COMPUTATIONAL MODEL OF THE PERIPHERAL AUDITORY SYSTEM FOR SPEECH RECOGNITION: INITIAL RESULTS

Ara Samouelian † and Clive D. Summerfield †
†School of Electrical Engineering,
The University of Sydney

ABSTRACT - This paper describes the design of a computational model of the peripheral auditory system, which is controlled via the AUDLAB Interactive Speech Signal Processing Package using a programmable harness to interface the AUDLAB command protocol and track file format to the structural model of the cochlear processor. A suite of signal processing modules, originally developed for speech synthesis research has been supplemented by a number of non-linear signal processing modules to model the transduction stage of the Cochlear. Some initial results of the cochlear processor model and its performance on real speech signals are presented.

INTRODUCTION

Reliable recognition of speech is fundamental to man-machine commmunications. Although there exists a number of different recognition strategies, such as dynamic time warping, Markov modelling, neural networks and phonetic feature extraction, all of these approaches rely on the effectiveness of the input signal processing to generate the parameters upon which the recognition is performed.

Traditionally, input speech signal analysis is performed by using linear signal processing algorithms, such as Fourier Transforms or Linear Predictive Coefficient analysis. However, recent results from extensive investigations by Seneff (1988) show that there is significant advantages in using models of the cochlear.

The logarithmic frequency scaling used in these mdels closely matches the frequency resolution known to be effective for speech perception. Moreover, the non-linear temporal properties of the model, enhances the acoustic features of the speech signal known to be of central importance in speech recognition. Ghitza (1986) also presents evidence that synchronous auditory based models exhibit significant improvement in recognition performance in high noise environment .

As well as these advantages, there are significant engineering motivations for developing a computational model of the peripheral auditory system. The computational model exhibits a high degree of structured regularity and process concurrency, and lend themselves to very efficient VLSI implementation using bit-serial design approach primarily reported by Lyons (1982).

Our ultimate aim is to investigate the design of a cochlear model VLSI chip which can be used as a front end signal processing module for a speech recognition unit, incorporating processes which enhance speech recognition performance. The design strategy utilises a hierarchical design approach, which has previously been used to implement VLSI formant speech synthesis ASIC (Summerfield, 1988).

The AUDLAB Interactive Speech Signal Processing software package was developed at the Centre for Speech Technology Research, at the University of Edindurgh and is designed primarily for speech recognition research. It extensively uses real-time data I/O capability of the MASSCOMP UNIX to record and reproduce speech signals and the colour graphics interface to display speech waveforms and analyse results. The package has a large inventory of speech signal processing programmes which can be easily combined to produce complex signal processing algorithms. Thus, AUDLAB provides an interactive, flexible and extremely versatile and productive research environment for cochlear model processing for speech recognition.

The present computational model used in this research has been implemented within AUDLAB. Our aim is by using this approach to provide a mechanism to investigate the trade-off between structural complexity and the performance of the model. AUDLAB also provides an ideal environment for evaluating and determining the most prominant features of speech, which need to be extracted to enhance the speech recognition signal.

SOFTWARE STRUCTURE OF THE COMPUTATIONAL MODEL

The computational model of the peripheral auditory system utilises a set of generic signal processing modules, written in the 'C' programming language. All the modules share a common generic communication protocol, which enables a structured model of the cochlear process to be constructed, using the UNIX piping, redirection and tee facilities. The suite of linear signal processing modules ,such as resonators (poles), anti-resonators (zeros) and differential filters, which were originally developed for speech synthesis research, has been supplemented by a number of non-linear signal processing modules, such as detectors, adaptors and automatic gain control to model the cochlear transduction stage. Construction, calibration and execution of the signal processing C-shell scripts is controlled by a software harness, which interfaces to the AUDLAB Interactive Speech Processing Package.

Figure 1 shows the software architecture of the AUDLAB harness. The cochlear processor architecture is controlled via the AUDLAB Interactive Speech Signal Processing Package using programmable harness to interface the AUDLAB command protocol and track file format to the structural model of the cochlear processor. The programmable harness is installed in the AUDLAB as an external signal processing module and is activated via the internal AUDLAB menu controls.

The harness calls the relevant programming modules and parses to it the appropriate parameters such as number of filters, bark spacing, and sampling rates to create C-shell UNIX scripts. It also controls AUDLAB track file format conversion and file header manipulation necessary to process the speech sample data files. It creates appropriate track files to display within AUDLAB the filter characteristics, spectogram of the wide band input speech signal and the cochleogram.

The harness can perform several processes from simple menu selection within AUDLAB. These in-

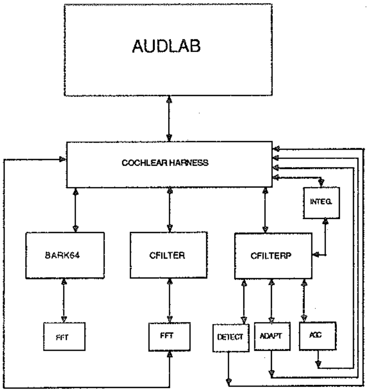

Figure 1: Software architecture of the AUDLAB harness

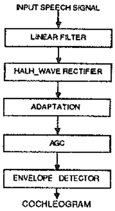

Figure 2: Computational model

clude the calibration of the auditory filterbank , calculation of the spectogram of the wideband speech input signal, the calculation of the output stage of each non-linearity module for evaluation and the calculation of the cochleogram, representing the output of the transconduction stage.

By utilising signal processing modules developed for the VLSI formant speech synthesis research and by adding new modules to represent the transconduction stage of the Organ of Corti, we can provide an established and proven development path from structural specifications to VLSI fabrication using FIRST language compiler and word level.

PERIPHERAL AUDITORY MODEL

The computational model of the cochlear is shown in figure 2. It consists of a set of auditory filters followed by the non-linearities, which capture the prominant features of the transformation from basilar membrane vibration to nerve fiber probabilistic response. The output is then envelope detected for display as cochleogram corresponding to the mean rate response of auditory neurons.

The Model presently under investigation consists of 64 overlapping critical band filters, spaced at 0.3

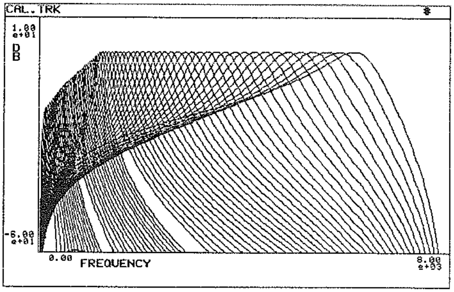

Figure 3: Frequency response of the auditory filters

Bark spacing and spanning a frequency range from 100 to 6310 Hz. Each filter has a bandwidth of 0.5 Bark, and an amplitude which is automatically calibrated to a given gain profile. Using the Bark spacing to determine the centre frequency of the filters produces a psychoacoustic masking filter instead of a sharp bandpass filter. Figure3 shows the frequency response of the auditory filters, plotted on a linear scale.

Following the auditory filterbank, each channel is processed by a half-wave rectifier and soft limiter (detector). This is followed by a short term adaptation circuit (adapt) and a final stage of rapid automatic gain control (agc). Since these elements are non-linear,ordering is important.The sequence used in this model is based on the structure proposed by Seneff's .

INITIAL RESULTS

Figure 4 shows an example of a sampled speech signal, for a short segment of a male speaker's voiced speech,during the /tᵊ/ of the word "centre". Figure 5 shows the spectogram of the wideband speech signal. The cochleogram representing the output of the transconduction stage is shown in figure 6.

The results shown in figures 5 and 6 show the high frequency contents of the burst release section of unvoiced alveolar stop, voicing and formant structures of the neutral swhar vowel. Note the enhanced onset properties of the cochlear model.

DISCUSSION

The main advantage of the design of the self calibrating computational model of the peripheral auditory system and its implementation on a MASSCOMP 5520 computer running a Real-time UNIX Operating System is that the model utilises a set of generic signal processing modules. This enables the cochlear signal processing function to be defined at the structural level, using high level signal processing functions, such as resonators, anti-resonators and differential filters, using piping and redirectional facilities of the UNIX Operating System. Cochlear processor architecture is controlled via the AUDLAB Interactive Speech Signal Processing Package using a programmable harness to

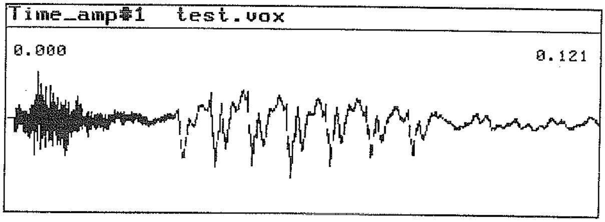

Figure 4: Sampled speech signal

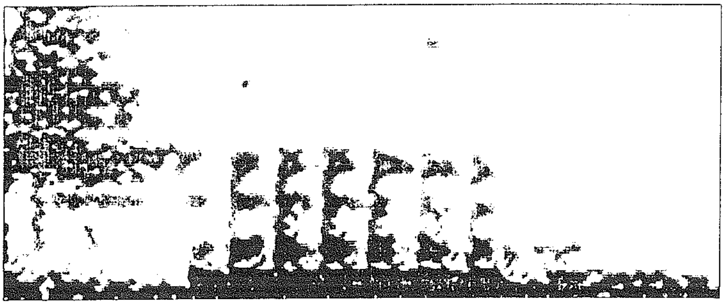

Figure 5: Wide-band spectogram

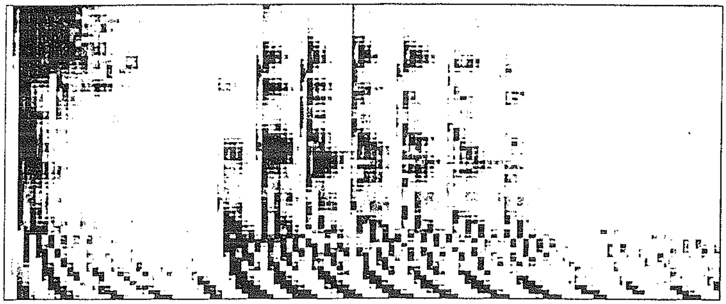

Figure 6: Cochleogram

interface the AUDLAB command protocol and track file format to the structural model of the cochlear processor. The suite of signal processing modules, originally developed for speech synthesis research has been supplemented by a number of non-linear signal processing modules to model the transconduction stage of the Organ of Corti.

Simple signal processing modules implemented in 'C' can be conveniently piped together to produce a peripheral auditory model, which has its variables such as number of filters, Bark spacing, sampling rate set externally through the programmable harness, which interfaces the AUDLAB command protocol and track format to the structural model of the cochlear processor.

The cochlear model is intended to be used as the the input signal processor of speaker independent speech recognition research and is still at an early stage of development. The scheme presently under development utilises transputer distributed processes for acoustic feature extraction and a knowledge based systems for the recognition task. Using the auditory model in combination with AUDLAB and the various signal processing modules provide an interactive environment for research into engineering aspects of a cochlear model for speech recognition.

REFERENCES

Ghitza O. (1986) "Auditory nerve representation as a front-end for speech recognition in noisy environments", Computer Speech and Language, 1, pp 109 -- 130.

Lyons R. F. (1982) "A computatinal Model of Filtering, Detection, and Compression in the Cochlea", IEEE Int. Conf. ASSP pp 1282 -- 1285.

Seneff S. (1988) "A joint synchrony/mean-rate model of auditry speech processing", Journal of Phonetics, 16, pp 55--76.

Summerfield C. D. & Jabri M. A. (1988) "A Formant Speech Synthesis ASIC: Functional Design", this conference.