TWO SPEECH PROCESSING SCHEMES FOR THE UNIVERSITY OF MELBOURNE MULTI-CHANNEL COCHLEAR IMPLANT PROSTHESIS*

Y.C. Tong, J.S. Chang, J.M. Harrison, J. Hugien, and G.M. Clark

Human Communication Research Centre, Department of Otolaryngology

University of Melbourne, Parkville 3052, AUSTRALIA

ABSTRACT

A series of speech perceptual studies were conducted on normally hearing subjects using synthesized acoustic signals based on two speech processing schemes: Zero-Crossing and Filter Bank. The Filter Bank scheme was shown to provide more speech information than the Zero-Crossing scheme. These schemes were implemented in a laboratory based speech processor which is described. A Low Power Switched Capacitor Speech Spectrum Analyzer embodying several novel design methodologies is also described. This spectrum analyzer can be used to implement the Filter Bank scheme in a wearable speech processor.

I. INTRODUCTION

In the design of speech processors that would convey useful speech information to cochlear implant recipients, extensive perception and engineering studies are required. These studies and work include:

1. Psychophysical studies to examine the nature of hearing sensations produced by electrical stimulation of residual auditory nerve fibres;
2. Formulation of speech coding strategies on the basis of psychophysical results;
3. Implementation of the chosen speech coding strategy using different signal processing schemes and different acoustic-to-electric parameter conversion algorithms;
4. Speech perceptual studies on normally hearing subjects using synthesized acoustic signals to assess the relative merits of potentially useful signal processing schemes; and
5. Speech perceptual studies on cochlear implant recipients to assess the relative merits and appropriateness of different signal processing schemes for electrical stimulation.

Results from psychophysical studies on cochlear implants have indicated that loudness increased with current level, pitch increased with electric pulse rate, and pitch and sharpness of the hearing sensation produced by individual electrode positions varied in accordance with the tonotopic organization of the cochlea [1,2]. On the basis of these psychophysical results, a logical speech coding is to convert sound intensity to electric current, the fundamental frequency of the speech signal to electric repetition rate, and the frequencies of spectral peaks of the speech signal to electrode positions. As far as the spectral peaks are concerned, formant frequencies may be estimated by counting the number of zero crossings at the output of formant filters, or spectral peaks may be detected by a peak picking algorithm at the output of a filter bank spectrum analyzer. In addition, the number of formants or spectral peaks would be one of the variables of the selected signal processing scheme.

The present signal processing scheme used in the University of Melbourne/Nucleus 22 channel cochlear implant [3] estimates the first and second formant frequencies by counting the number of zero crossings at the output of two formant filters, and converts these two frequencies into electrode positions [2]. This scheme will be referred to as the Zero-Crossing (ZC) scheme in this paper. A wearable speech processor has been developed and encouraging speech perception results were obtained from cochlear implant recipients [4].

In order to provide additional spectral information, a different speech processing scheme estimates the frequencies of 4 spectral peaks at the output of a filter bank spectrum analyzer and converts these spectral peaks to 4 electrode positions. This scheme will be referred to as the Filter Bank (FB) scheme.

In this paper, the amount of useful speech information provided by the two speech processing schemes, ZC and FB, will be compared [5]. A series of speech perceptual studies were conducted on four normally hearing subjects using synthesized acoustic signals. This study corresponds to that outlined in 4. above. The speech processing schemes were implemented in a real-time laboratory based speech processor which will be briefly described. The development of a Low Power Switched Capacitor speech spectrum analyzer integrated circuit [6] which can be used to implement the FB speech processing scheme in a wearable speech processor will also be described.

II. A REAL-TIME LABORATORY BASED SPEECH PROCESSOR

A real-time laboratory-based speech processor has been developed to permit speech processing and perceptual studies. The processor is interfaced with acoustic transducers for studies on normally hearing subjects, and cochlear implant hardware for cochlear implant recipients. It comprises 14 TMS32010 digital signal processors (DSPs) interfaced to an IBM compatible personal computer. Seven of the DSPs are used in the analysis front-end, while the remaining seven are used in the synthesis output section.

The ZC scheme, depicted in figure 1, used estimates of the first and second formant frequencies (F_1 and F_2) and their amplitudes (A_1 and A_2) to select the channel number and amplification factor of two out of twenty synthesis filters. The fundamental frequency, F_0 , and the presence of voicing were also determined. For voiced speech, the two synthesis filters were excited at a rate equal to F_0 , while for unvoiced speech, a random rate was used. The outputs of the two filters were summed to form the synthesized speech signal.

The FB scheme, depicted in figure 2, estimated a running spectrum of the speech signal using a 24 channel speech spectrum analyzer, each comprising a Bandpass filter, Full-Wave Rectifier, and Lowpass filter. The analysis front-end of the FB scheme implemented in the laboratory-based speech processor emulates the function of the Switched Capacitor Speech Spectrum Analyzer described in Section V. below. The 24 spectrum channels were processed by a peak picking and synthesis filter selection algorithm which selects four (out of 20) synthesis filters and their corresponding amplification

* The authors gratefully acknowledge the financial support of The Australian National Health and Medical Research Council, The Human Communication Research Centre (University of Melbourne), and The National Institutes of Health (USA) Contract No. 1-NS-5-2388 : 'Speech Processors for Auditory Prostheses'

factors. The rate of excitation was the same for the four filters, and was determined by F0 and the presence or absence of voicing as described for the ZC scheme. The outputs of the four filters are summed.

III. TESTING AND TRAINING SCHEDULES

Each subject was initially exposed to two sessions of synthesized speech containing only F0 information. These sessions were conducted to accustom the subject to degraded and synthesized speech before formal assessment and training so that order effects could be reduced. After these introductory sessions, the study followed a pre-training assessment, training and post-training assessment schedules for each scheme. Two of the four subjects were first tested and trained using the ZC scheme while the other two used the FB scheme. The test and training schedules were repeated for the unused scheme.

IV. RESULTS

The post-training percentage scores for consonant perception in noise are shown in figure 3 where 16 consonants in the /a/-/c/-/a/ frame were used. For all four subjects, scores were higher for the FB scheme than for the ZC scheme at signal-to-noise-ratios ranging from 5 to 20 dB. The results obtained using unprocessed (normal) speech materials are also shown in figure 3. At signal-to-noise ratios of 5 and 10 dB, it can be seen that the scores were highest for the unprocessed speech, medium for the FB scheme, and lowest for the ZC scheme. From figure 3, it can also be seen that the intersubject variations in the consonant perception were small for each of the three conditions : unprocessed speech, FB and ZC. These results indicate that the FB scheme provided more information that is pertinent to consonant perception than the ZC scheme. However, the performance for the FB scheme is far from perfect as indicated by the much better performance for the unprocessed condition.

Figure 4 depicts the pre-training and post-training percentage correct scores for the 16 consonant in the /a/-/c/-/a/ frame when no noise was added to the signal. As in figure 3, there were only small intersubject variations in the post-training scores. It is important to note that for all subjects, the effects of training, as indicated by the difference between the pre- and post-training scores, were larger for the speech processing scheme that was tested first. As an example, for subject 1, FB was trained and tested before ZC, the improvement in performance from pre- to post-training was much larger for the (FB) scheme that was tested first than for the second (ZC) scheme.

Vowel perception performance was similar across subjects and speech processing schemes for the 11 vowels in the /hvd/ format. There were also small improvements from pre-training to post-training. These results indicate that information pertinent to vowel perception was provided by both schemes.

Figure 5 depicts the pre- and post-training percentage correct scores for open set monosyllabic CNC words. For all four subjects, CNC performance was better for FB than for ZC, and as in the case of the consonant results in figure 4, the effects of training were larger for the speech processing scheme that was trained and tested first. For both FB and ZC, subjects 3 and 4 scored higher in CNC words than subjects 1 and 2.

In summary, the FB scheme provides more information than the ZC scheme for normally hearing subjects. Speech perception studies on cochlear implant patients are now being conducted to compare the performance of these two speech processing schemes.

V. A LOW POWER SWITCHED CAPACITOR SPEECH SPECTRUM ANALYZER

The development of a Low Power Switched Capacitor (SC) Speech Spectrum Analyzer which can be used to implement the FB scheme in a wearable speech processor is now described. The primary requirements are minimum chip area and power dissipation is now described. A Time-Multiplexed approach has been adopted as it allows operational amplifiers

(op amps) and some capacitors to be shared, hence reducing hardware requirements.

It comprises 24 channels, each of which consist of a cascade of Bandpass filter (BPF), Full-Wave Rectifier (FWR) and Lowpass filter (LPF). Several novel design methodologies have been employed in order to satisfy the requirements of the speech spectrum analyzer.

A transitional maximally flat magnitude (Butterworth) and maximally flat group delay (Bessel) filter approximation is used to in the SC BPF synthesis. This approximation provides a suitable compromise between the frequency and temporal resolutions.

DC offsets between different BPF channels is an important consideration as they would inadvertently be measured as energy of the bandlimited BPF outputs. Methods cited in literature to reduce these DC offsets include cascading a Highpass filter section to the preceding Bandpass filter section [7] and the use of resistive strings [8] to provide voltage division so that capacitor ratios of all BPF channels are made equal. The latter method results in identical DC transfer functions from the input of each biquadratic filter (biquad) op amp to the output of the biquad for all BPF channels. However, these methods are hardware inefficient in terms of chip area and power dissipation.

A solution proposed here is to design biquads such that the above mentioned DC transfer functions are *independent* of capacitor ratios. Figure 6(a) depicts such a Time-Multiplexed biquad whose input/output transfer function (V_{out}/V_{in}) and transfer functions from the inputs of op amp 1 and 2 to biquad output (V_{out}/V_{in1} , V_{out}/V_{in2}) are given in figure 6(c). At DC, V_{out}/V_{in1} and V_{out}/V_{in2} are shown to be -1 and 0 respectively. This result is significant because the BPF section is now not only hardware efficient but also micropower compatible without serious DC offset differences between BPF channels. A circuit is termed micropower compatible if all its op amps satisfy the following criteria : when an op amp samples an input, its output is not sampled by another op amp during the same instant. Usual design techniques are employed to further minimize DC offsets including using minimum sized complementary switches, clock signals with fairly slow turn off rates, and capacitors as large as tolerable. The Time-Multiplexed biquad uses the clock signals depicted in figure 6(b).

Capacitors of the BPF biquad vary considerably across channels due to the different BPF transfer functions. In figure 6(a), the non-integrating capacitors 'A', 'F', 'I' and 'U' do not carry charge information from one local clock period to the next. As the value of 'J' remains invariant for all channels, it is shared amongst all channels as depicted in figure 6(a), hence achieving some chip area saving.

A modular layout where equal areas are allocated to each channel of the spectrum analyzer, is desirable in an integrated circuit implementation to simplify interconnection and reduce the area due to interconnection. These equal allocated areas are primarily taken up by capacitors and because of the large differences in the total capacitance of the different channels, the allocated areas are quite large. For the Time-Multiplexed biquad used, the non-integrating capacitor 'A' has the largest capacitance variations across channels. The large variation in total capacitance of the BPF biquad can be reduced by employing a new capacitor sharing technique (figure 6(a)) applied to capacitor array 'A' which is now described. First, a common capacitor 'A0' of 1 unit capacitance is shared by all filter channels as the smallest capacitor of the 'A' array is greater than 2 units. In this manner, an 'A' capacitor is made up by connecting 'A0' in parallel to a residual 'Ax' capacitor (x being the BPF channel in consideration). The size of 'A' array can be further reduced if channels with large 'A' values, channels 9-24, share a further common capacitor 'AA' of 11 units. In these cases, the 'A' capacitor is realized by a parallel combination of 'A0' + 'AA' + 'Ax'. In this fashion, 'Ax' for channels 10-24 are reduced by 11 units each, hence the Time-Multiplexed BPF biquad capacitance is reduced by a notable

23%. This capacitor sharing corresponds to an approximate 15% chip area saving for the Time-Multiplexed BPF section. A clock signal Pd9-24, shown in figure 6(b), that goes high at the commencement of Pd9 and low at the end of Pd24, is used to connect 'AA' into the circuit, or alternatively, a bank of switches may be used. The proposed capacitor sharing technique is also applicable to array 'U' which has been simplified to a two capacitor array shown figure 6(a). Use of this array results in a chip area saving of 5% and simplifies the chip layout. Furthermore, as the smallest 'D' capacitor is increased to maintain the same capacitor ratio at the input of op amp 1, the worst case noise performance of the Time-Multiplexed BPF biquad is improved.

The SC FWR which is also new, is shown in figure 7. Op amp 1 and comparator 2 are autozeroed during the even phase, hence DC offset compensated. In the following odd phase, op amp 1 serves as a delay-free unity gain inverting amplifier and comparator 2 compares the FWR input and its inversion (obtained from the unity gain inverting amplifier). If the input is positive (negative) with reference to analog ground, the output is simply the input signal via 'path +' ('path -') in figure 7. The circuit thus functions as a FWR, i.e. $v_{out}(nT) = v_{in}(nT) \cdot \text{sign}\{v_{in}(nT)\}$. As a result of autozeroing during the even phase, all capacitors do not retain any charge information pertaining to the input of the FWR. The proposed FWR may therefore be used in a Time-Multiplexed application where one FWR services all filter channels, hence achieving substantial chip area saving. The sensitivity of the FWR is enhanced by designing the comparator to sense the differential input signal (instead of the FWR input and analog ground). It is also evident that the FWR is parasitic insensitive and jitter-free. SC FWRs cited in literature to date lack one or more of these features.

The LPF synthesis uses the Bessel approximation. All channels of the LPF section have identical transfer functions, a cutoff of 35 Hz. The LP filters are sampled at the same rate as the preceding FWR for aliasing considerations. This high sampling rate to cutoff frequency ratio results in a large capacitor ratios, hence requiring large chip areas. A multi-rate sampling technique is therefore adopted where the first order LP section serves as an antialiasing filter to the following biquad which is down sampled, hence reducing the capacitance ratios. The Time-Multiplexed LPF depicted in figure 8 is also micropower compatible.

A worthy feature of the complete speech spectrum analyzer comprising a cascade of the BPF section, FWR and LPF section presented in figures 6-8 is its *micropower compatibility*. Consequently, the speech spectrum analyzer can be realized with low power dissipation.

An experimental four channel speech spectrum analyzer has been fabricated using a 5 micron double polysilicon CMOS process. The integrated circuit layout depicted in figure 9 is very regular with an active area of 1.4 mm x 3 mm. Figures 10 and 11 respectively depicts the close agreement between the theoretical and measured frequency responses of the BPF and LPF sections.

Fig. 1 : ZC Scheme

Fig. 2 : FB Scheme

The accuracy of the running spectrum measurement of an SC speech spectrum analyzer is limited by the non-idealities of the BPF, FWR and LPF sections. The errors due to non-idealities, DC offsets of these sections and the finite sensitivity of the FWR, has been derived in [6]. Using the measured values obtained from the prototype integrated circuits and expressions derived in [6], it can be shown that the typical error of the running spectrum measurement of the spectrum analyzer is 0.9% for a 1V peak input signal. In the authors' view, this error is negligible for most speech applications.

The BPF dynamic range using $\pm 1.5V$ supplies is 64dB. The crosstalk between the BPF channels is -45dB. The measured impulse responses of the BPF section settle within 20 ms. These results satisfy typical speech spectrum analyzer specifications. The PSRR of the BPF section and FWR at 1 KHz is -20 dB and -25 dB respectively. PSRR of the LP section at DC is -35 dB.

With a $\pm 1.5V$ supply, the current drain of the prototype chip is 500 μA . The op amps used in this realization were deliberately over-designed for measurement purposes. An integrated circuit comprising 24 channels (168 poles and 24 FWRs), now under design, is expected to draw less than 2.5 mW with a $\pm 1.5V$ supply, hence suitable for its application in a wearable speech processor.

VI. CONCLUSIONS

Speech perceptual studies have been conducted which show that the Filter Bank speech processing scheme provides more speech information than the Zero-Crossing scheme. These schemes were implemented in a laboratory based speech processor which has been described. A Low Power Switched Capacitor Speech Spectrum Analyzer embodying several novel design methodologies has also been described.

REFERENCES

1. Y.C. Tong, R.C. Dowell, *et al.*, 'Psychophysical Studies evaluating the feasibility of a Speech Processing Strategy for a Multi-Channel Cochlear Implant', *J. Acoust. Soc. Am.*, 74, pp. 73-80, 1983.
2. Y.C. Tong, R.C. Dowell, *et al.*, 'Two Component Hearing Sensations produced by Two Electrode Stimulation in the Cochlea of a Totally Deaf Patient', *Science*, Vol. 219, pp. 993-994, 1983.
3. P.J. Blamey, R.C. Dowell, *et al.*, 'A Formant-Estimating Speech Processor for Cochlear Implant Patients', *Speech Communications*, Vol. 6, pp. 293-298, 1987.
4. G.M. Clark *et al.*, *The University of Melbourne-Nucleus Multi-electrode Cochlear Implant*, Ed. C.R. Pfaltz, S. Karger, Switzerland, 1987.
5. Y.C. Tong, J.S. Chang, *et al.*, National Institutes of Health (USA) Quarterly Progress Reports: 'Speech Processors for Auditory Prostheses', Contract N01-NS-5-2388, 1986-1988.
6. J.S. Chang and Y.C. Tong, 'A Low Power Switched Capacitor Speech Spectrum Analyzer', Submitted to IEEE JSSC.
7. L.T. Lin *et al.*, 'A Monolithic Audio Spectrum Analyzer', *IEEE ASSP*, Vol. 31, pp. 288-293, 1983.
8. Y. Kuraishi, *et al.*, 'A Single-Chip 20 Channel Speech Spectrum Analyzer Using a Multiplexed Switched Capacitor Filter Bank', *IEEE JSSC*, Vol. 19, pp. 964-970, 1984.

Fig. 3

Fig. 4

Fig. 5

Fig. 6 : Time-Multiplexed BP Biquad

Fig. 7 : FWR

$$\frac{V_{out}(z)}{V_{in}(z)} = \frac{Jz^{-1}(1-z^{-1})}{(1+F) + z^{-1}(AU-F-2) + z^{-2}}$$

$$\frac{V_{out}(z)}{V_{in}(z)} = \frac{-AUz^{-1}}{(1+F) + z^{-1}(AU-F-2) + z^{-2}}$$

$$\frac{V_{out}(z)}{V_{in}(z)} = \frac{(A-F) + z^{-1}(-A-F+J) - Jz^{-2}}{(1+F) + z^{-1}(AU-F-2) + z^{-2}}$$

$$\left. \frac{V_{out}(z)}{V_{in}(z)} \right|_{z=1} = -1$$

$$\left. \frac{V_{out}(z)}{V_{in}(z)} \right|_{z=-1} = 0$$

(b)

Fig. 8 : Time-Multiplexed LP Filter

Fig. 9 : Microphotograph of 4 Channel SC Spectrum Analyzer

Fig. 10

Fig. 11

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Tong, Yit C.;Chang, J. S.;Harrison, J. M.;Hugien, J.;Clark, Graeme M.

Title:

Two speech processing schemes for the University of Melbourne multi-channel cochlear implant prosthesis

Date:

1989

Citation:

Tong, Y. C., Chang, J. S., Harrison, J. M., Hugien, J., & Clark, G. M. (1989). Two speech processing schemes for the University of Melbourne multi-channel cochlear implant prosthesis. In IEEE International Symposium on Circuits and Systems, Portland, Oregon.

Persistent Link:

<http://hdl.handle.net/11343/26837>