ultimate aim is to develop a user friendly system, to serve the Australian Community at large. This paper reports the preliminary results obtained from a speech generation module currently under development at the University of South Australia and projects the future directions of research in this application area. To our knowledge, our project is the first one making an attempt to develop speech synthesis system for Pitjantjatjara language in Australia.

INTRODUCTION

Pitjantjatjara and Yankunytjatjara are 'sister dialects' of a larger language group called the Western Desert Aboriginal Language. There are approximately 1600 Pitjantjatjara and Yankunytjatjara people living on the Pitjantjatjara freehold lands in South Australia or just over the borders in Western Australia and Northern Territory (Eckert and Hudson 1992). The use of Pitjantjatjara as the language of instruction in schools was begun at Ernabella in 1940 and this has led to a reasonably high proportion of Pitjantjatjara people today being literate in their own language. Ptjantjatjara language is still in everyday use by a significant number of speakers.

The past few years have seen unprecedented growth in information technology due to rapid developments taking place in the field of computer hardware and software. In particular, as applied to speech , text-to-speech synthesis systems employing different techniques have been developed for Japanese language(Hakoda 1990, Ishikawa and Nakajima 1994), Korean language (Sang-Hun and Jung-chul Lee 1994) and other European Languages. In this project, an attempt is being made to realise a text-to-speech system for Pitjantjatjara language (spoken only in Australia by small groups) primarily to facilitate flexible learning of this language by students, children and adults. In a text-to-speech system, the basic acoustic units to be used can be whole sentences, words, syllables or phonemes. In this project, we first made an initial attempt to synthesise the selected words from syllabic units derived from Pitjantjatjara words. It is envisaged that the intended system will eventually be phoneme based and consists of a language processing module and a prosodic processing module in addition to a speech generation module. The prosodic processing module and allophonic details are intended to be developed in the next phase of the project.

In this paper, we briefly provide the details of the investigative analysis on : the linguistic structure of the words in Pitjantjatjara language from the perspective of signal processing for the selection of appropriate synthesis units; give an overview of the system currently under development taking into consideration the unique linguistic aspects of the language; present the initial methodology adopted for the synthesis of selected words in Pitjantjatjara language and report the preliminary experimental results obtained with illustrated examples.

MAIN CONSIDERATIONS

The main objective of the project was to initially study the feasibility of synthesising the words in Pitjantjatjara language and this necessitated the investigation of the linguistic aspects of the language necessary for the basic understanding of the phonemic and syllabic structure of words. Human speech organs can make a large range of sounds(Blake 1984), but each language recognises comparatively few of these sounds as being different from each other(Blake 1984)). English language distinguishes 44 and Pitjantjatjara language 25(including 2 diphthongs)(Kirke 1987).

BASIC LINGUISTIC ANALYSIS

In the Pitjantjatjara language , sounds are divided into two groups: vowels and consonants. There are three short vowels 'a', 'i', 'u' , 3 long vowels 'a:','i:','u:'(also written as 'aa','ii','uu') and 2 diphthongs 'ai' and 'au'(Kirke 1987). Phonemically, vowels may be written as /a/ , /aa/ and so on. Most Pitjantjatjara words end in a vowel(for example, 'mutuka' meaning 'motor car' and 'taraka' meaning 'truck'). There are 17 consonants in Pitjantjatjara language as listed below:

k, l , l̲ , ly , m , n , n̲ , ng , ny , p , r , r̲ , t , t̲ , tj , w and y .

It was also observed that the Pitjantjatjara language has a limited vocabulary consisting approximately of 2400 words (all words used in the language not just the nouns) , approximately 300 hyphenated words such as 'muu-muu', 'nyii-nyii', 'ut̲i-ut̲i', 'tjilyi-tjilyi', and approximately 30 case endings(eg -kutu and -lu) (Goddard 1992).

Consonants are rare at the end of words in Western Desert language, although 'n', 'l' and 'r' occasionally occur , typical examples of such words being 'ngaan', 'raitjin' and 'pintjin'(Kirke 1987).

A vowel(V) is usually the syllable nucleus while a consonant(C) usually represents the syllable margin. A long vowel counts as two syllables in Pitjantjatjara, for example, word like 'taanpa' meaning 'rise' or 'outcrop' is really a three syllable word. Aboriginal words usually have atleast two syllables(Blake 1984). For Aboriginal words, the accent generally falls on the first syllable.

An investigation into the structure of 2 to 5 letter words collected from a Dictionary(Goddard 1992) in Pitjantjatjara and other *'Pitjantjatjara Readers'* books used by the Education Department of South Australia revealed that word patterns CV, VCV, CVCV, VCCV, CVCCV, VCVCV are quite commonly found in the speaking language. Longer units such as VCV syllables, CVC syllables, triphones and quadriphones found in the language could therefore be considered as basic synthesis units for a text-to-speech system.

CREATION OF DATABASE

As no speech database is currently available for the Pitjantjatjara language, digitized samples of the selected words were obtained from previously recorded analogue tapes supplied for Pitjantjatjara language course by the Faculty of Aboriginal and Islander studies of the University of South Australia. The digitisation was carried out using ' Goldwave' software package and speech segments sampled at 8000 samples per second(each sample being 16 bits in resolution). A digital data base was created for nearly 50 -60 selected isolated words from analogue tapes where continuous sentences spoken were by the same speaker.

SYNTHESIS METHODOLOGY

The linguistic analysis influenced the choice of synthesis unit to be a 'syllable' for our preliminary investigation. Our method uses a set of pre-recorded sub-syllabic synthesis units. The synthesis technique used is a time-domain approach based on concatenation of the digital data that represents synthesis units. As a first trial , in our data base, we stored only limited synthesis units consisting of open-syllables (consonant-vowel), triphones and vowel-consonant-vowel units. An open syllable is designated as CV , not ending in a consonant.

Flow chart in Figure 1 shows the various functions performed by our experimental text-to-speech system under development . The text entered by the user as phonemic transcription such as /mutuka/ is pre-processed . Each character of the input text is analysed to determine whether it is a consonant or vowel. A 'c' or a 'v' is concatenated with each character depending on whether a consonant or a vowel is encountered.

The output of the preprocessing results in a new array 'mcuvtcuvkcav' (letter 'c' concatenated with the consonant 'm' and the letter 'v' concatenated with the vowel 'u' and so on).

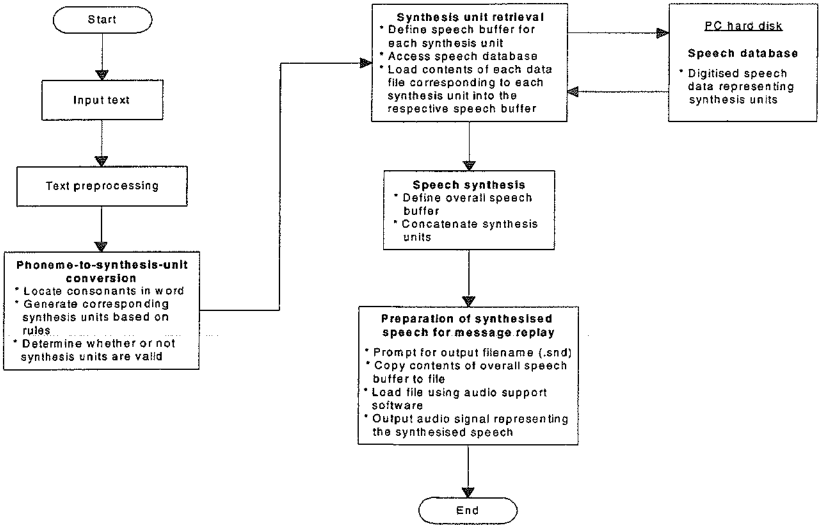

Start
Input text
Text preprocessing
Phoneme-to-synthesis-unit conversion
• Locate consonants in word
• Generate corresponding synthesis units based on rules
• Determine whether or not synthesis units are valid
Synthesis unit retrieval
• Define speech buffer for each synthesis unit
• Access speech database
• Load contents of each data file corresponding to each synthesis unit into the respective speech buffer
PC hard disk
Speech database
• Digitised speech data representing synthesis units
Speech synthesis
• Define overall speech buffer
• Concatenate synthesis units
Preparation of synthesised speech for message replay
• Prompt for output filename (.snd)
• Copy contents of overall speech buffer to file
• Load file using audio support software
• Output audio signal representing the synthesised speech
End

Figure 1. Flow chart of system functions

PHONEME-TO-SYNTHESIS UNIT CONVERSION AND GENERATION OF SYNTHESIS UNITS

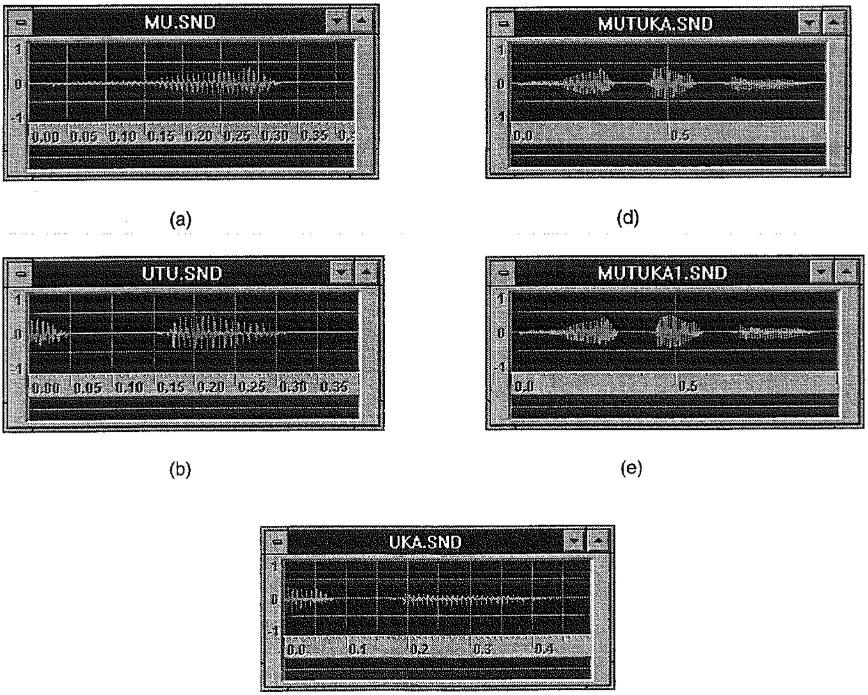
For each consonant in the phonemic transcription, there exists one synthesis unit and hence, the utterance contains as many synthesis units as it does consonants. In the phonemic description of 'mutuka', there are three consonants 'm','t', and 'k' and this results in the generation of three synthesis units 'mu', 'utu' and 'uka'. The procedure for the generation of these synthesis units is the same as that used for the generation of synthesis units for the Arabic language(Yousif 1990).

RESULTS AND DISCUSSION

Program code was written in 'C' on a PC for : text preprocessing, phoneme-to-synthesis unit conversion, synthesis unit retrieval and speech synthesis. The message replay of the synthesised word was carried out by the 'Goldwave' software.

EXPERIMENT 1

In this experiment , phonemic transcription of the word /mutuka/ was used as input text. The word 'mutuka' in Pitjantjatjara meaning 'motor car' in English language is a three syllable word, three syllables being 'mu','tu','ka'. The program generated the synthesis units /MU/(CV), /UTU/(VCV), /UKA/(VCV) as shown in Figure 2a, 2b and 2c. The same word was resynthesised from the synthesis units, by sliding the succeeding synthesis unit on the previous synthesis unit in the backward direction in the time domain to determine the point of onset of the vowel for alignment of the same vowel in the succeeding synthesis unit and concatenating the succeeding synthesis unit with the previous synthesis unit from the point of alignment. In Figure 2d, the acoustic waveform of the original word 'mutuka'(mutuka.snd) is shown and in Figure 2e , the waveform is displayed for the synthesised word(mutuka1.snd). Both the original and synthesised words were intelligible and the quality was more than satisfactory.



(a)

(d)

(b)

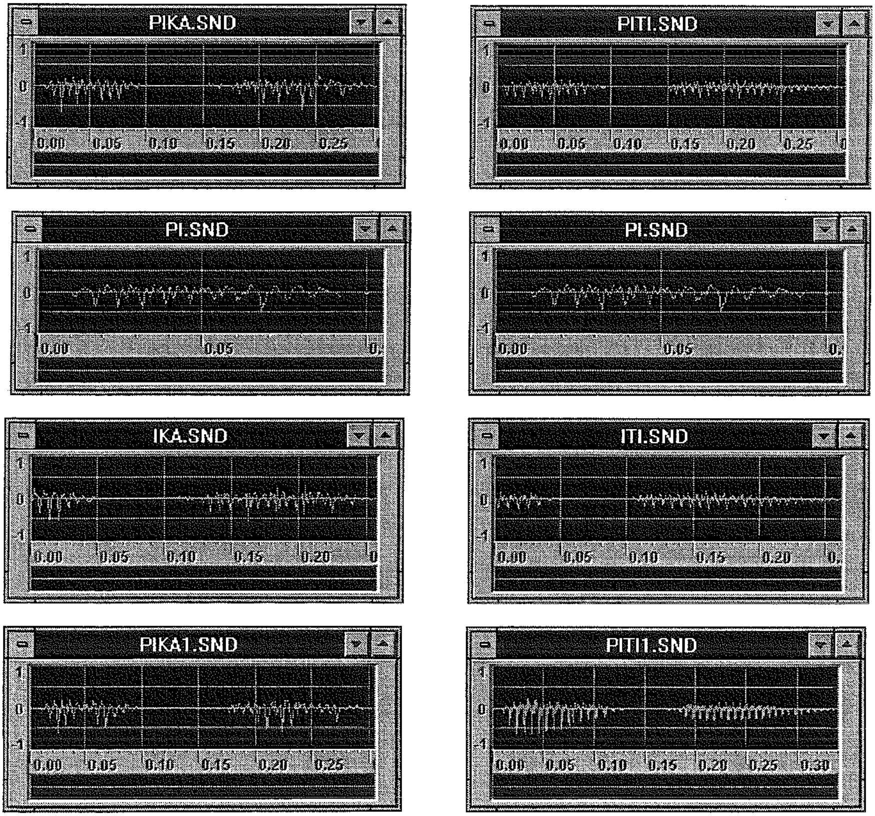
(e)

(c)

Figure 2. Synthesis units of the word /MUTUKA/ (a) VC unit /MU/ (b) VCV unit /UTU/(c) VCV unit /UKA/ (d) Original word /MUTUKA/ (e) Synthesised word /MUTUKA/

EXPERIMENT 2

In this experiment, two syllable words ‘pika’ and ‘piti’ were taken and applied to the synthesis unit generation program. From the word ‘piti’, the synthesis units /PI/(CV), and /ITI/(VCV) were generated. In a similar way, from the word ‘pika’, the synthesis units /PI/(CV), and /IKA/(VCV) were generated. Synthesis unit /pi/ generated from /piti / was discarded. By running the synthesis program, the word /piti/ was synthesised from the synthesis unit /pi/ derived from /pika/ and the synthesis unit /iti/ derived from the word ‘piti’(PITI.SND shown in Figure 3b). On comparing with the original word ‘piti’ with the synthesised word, it was observed that the /p/ sounded like /b/ in English. In Pitjantjatjara, ‘p’ has no puff of air and may sound like ‘b’ unlike the English ‘p’ in words like pin, pull, pat(for example papa, apu, ipi, piti). Figure 3(a) shows the synthesised word /pika/ from the synthesis unit /pi/ from ‘piti’ and ‘the synthesis unit /ika/ from the word /pika/. Figure 3(b) shows the waveform of the synthesised word /piti/.



(a) (b)

Figure 3 Synthesis of the words /PIKA/ and /PITI/ using cross-synthesis units

CONCLUSION AND FURTHER WORK

This paper reported the preliminary results obtained in our initial attempt to realise a text-to-speech synthesis system for the Aboriginal language 'Pitjantjatjara'. The quality of the synthesised words using longer acoustic units such as syllables, diphones, triphones match closely with the natural speech. However, even for a language with limited vocabulary, storage space need to be optimised. The relative simplicity of the language should lead to good results. It may be necessary, however to code phonemes as lattice reflection filter coefficients and excitation sequences to allow interpolation between acoustic units. It is not clear yet to what extent allophonic information will have to be incorporated. Alternatively, an LPC-based diphone synthesis approach would be expected to yield good results and may be investigated at a later stage. Further work is required to produce speech synthesis using phonemes as basic acoustic units.

ACKNOWLEDGMENTS

The authors wish to thank Mr. Jim Harvey, Dean -UniSA, Whyalla Campus for the support given to this project and for the permission to publish the preliminary results. We express our appreciation to Dr. Brian Trainor and Dr.June Stephenson , members of the Small Grants Committee for providing the financial support. Special thanks to Mr. Kevin McDermott for approving the project under Advanced Study 1 subject. We are grateful to Mr. Malcolm Raymond for his initial suggestions and Ms. Loraine Lerc, Mr. Henry Croft and Mr. David Wilson for their support to continue this useful activity further.

END NOTE # The authors are located at the Whyalla Campus of the University of South Australia.

REFERENCES

Blake, B.J (1984) *Australian Aboriginal Languages: A General Introduction*, Angus and Robertson Publishers, Sydney.

Eckert, P. & Hudson, J. 1992, *Wangka Wiru: A Handbook for the Pitjantjatjara Language Learner*, Aboriginal Studies and Teacher Education Centre (ASTEC), University of South Australia, Underdale.

Goddard, C. 1992, *Pitjantjatjara/Yankunytjatjara to English Dictionary*, 2nd edn, Institute for Aboriginal Development, Alice Springs.

Hakoda,H., et al .(1990) *A New Japanese Text-to-Speech Synthesizer based on COC Synthesis Method*, Proc.ICSLP '90, pp. 809-812.

Ischikawa,Y. & Nakajima,K.,(1994) *On Synthesis Units for Japanese Text-to-Speech Synthesis of Japanese* , Proc.ICSLP '94, pp. 1751-1754.

Kirke, B. (Cartoons by Cane, J.) 1987, *Wangka Kulintjaku (Talk so as to be understood): An introductory self-instruction course in Pitjantjatjara (a dialect of the Western Desert Aboriginal language)*, 2nd edn, Aboriginal Research Institute, Faculty of Aboriginal and Islander Studies, University of South Australia, Adelaide.

Sang-Hun Kim. & Jung-Chul Lee (1994) *Korean text-to-speech system using time domain-pitch syncronous overlap and add method*, Proc.of the fifth Australian International Conference, SST-94, pp 587-592.

Yousif A. El-Imam(1990), *Text-to-Speech Conversion on a Personal Computer* , IEEE Micro, August 1990.