THE IMPORTANCE OF F0 OR VOICE PITCH FOR PERCEPTION OF TONAL LANGUAGE: SIMULATIONS WITH COCHLEAR IMPLANT SPEECH PROCESSING STRATEGIES

Robert Alexander Fearn B.E.Electrical (HONS)

School of Physics, University of New South Wales, Sydney, Australia

ABSTRACT

Cochlear implants were developed and were initially distributed in Western countries, in particular Australia, North America and parts of Europe. The speech processing strategies that have been developed to drive the implants provide detailed information about the spectral envelope and transients. This information is necessary to identify the phonemes of English and other Western languages.

Cochlear implants have more recently been distributed in some Eastern countries such as China. Dialects of Chinese such as Mandarin and Cantonese are called tonal languages. In addition to spectral envelope and transient information, tonal languages require finer resolution of pitch to distinguish different words. Speech processing strategies only provide the cochlear implant user with relatively low resolution of pitch.

This study investigates the importance of voice pitch (F0) in detection of phonetically identical Cantonese words, which vary in pitch. Simulations of speech processing strategies were performed with a normal hearing subject and the results suggest that F0 is very important for correct identification of tonal words and without it, identification reduces dramatically.

1. INTRODUCTION

Unlike Western languages, many Eastern languages rely on separate tones to identify words that are phonetically identical. A tone can be described as a distinct level of pitch or change in pitch throughout the duration of the word.

One quarter of the world's population speaks Chinese. There are many dialects of Chinese, two of these include Mandarin and Cantonese. Mandarin is spoken primarily in mainland China especially the north. Cantonese is spoken in the south of China and in other Chinese settled countries like Malaysia and Hong Kong (Comrie, 1987). Mandarin officially consists of four different tones and Cantonese of six different tones.

Cochlear implants ("**CI**") have been developed in Western countries like Australia, America and parts of Europe whose languages carry little information in the change of tone except for inflection to indicate a question or statement. Most of the information necessary to identify the phonemes in Western language is found in the spectral envelope, particularly formant shapes, and transients.

Speech processing strategies ("**strategies**") that drive the implants have been designed to provide information about these broad spectral envelopes and transients. These strategies have been reasonably successful with users averaging 80% correct in open set word recognition tasks and some users nearly scoring 100% (Clark,1992). Changes in pitch are transmitted rather poorly by current strategies, with CI users only able to achieve discrimination levels of less than half an octave (30-40% in frequency) (Fearn, unpublished thesis). In comparison normal hearing subjects can discriminate up to 0.5% in frequency (Roederer, 1973).

The poor levels of pitch discrimination by CI users has implications not only for tonal languages but also for the perception of music for all CI users.

Studies have been carried out in improving the filterbank frequency boundaries for non-English, non-tonal Western languages like such as Spanish (Aronson & Arauz, 1995). Filter boundaries were adjusted so that vowels with similar formant positions could be identified more easily. This technique significantly improved vowels and bisyllable discrimination by 11% (73% to 84%) and 8% (64% to 72%) respectively.

Studies with CI users speaking tonal languages like Mandarin and Cantonese have used strategies where the fundamental frequency ("**F0**") was extracted and presented to the users as the rate of stimulation.

Kwok et al (1991), investigated 8 subjects using a single channel device and whose language was Cantonese. Their study found that the mean recognition of the 6 tones of the phoneme /fu/ was 38% (chance level of 17%).

Tang et al (1990), found similar results studying 4 subjects, using a single channel device and whose language was Cantonese. They found that the mean recognition of the different tones of phonemes was 30% (chance level of 17%).

Xu et al (1988), based their study on a subject whose first language was Cantonese and who also spoke English. With a multichannel device and where formant information was presented at the F0 rate, this subject could remarkably score 100% (33/33) for a tonal word test in Cantonese and 100% (20/20) in detecting the direction of inflection in an English Question/Statement test.

Huang et al (1996) speculates that the cues of F0 may be extracted from the rate pitch signal of the stimulation. This study was based on 4 Mandarin speaking subjects using a multichannel device. It was found that the mean recognition of the 4 tonal phonemes of Mandarin was 65% (chance level of 25%).

The single channel device studied by Kwok et (1991) and Tang et al (1990) did not provide as accurate tonal information as the multichannel device studied by Xu et al (1988) and Huang et al (1996), where spectral information as well as the F0 rate was presented.

A multichannel CI consists of an array of electrodes, up to 22, inserted into the scala tympani in the cochlea. These 22 electrodes are stimulated by passing current between them to activate the surviving nerves in the organ of corti or more possibly the spiral ganglion cells in the modioulus. A typical healthy cochlea may have 30 000 nerve cells stimulated by the motion of the basilar membrane. A typical CI user may have only 10 000 surviving nerve cells (Clark,1992) stimulated electrically, by at most, 22 electrodes. As a result, a large reduction in spatially disccriminable stimuli would be expected. There exists however, the temporal component of coding the input signal which may accurately convey 'pitch' up to 800 Hz (Pijl and Schwarz, 1995).

There are recent strategies that do not attempt to extract and present FO information to the CI by rate of stimulation. These strategies choose an arbitrary rate and rely on the tonotopic arrangement of the cochlea to convey the pitch of the incoming signal (Whitford et al, 1995). (See Clark (1992) and Blamey et al (1985) for a complete description of recent strategies and results achieved). These recent strategies that present detailed spectral information have improved speech scores from previous formant picking strategies. It remains to be seen whether these new strategies that do not present rate pitch can transfer enough information in the spectrum to distinguish tonal language.

This study aims to investigate two different strategies by simulation. One strategy utilises the spatial tonotopic arrangement of the cochlea and presents stimulation at an arbitrary rate. The other strategy uses the spatial tonotopic arrangement of the cochlea as well as temporal coding of F0. These strategies are simulated in the acoustic domain using information known about electrical stimulation of the ear and functions of the auditory system.

2. MATERIALS AND METHODS

2.1 Lexical Tone Perception Using Simulated Cochlear Implant Speech Strategies

The normal hearing subject who participated in this experiment was a native Cantonese speaker from Hong Kong, who now resides in Australia and speaks English as a second language. The subject has normal hearing levels and **does not** have a cochlear implant.

The Cantonese phoneme /si/ was chosen for investigation which has six different meanings depending on the tone used. The accepted tones and tone numbers are:-

Table 1: Cantonese Tones and Meanings for the phoneme /si/.

Tone Number	Tone Name	Meaning
1	High Level/Falling	poem
2	High Rising	history
3	Middle Rising	test
4	Low Rising	time
5	Low Falling	city
6	Low Level	event

A tape recorder was used to record the subject. The subject recorded the six tones alone and six tones in contextualised phrases. This data was then transferred to a computer, stored as 'wave' files and filtered to reduce noise and hiss.

These six words and six phrases were then processed using two schemes. The first scheme used a filterbank ("**FB**") strategy that filtered the sound data into sixteen channels. The six channels with the largest energy were chosen and the output calculated by combining the appropriately weighted, filtered impulse responses from the chosen filterbank. This was calculated with a window period of 4ms (250Hz). The second scheme was identical except the lowest filter in the filterbank was discarded and replaced with the F0 frequency of the word or phrase ("**FB+F0**"). This data was then converted back into audio files.

The audio files were then presented to the subject for identification, in 6 groups. These groups were: 'Words using raw audio', 'Words using FB', 'Words using FB+FO' 'Phrases using raw audio' 'Phrases using FB' and 'Phrases using FB+F0'.

In each of the groups the subject listened to the six words or six phrases, a total of 3 times each, presented in a random order. For each audio presentation the subject had to identify the correct word or phrase from a closed group of the six words or phrases.

Cantonese characters used in the tests and their meaning:-

1. 詩 /si/ poem 2. 史 /si/ history 3. 試 /si/ test

4. 時 /si/ time 5. 市 /si/ city 6. 事 /si/ event

Phrases used in the testing and their meaning:-

1. 一首詩 A poem
2. 歷史 Past history
3. 考試 An exam
4. 幾時 What time?
5. 大城市 A large city
6. 大件事 A big event

3. RESULTS

A spectrogram of the six tonal versions of /si/ are shown in Figure 1. It is clear how the broad spectral features of these six phonemes are almost identical.

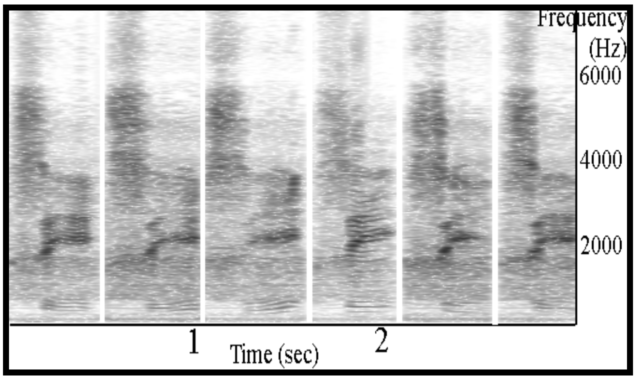

Figure 1: Spectrogram of the six different tonal phonemes /si/.

A more detailed spectrogram of the low frequencies was used with the six tonal phonemes and the plots of changes in voiced frequency are shown in Figure 2.

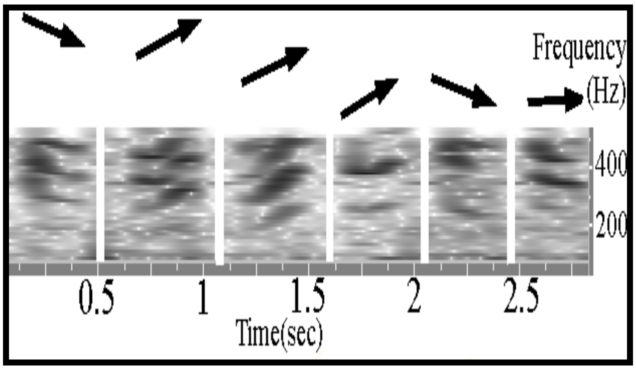

Figure 2: Detailed spectrogram of 0 to 500 Hz of the six words used in the tests. Bar lines indicate the end of each word. Arrows indicate the direction of the change in pitch.

Results of word and phrase identification are shown in Figure 3.

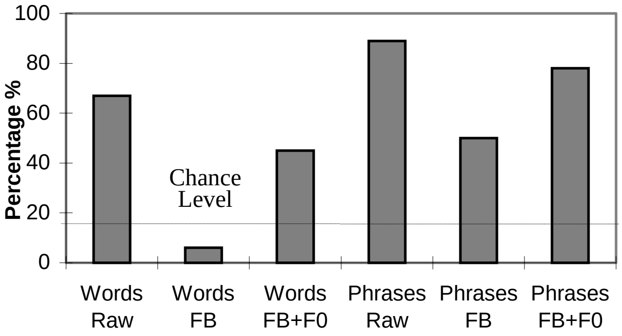

Figure 3: Word and Phrase Test Results

The results clearly show that identification of these isolated unprocessed tonal phonemes is reasonably difficult, with the mean recognition of raw audio being 67% (12/18). The identification of the tonal phonemes drops to 6% (1/18) when processed by the FB strategy but improved to 45% (8/18) when processed by the FB+F0 strategy.

When the subject was presented with the tonal phonemes in a short contextualised phrase, the mean recognition of raw audio was 89% (16/18). When the FB strategy was applied the identification rate drops to 50% (9/18). When the FB+F0 strategy was applied the identification rate improved to 78% correct (14/18).

4. DISCUSSION

The results suggest that even though tonal phonemes rely on the voice pitch for identification, that alone, is not enough for 100% identification. The subject scored 67% in a closed set task using unprocessed audio files of the subject's previously recorded voice. This is well above chance of 17% and shows that the subject was using the tonal features of the phonemes. Some ambiguity exists with these tonal phonemes as the recognition was not 100%.

When the FB strategy was used with the isolated words that did not deliver fine resolution of F0, the subject scored 1 correct out of 18 (6%). This is well below chance and the subject admitted they were clearly guessing as each word sounded very similar. This shows how much information was lost by only transmitting broad channel information. The overall phonetic sound of the word was preserved, due to this strategy transmitting detailed information about transients and broad spectral information. It was quite clear to the subject which phoneme was being said, but difficult to distinguish which tonal phoneme it was.

When this strategy was changed to include finer resolution F0 details, as would be done by F0 rate presentation to CI users, the subject's score improved dramatically to 45%.

The identification of the phrases when unprocessed was 89% which demonstrates how the contextual cues improve the recognition of the tonal phoneme. When the phrase was processed by the FB strategy the identification dropped to 50%, which was much improved on the 6% scored when using isolated words alone. The mean identification improved to 78% when the phrase was processed by FB+F0. All the phrase scores were higher compared to the word scores indicating the importance of context.

5.CONCLUSIONS

Audio simulations of the electrically stimulated ear can never fully represent what the CI user might hear. The separation of rate and place of stimulation information that a CI user may receive is virtually impossible to reproduce in the acoustical domain. These simulations may however, provide some idea of what parameters may be important in a strategy and are worth pursuing with a full study using CI users.

The results demonstrate the importance of F0, namely, that without F0 extraction the identification of phonetically identical words with different tones is almost zero. CI users speaking tonal languages are likely to score better on contextual sentences than isolated words. Contextual sentences allow a subject to use the surrounding words to assist in determining the meaning of the phoneme. This ambiguity in word discrimination must reduce speech rate understanding and add to the already reduced information that CI users receive. By attempting to improve the recognition of tonal phonemes the overall speech rate may increase, which is likely to result in a greater understanding of tonal languages.

A more detailed study is presently being conducted to compare strategies with CI users that speak a tonal language in order to fully identify any improvements that could be made.

6. ACKNOWLEDGEMENTS

The author highly appreciates Louis Wong for his time for recording and testing these words and phrases in Cantonese. The author is also in gratitude to Theresa Fearn for her time and assistance with the editing.

7. REFERENCES

1. Aronson, L., Arauz, S.L. “Fitting the Nucleus-22 cochlear implant for Spanish speakers”, *Clark and Cowan, International cochlear implant, speech and hearing symposium, Melbourne, Annals of Otology, Rhinology and Laryngology* Vol 104, No 9,Pt 2, Supp 166:75-76, 1995
2. Blamey, P.J., Martin, L.F.A., Clark, G.M. “A comparison of three speech coding strategies using an acoustic model of a cochlear implant”, *Journal of the Acoustical Society of America,* Vol 77,No 1:209-217, 1985
3. Clark, G.M. “The development of speech processing strategies for the University of Melbourne/Cochlear multiple channel implantable hearing prosthesis”, *JSLPA,* Vol 16, No2:95-107, 1992
4. Huang, T.-S., Wang, N.-M., Liu, S.-Y. “Nucleus 22-channel Cochlear mini-system implantations in Mandarin-Speaking patients”, *The American Journal of Otology* Vol 17: 46-52, 1996.
5. Kwok, C. L., Wong, C. M., So, K. W., Yiu, M. L., Lau, C. C., Luk, W. S., Tang, S. O. “Speech and lexical-tone perception in Cantonese-speaking cochlear implant patients”, *Australian Journal of Human Communication Disorders* Vol 19, No 2: 77-90, 1991.
6. Pijl, S., Schwarz, D.W.F. “Melody recognition and musical interval perception by deaf subjetcs stimulated with eletrical pulse trains through single cochlear implant electrodes”, *Journal of the Acoustical Society of America*, Vol 98 No 2 Pt :886-895, 1995.
7. Roederer, J. “Introduction to the Physics and Psychophysics of Music”, The English Universities Press, London 1973.
8. Tang, S. O., Luk, W. S., Lau, C. C., So, K. W., Wong, C. M., Yiu, M. L., Kwok, C. L. “Cochlear implant in Hong Kong Cantonese”, *The American Journal of Otology,* Vol 11, No 6: 421-426, 1990.
9. Whitford, L.A., Seligman, P.S., Everingham, C.E., Antognelli, T., Skok, M.C., Hollow, R.D., Plant, K.L., Gerin, E.S., Staller, S.J., McDermott, H.J., Gibson, W.R., Clark, G.M. “Evaluation of the Nucleus Spectra 22 Processor and the new speech processing strategy (SPEAK) in a postinguistically deafened adults”,*Acta. Otolaryngol (Stockh),*Vol 115:1-9, 1995.
10. Xu, S. A., Dowell,, R. C., Clark, G. M. “Results for Chinese and English in a multichannel cochlear implant patient”, *Clark and Busby, International cochlear implant symposium and workshop, Melbourne, Annals of Otology, Rhinology and Laryngology*, Vol 96,No 1,Pt 2,Supp 128:126-127,1987