AN AUDITORY MODEL ASIC FOR SPEECH RECOGNITION: FUNCTIONAL DESIGN

A. Samouelian † A. Markus † and C. D. Summerfield ‡

†School of Electrical Engineering,
The University of Sydney
‡Syrinx Speech Systems Pty Ltd

ABSTRACT - This paper describes the functional design of an ASIC auditory model for use as a front-end signal processor for speech recognition. The model consists of a set of critical band auditory filters, followed by non-linearities, representing the transduction stage of the Organ of Corti. The initial functional design was performed using the Denyer and Renshaw FIRST Silicon compiler. The compiler was modified to accommodate look-up tables to implement the compressive rectifier. Initial simulation results indicate that a bank of 32 filters, spanning a centre frequency range of 100 to 6003 Hz, using a Bark spacing of 0.6 can be implemented on a single chip. Results of the simulation and its performance on real speech signals are presented.

INTRODUCTION

Computational models based on the processes of the cochlea have come into prominance in recent years as alternative front-end signal processors for speech recognition. One of their main attraction is that they have been shown to offer superior performance in high ambient noise environment (Ghitza 1986). This is specially attractive for deploying speech recognition in Telecommunication applications. Computational models of the cochlea take into account the logarithmic frequency scaling of the ear and the non-linear temporal properties of the cochlea. Such a computational model of the peripheral auditory system, has been under development at Sydney University for the past year (Samouelian 1990). Since the model exhibits a high degree of structured regularity and process concurrency, it lends itself to efficient ASIC implementation.

The model consists of complex algorithms combining linear and non-linear elements. In order to achieve an efficient design, the algorithm is partitioned and simulated using generic signal processing modules, such as complex pole pairs (resonators), complex zero pairs (anti-resonators) and real zeros (differantiator), which were originally developed for speech synthesis research. These have been supplemented by a number of non-linear signal processing modules, such as compressive rectifiers, adaptors and automatic gain controls to model the cochlea transduction stage.

The motivation behind the development of the model using these modules was based on the opportunity it offered to use the established signal processing elements, which have been implemented

in ASIC form, thus providing us with a proven path from functional specification to ASIC, as was demonstrated in the speech synrhesis ASIC (Summerfield & Jabri 1989).

The basis for the ASIC design is a bit-serial design approach. Speech signals have relatively narrow bandwidth (5 kHz), large dynamic range (60 to 100 dB) and complex spectral and temporal structure. The bit-serial ASIC design allows processing bandwidth and dynamic range trade-offs to be exploited without unduly effecting chip size. Furthermore, using a bit-serial ASIC minimises communication overheads in functionally complex processes.

The VLSI auditory model has been developed using the method of "functional design" as implemented in the Denyer/Renshaw FIRST Silicon Compiler (Denyer & Renshaw 1985). FIRST offers a hierarchical design environment which is designed for the rapid implementation of fully synchronous bit-serial signal processing architectures.

FUNCTIONAL DESIGN

The cochlea model consists of a bank of linear filters followed by the non-linearities. The implementation of the filterbank is shown in Figure 1. This was originally implemented using a set of generic signal processing modules, written in 'C'. All the modules shared a common generic communication protocol, which enabled a structured model of the cochlea processes to be constructed, using the UNIX piping, redirection and tee facilities (Samouelian & Summerfield 1989).

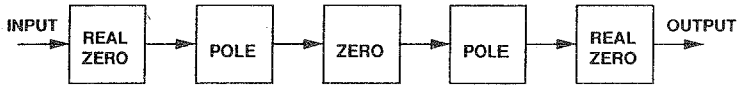

Figure 1: The model of a single filter

The above is the model of a single filter, which was multiplexed 32 times to produce the cochlea model filterbank. This particular sequence was adopted as it gave optimal configuration for 16 bit integer arithmetic.

Figure 2 shows the stability of the filter model. It can be seen that the maximum input signal value (sinmax) that may cause integer overflow is frequency dependent. Since a significant portion of information in a speech signal is contained in the region above 1 kHz, the performance of the filter structure was optimised to perform over this range.

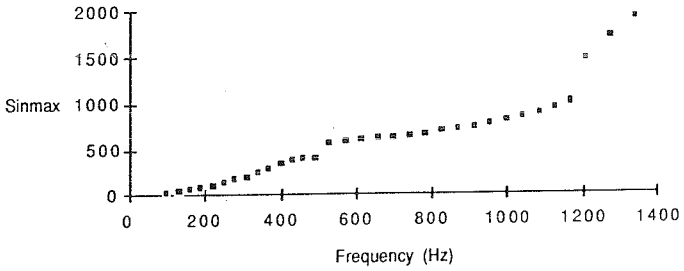

Figure 2: Stability of the filter model

FUNCTIONAL OPERATOR STRUCTURE

The operator structure of the filter model is shown in figure 3. It consists of a differential filter that provides real zero at the input to the core signal processing unit and a second differential filter that

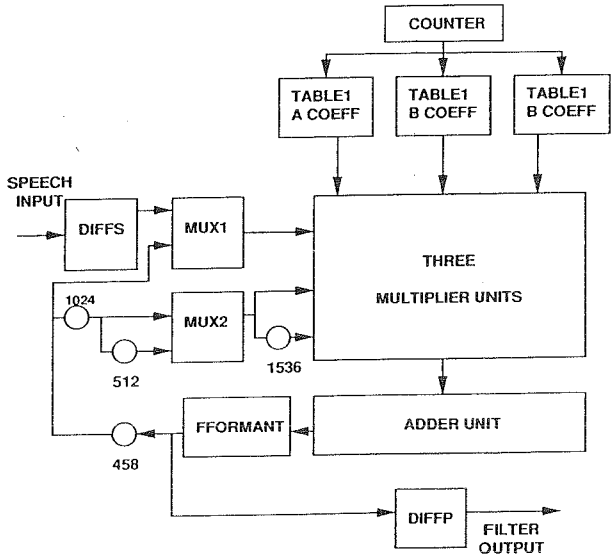

Figure 3: Operator structure of the filter model

provides a real zero at the output. The core signal processing unit consists of a function primitive operator containing three independent double precission 16 bit serial multipliers, followed by a network of double precission bit-serial adders, which implement the pole and zero difference equations.

The complete circuit is controlled by a 96 phase multiplier clock, which enables each of the three second order filters in each cochlea model filter bank to be computed sequentially.

The sequencing of the filter calculations is important to minimise the number of delay line elements used in the design. Sequencing is arranged to compute the all input pole filters first, followed by the zero filters, followed by the final pole filters. The complete filter calculation is 1,536 clock calculations. This translates to an operating frequency of 15.36 MHz (for a 10 kHz sampling rate). This is well within the clocking rate for contemporary CMOS technology.

RESULTS

The filterbank consists of 32 critical band auditory filters, spaced linearly at 0.6 Bark spacing and spanning a frequency range from 100 to 6003 Hz. Each filter has a bandwidth of 0.5 Bark. Figure 4 shows the theoretical frequency response of the 32 auditory filters, and the results of the filter model simulation using an input pulse of an amplitude of 8191.

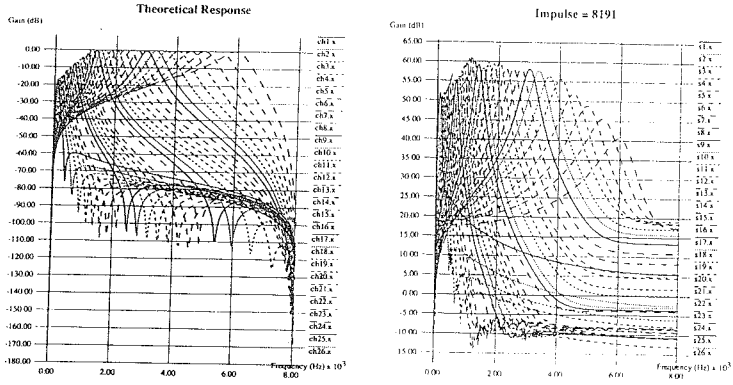

Figure 4: Frequency response of 32 auditory filters

Figure 5 show the performance of the ASIC simulation on the word "speech" as spoken by an male speaker. It should be noted that only the filterbank section has been fully simulated, the output

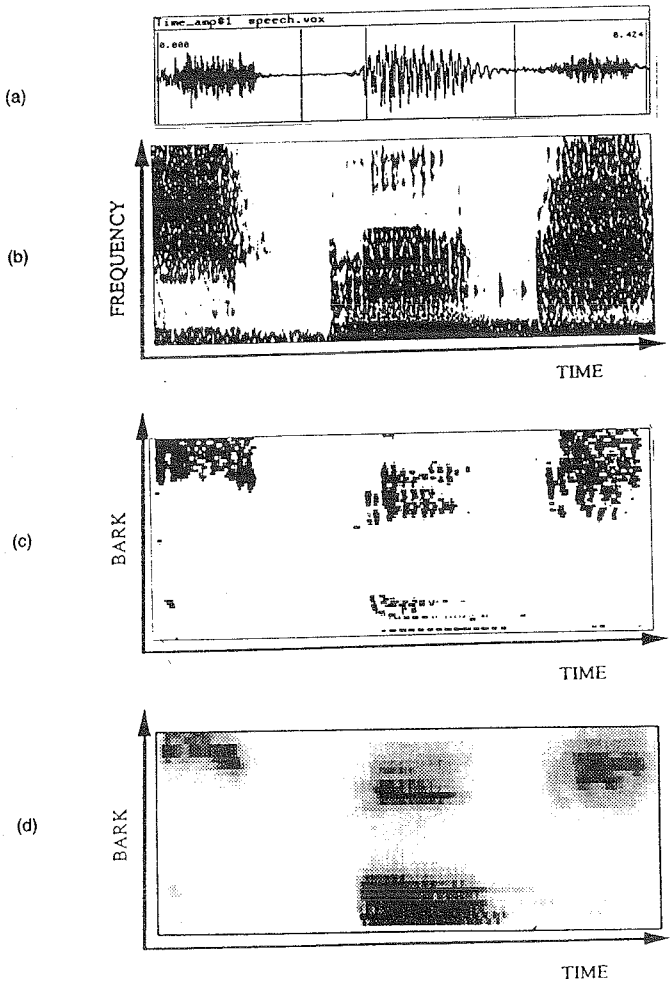

Figure 5: Performance of the ASIC simulation on the word "speech"

demultiplexed and used as the input signal to the suite of non-linear signal processing modules, implemeted in 'C'. Figure 5(a) show the segment of the speech waveform. Traditional wideband speech spectrogram calculated using 256 point FFT is shown in Figure 5(b). Figure 5(c) shows the cochleogram obtained by processing the signal through the linear filterbank and non-linearities, and then envelope detected for display. The output of the ASIC simulation is shown in Figure 5(d).

CONCLUSION

This paper has shown that by utilising signal processing modules developed for speech synthesis ASIC, we can implement two sets of 32 auditory filterbanks on a single ASIC, using bit-serial compilation techniques. Additional work is required to implement the non-linearities in Silicon. Work is progressing on the implementation of the ASIC cochlea as a robust front-end signal processor for speaker independent speech recognition.

ACKNOWLEDGEMENTS

This work was funded by an Australian Research Council (ARC) grant No. A48830416.

REFERENCES

Denyer P. & Renshaw D. (1985). "VLSI signal processing: a bit serial approach", Addison-Wesley Publishing Company. Australia, pp 112-115.

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

Samouelian A. (1990). "Speech recognition front-end using auditory model", to be published in Int. Conf. on Signal Proc. '90 proceedings, 22-26 Oct., 1990, Beijing, China.

Samouelian A. & Summerfield C. D. (1989). "Front-end speech signal processor for speech recognition", Proc. IREECON89 Int. Conf., September 1989, Melbourne, Australia, pp 112-115.

Samouelian A. & Summerfield C. D. (1988). "Computational model of the peripheral auditory system for speech recognition: Initial results", Proc. Second Australian Int. Conf. on Speech, Science and Technology, November 1988, Sydney Australia, pp 234-239.

Summerfield C. D. & Jabri M. A. (1989). "Design and implementation of a Formant speech synthesiser ASIC", ICASSP May 1989, Glasgow, Scotland.