

Unified Theory of Speech : Speech Processing for the Hearing Impaired and Beyond

N. Magotra, F. Livingston, S. Savadatti
Texas Instruments Incorporated
12203 Southwest Freeway
Stafford, TX 77477

Abstract

This paper presents some speech processing algorithms that were originally developed for hearing aid applications. However these algorithms are also applicable for other speech and audio applications. Considering that the basic properties of speech remain invariant across applications, it is logical to consider these algorithms under the broader umbrella of 'unified theory of speech.' These algorithms have been implemented on Texas Instruments (TI) floating-point processor, the TMS320C3X (C3X). Portable C3X units were tested on hearing impaired subjects to assure the efficacy of the various speech processing algorithms. As these algorithms are being ported to TI's low-power fixed-point Digital Signal Processing (DSP) chips – TMS320C54X and TMS320C55X, they are also being made XDAIS compliant. XDAIS is the acronym for eXpress DSP Algorithm Standard. It is a software tool developed by TI which simplifies algorithm integration in today's complex DSP algorithm systems and also makes it easier to port algorithms across applications.

1. Introduction

The ultimate goal of this research is to facilitate the development of a fully programmable portable digital hearing aid that conditions the input speech signal based on environmental circumstances and the hearing impaired person's hearing loss - a fully customized digital hearing aid. Over the past several years, we have developed a prototype Programmable Digital Hearing aid (PDH) system consisting of a laboratory-based PC system with a TMS320C3X DSP card and a wearable unit also based on the TMS320C3X DSP chip. The PC-based system is used to customize the hearing aid for a particular patient and the final algorithm is downloaded to a read-only memory (ROM) chip on the wearable unit [1].

These Algorithms are also being tested for other speech and audio applications. Since they are being developed using TI's XDAIS tool, they are easy to port across diverse applications.

We are also porting these algorithms to our fixed-point DSP chips. Typically, floating-point chips are larger and consume more power. However, it is usually easier to implement DSP algorithms on these chips since

quantization effects are negligible for most applications. Fixed-point chips on the other hand tend to be smaller in size and have lower power requirements. However, the algorithms have to be analyzed for the effects of quantization noise on their performance and modified if necessary prior to porting to the fixed-point platform. The critical finite register length effects are [2]: errors due to A/D conversion, errors due to arithmetic round-off, quantization of system parameters and bounds on the signal levels in order to prevent overflows. We are currently carrying out detailed quantization analysis of these algorithms.

2. Algorithms

PDH is a binaural system capable of sampling two input microphone signals at sampling rates of up to 32 kHz per channel. The minimum bandwidth of the hearing aid is 10 kHz. At present we are researching three critical algorithms - Spectral Shaping (SS), Adaptive Noise Reduction (NR) and Multiband Dynamic range Compression (MDC). Figure 1 represents a palate of speech processing algorithms that could possibly be implemented on PDH because of its programmable design. A brief overview of the SS, NR and MDC algorithms is presented below.

Frequency Shaping

This signal processing stage shapes the speech spectrum (0 up to 16 kHz) to compensate for the patient's hearing loss. The speech data is input to a binaural equalizer implemented as two banks of bandpass filters, one for each channel (ear). These filters are designed to have perfectly linear phase and high (>80 dB) band isolation. The therapist can interactively (in real-time) choose the number of filters in each bank and select their critical frequency characteristics namely, their cutoff frequencies and isolation between different frequency bands. Figure 2 shows the resulting magnitude plot tuned for a particular patient's hearing loss. Because of the design flexibility built into the SS algorithm, the same PDH unit can be reprogrammed for different languages [1].

Adaptive Noise Reduction

The noise reduction algorithm used in PDH has been designed to work with just one input data channel. Hence

PDH processes the two (right and left ears) single input-single output channels independently. For each channel, the input signal is first high pass filtered to compensate for the low frequency spectral tilt [3] in speech signals. The high pass filter is a simple first order infinite impulse response (IIR) filter with tunable cutoff frequency. The noise reduction algorithm is referred to as Real-time Adaptive Correlation Enhancer (RACE). RACE is essentially an adaptive finite impulse response (FIR) filter. The autocorrelation coefficients of the input speech are used to update the RACE (FIR) coefficients after applying suitable gain control [4]. Figure 3 shows a generic block diagram for the RACE algorithm. The autocorrelation function is estimated using a recursive estimator represented by the equation below

$$\hat{R}_{xx}(n, k) = \mathbf{b} \hat{R}_{xx}(n-1, k) + (1 - \mathbf{b})x(n)x(n+k)$$

where $\hat{R}_{xx}(n, k)$ and $\hat{R}_{xx}(n-1, k)$ are the autocorrelation estimates of the input $x(n)$ at lag value k and at times n and $n-1$ respectively and \mathbf{b} is a smoothing constant whose value lies between 0 and 1. The autocorrelation coefficients are estimated for lag values of k ranging from $-L$ to $+L$ where L is known as the maximum lag. This results in a set of $(2L+1)$ autocorrelation coefficients. Typically 'L' is chosen to lie between 5 and 7. Both stability and rate of convergence are dependent on a single parameter, namely the smoothing constant \mathbf{b} . A large value of \mathbf{b} implies slow adaptation while a small value of \mathbf{b} implies faster adaptation. Typically we set its value to correspond to a time constant of 4 msec for a 20 kHz sampling rate. This is less than the short term stationarity of speech which is generally between 5-20 msec. Hence the values of \mathbf{b} and L which determine the length of the adaptive FIR filter, should be chosen, so as not to exceed the short term stationarity assumption of speech [3].

Figure 4 shows the Power Density Spectra (PDS) corresponding to three time traces – a clean sinusoid, a signal (input to RACE) consisting of the sinusoid corrupted with additive White Gaussian Noise (AWGN), and the RACE output signal for this particular input. From the figure it is evident that RACE reduces the amount of background white noise considerably. The noise floor is reduced by approximately 17 dB.

Similar tests conducted using CUNY nonsense syllables yielded SNR improvements on the order of 6 dB.

Amplitude Compression

In addition to frequency shaping and noise suppression, the platform also permits real time implementation of multiband amplitude compression. Speech amplitude compression is essentially the task of controlling the overall gain of a speech amplification system. It essentially

"maps" the dynamic range of the acoustic environment to the restricted dynamic range of the hearing impaired listener. Amplitude compression is achieved by applying a gain of less than one to a signal whenever its power exceeds a predetermined threshold [6]. Amplitude compression is based on the average power in the signal. The time constant of power estimation is used to modify the attack/release time of the compression algorithm [6]. If $x(n)$ is the discrete time input signal, its estimated power $p(n)$ is given by,

$$p(n) = \mathbf{b}p(n-1) + (1 - \mathbf{b})x^2(n)$$

As long as the input power, to the compressor, is less than a preset threshold no compression takes place and the input is equal to the output. When the input power exceeds the threshold, the signal is attenuated. Once amplitude compression is being applied, if the attenuated input power exceeds a specified saturation power level the output power is held at a constant (saturated) level.

The simulation data plots shown in Figure 5 demonstrate the efficiency of the compression algorithm described above. The top trace in Figure 5 corresponds to the nonsense syllable 'za'. The lower plot in this figure represents the corresponding compressed output with a lower threshold of 0.15. Note that there is no 'spiking' in the time domain - a common artifact in amplitude compressed signals. Typically these spikes are representative of the compression algorithm missing the first few cycles of the loud sound. However, the power estimation employed in our implementation circumvents this problem.

3. Conclusions and Future Work

The research summarized in this paper indicates that we can successfully implement various speech processing algorithms on Texas Instruments family of DSP chips – floating-point as well as fixed-point. The results indicate that we can do so without sacrificing speech output quality. This is quite critical since the TMS320C55X chip is on target to be low enough in power consumption and small enough in footprint, to be used in developing *true* digital hearing aids that can be totally customized to individual hearing impaired subjects. Keeping in mind that these algorithms can be ported for other applications, they are being developed using TIs XDAIS tools.

An added benefit of this portable prototyping device is that the efficacy of various speech processing strategies, including the algorithms described in this paper, can be measured using test subjects under real world conditions.

REFERENCES

- [1] N. Magotra, S. Sirivara, "Real Time Digital Speech Processing Strategies for the hearing impaired," ICASSP '97, Munich, Germany.
- [2] C. T. Mullis, R. A. Roberts, "Digital Signal Processing," Addison Wesley, 1987.
- [3] L. R. Rabiner, R. W. Schafer, "Digital Processing of Speech Signals," Prentice-Hall, 1978.
- [4] N. Magotra, F. Livingston, S. Rajagopalan, "Single Channel Speech Enhancement in Real-time," 27th

Asilomar Conference on Signals, Systems and Computers, Asilomar, CA, Nov 1993.

- [5] S. Bangalore, "Quantization Analysis for Fixed-Point Implementation of Speech Processing for the Hearing Impaired," MS Thesis, EECE Department, University of New Mexico, 1999.

[6] Sudheer Sirivara. Digital signal processing strategies for the hearing impaired. M.S Thesis, Electrical and Computer Engineering Department, University of New Mexico, July 1997.

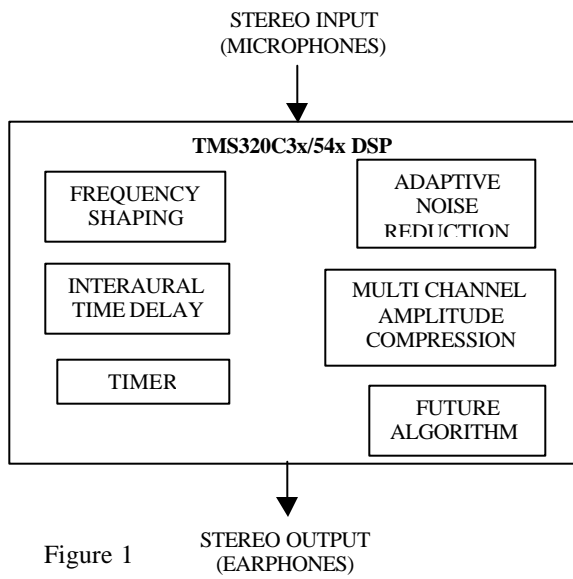


Figure 1

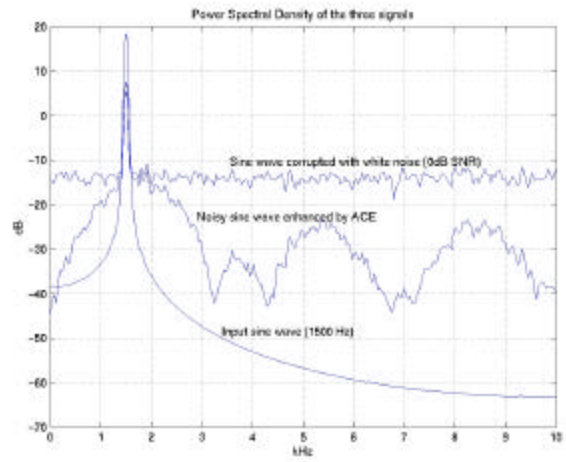


Figure 4

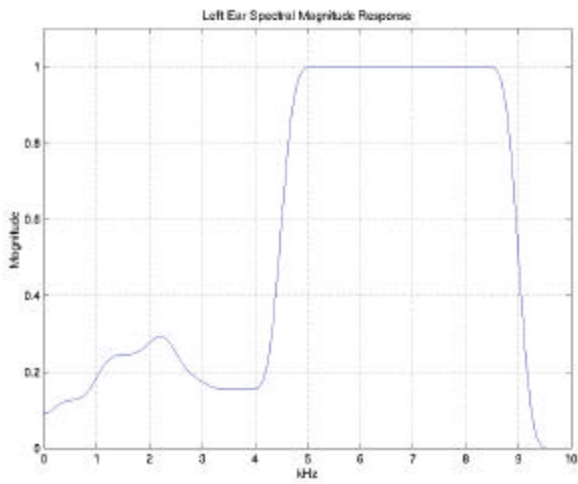


Figure 2

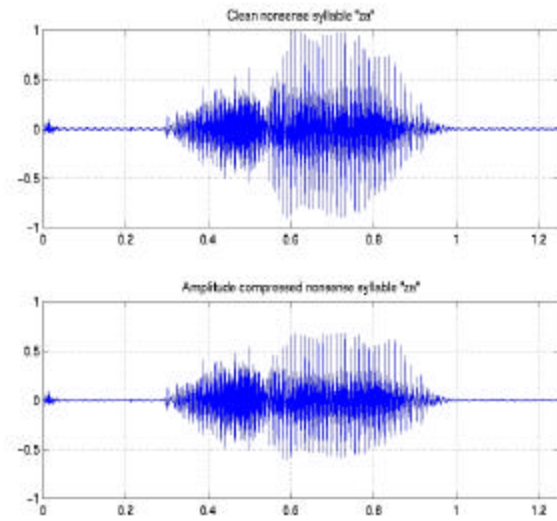


Figure 5

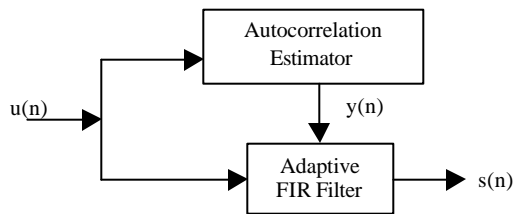


Figure 3