Enhancement of Electrolaryngeal Speech

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Master of Technology

by

S. Khadar Basha
(Roll No. 08307R01)

under the supervision of

Prof. P. C. Pandey

Department of Electrical Engineering
Indian Institute of Technology Bombay
June 2011

Indian Institute of Technology Bombay

M.Tech. Dissertation Approval

This dissertation entitled **"Enhancement of electrolaryngeal speech"**
by **S. Khadar Basha** (Roll No. **08307R01**)
is approved, after the successful completion of *viva voce* examination,
for the award of the degree of **Master of Technology in Electrical Engineering**.

Supervisor	.....	(Prof. P. C. Pandey)
Examiners	.....	(Prof. Preeti Rao)
	.....	(Prof. M. S. Shah)
Chairperson	.....	(Prof. R. K. Joshi)

Date: 27 June, 2011
Place: Mumbai

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(Shaik Khadar Basha)

Date: 26 June, 2011

Place: Mumbai

S. Khadar Basha / Prof. P.C. Pandey (Supervisor): “Enhancement of electrolaryngeal speech”, *M.Tech. dissertation*, Department of Electrical Engineering, Indian Institute of Technology Bombay, June 2011.

ABSTRACT

An electrolarynx, a verbal communication aid used by laryngectomy patients, is a vibrator held against the neck tissue to provide excitation to the vocal tract, as a substitute to that provided by the glottal vibrations. Electrolaryngeal speech suffers from a monotonic nature, low-frequency spectral deficit, and background noise due to leakage from the vibrator. While the first two factors affect the speech quality, the background noise also affects the intelligibility. This project presents the investigations for enhancement of electrolaryngeal speech and real-time implementation of the enhancement techniques. Pitch-synchronous application of generalized spectral subtraction is used for reducing the background noise. Effects of different noise estimation and phase estimation techniques are investigated. Spectral compensation, and introduction of jitter and shimmer in the speech signal, using LPC based analysis-synthesis, is investigated for improving its naturalness. Two real-time implementations of the spectral subtraction for enhancement of electrolaryngeal speech are carried out using 16-bit fixed-point processors: first using dsPIC33FJ128GP804 based single chip circuit with sampling frequency of 10 kHz and subsequently using TMS320C5515 based board with sampling frequency of 12 kHz. In both the implementations, input and output are handled using DMA and memory buffers for block processing using two-pitch period analysis window with 50 % overlap. The noise is estimated using 3-point 4-stage median and speech is resynthesized using noisy phase. In both the implementations, the noise reduction is compatible with that obtained by Matlab based simulations.

CONTENTS

List of symbols

Symbol	Explanation
b_m	FIR filter coefficients
$c(n)$	Cepstral coefficient
$e(t)$	Excitation pulse
$h_v(t)$	Impulse response of vocal tract
$h_l(t)$	Impulse response of leakage path
$l(t)$	Leakage sound
$L(k)$	Average magnitude spectrum of noise
N	Window length
$s(t)$	Speech sound
$x(t)$	Noisy speech
$y\ (t)$	Cleaned speech
α	Over-subtraction factor
β	Spectral floor factor
γ	Subtraction power
μ	Convergence parameter

List of abbreviations

Abbreviation	Explanation
ABNE	Average based noise estimation
ADC	Analog-to-digital converter
AMT	Auditory masking threshold
CPU	Central processing unit
DAC	Digital-to-analog converter
DMA	Direct memory access
DSK	Digital starter kit
DSP	Digital signal processor
EDMA	Enhanced direct memory access
EMIF	External memory interface
FFT	Fast Fourier transform
FIR	Finite impulse response
IFFT	Inverse fast Fourier transform
I/O	Input/output
LMS	Least mean square
MAC	Multiply-accumulate
MBNE	Median based noise estimation
McBSP	Multi-channel buffered serial port
MSBNE	Minimum statistics based noise estimation
QBNE	Quantile based noise estimation
RISC	Reduced instruction set computing
RTC	Real-time clock
SAMT	Supplementary auditory masking threshold
SNR	Signal-to-noise ratio
TI	Texas Instruments
USB	Universal serial bus

List of Figures

Chapter 1

INTRODUCTION

1.1 Problem overview

Laryngeal cancer sometimes necessitates the removal of larynx [1]. Having lost the natural voicing source, the patient needs a voicing aid for producing speech. The artificial larynx [2], [3] is a prosthesis meant for providing vibrations as a substitute to those provided by the natural larynx. Several types of artificial larynges are available, and the electronic artificial larynx or the electrolarynx is the most widely used type. It is a vibrator held against the neck tissue to produce vibrations required for the generation of speech. Pulses from its vibrating diaphragm, held against the throat, get transmitted through the neck tissue to the vocal tract. The resonances of the time-varying vocal tract filter dynamically shape the harmonic spectrum of the vibrations. The resulting speech is known as electrolaryngeal speech. The electrolaryngeal speech enables the laryngectomee patients to communicate verbally, but the speech suffers from the problem of background leakage noise, deficiency of low frequency content, and monotonic nature.

1.2 Project objective

Several signal processing techniques [4]-[10] have been reported for enhancement of electrolaryngeal speech. The objective of this project is to (i) investigate effect of different noise estimation and phase estimation techniques for use with spectral subtraction for suppressing the background noise in electrolaryngeal speech, (ii) develop a signal processing technique to compensate the low frequency deficit, and to reduce the monotonicity of the electrolaryngeal speech, and (iii) implement a real-time system for enhancement of electrolaryngeal speech which removes the background noise in the electrolaryngeal speech using a dynamic estimation of noise.

1.3 Dissertation outline

Chapter 2 reviews the literature related to enhancement of electrolaryngeal speech.Chapter 3 describes the effect of estimating the phase spectrum with different techniques in resynthesizing the clean speech. LPC based analysis-synthesis for introduction of jitter, shimmer, and spectral compensation is also described in this chapter. Real-time implementation of spectral subtraction algorithm using dsPIC33FJ128GP804 is described in Chapter 4. The subsequent chapter describes the real-time implementation of spectral subtraction algorithm using TMS320C5515 eZdsp USB stick. The last chapter gives the summary of the work done and suggestions for the future work.

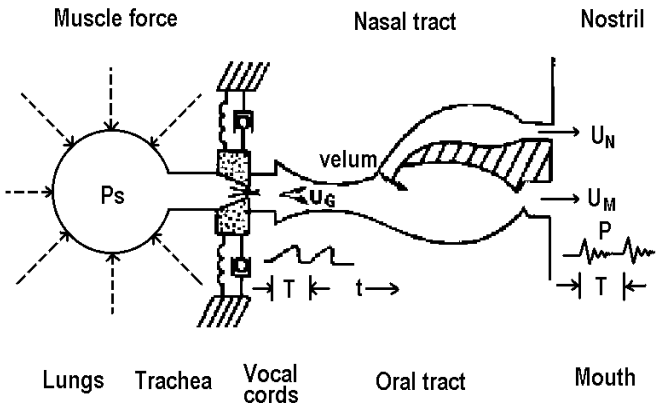
Chapter 2

ENHANCEMENT OF ELECTOLARYNGEAL SPEECH

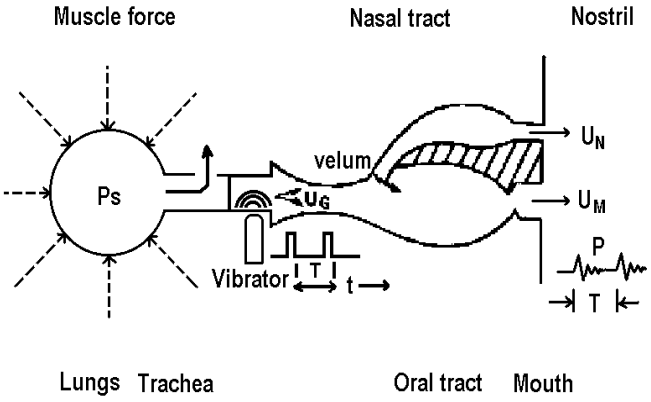
2.1 Introduction

It is sometimes necessary to surgically remove the larynx of a patient suffering from laryngeal cancer. After removal of the larynx there will not be any natural means to produce vibrations required for the generation of speech. Different artificial larynges have been developed [2], [3] and the external electronic artificial larynx, or the electrolarynx is the most widely used type. It is a vibrator held against the neck tissue to couple the vibrations to the air in the vocal tract, as a substitute to that provided by the vibration of vocal folds in the larynx. The speech produced using it is known as the electrolaryngeal speech. Schematics of normal speech production and that using this device are shown in Fig. 2.1 (a) and (b), respectively. The device generally permits setting of the vibration level and pitch by the user. However, a dynamic control of level, voicing, and pitch during speech production is very difficult. In addition to this basic limitation, the electrolaryngeal speech suffers from (i) presence of background noise caused by leakage of acoustic energy from the vibrator and vibrator-tissue interface, (ii) low-frequency spectral deficiency due to attenuation of the lower harmonics in transmission through the neck tissue, and (iii) unnatural quality due to constant pitch and level. Background noise decreases the intelligibility, while the other two factors affect the speech quality [1], [12], and [13]. Weiss *et al.* [1] reported that electrolaryngeal speech has a stronger concentration of energy between 400 and 800 Hz, resulting in a confusion in the identification of vowels due to auditory masking of the vowel formants.

Norton and Bernstein [14] reported that application of a foam shield around the device resulted in a reduction in the background noise. Later Espy-Wilson *et. al.* [4] reported that acoustic shielding of the vibrator assembly could reduce the leakage of the acoustic energy from the vibrator, but the shielding effect of the insulation was counterbalanced by mechanical damping and it was not effective in reducing the leakage from the vibrator-tissue interface. Several signal processing techniques for enhancement



(a)



(b)

Fig. 2.1 Schematic of (a) normal speech production [11], (b) speech production with an electronic artificial larynx [6].

of electrolaryngeal speech have been reported and some of these techniques are reviewed in the following section.

2.2 Adaptive filtering method

Espy-Wilson *et. al.* [4] reported a two-input noise cancellation method using adaptive filtering for the reduction of background noise. This method is based on the assumption that, in addition to the noisy signal, a reference correlated with the noise and uncorrelated with the desired signal is available. In the implementation, electrolaryngeal speech from near the lips $x(n)$, and background noise from near the electrolarynx $r(n)$ were recorded simultaneously. The signal $r(n)$ is passed through a filter and its output $y(n)$ is subtracted from $x(n)$ resulting in an error e(n). The filter coefficients are updated at each sample

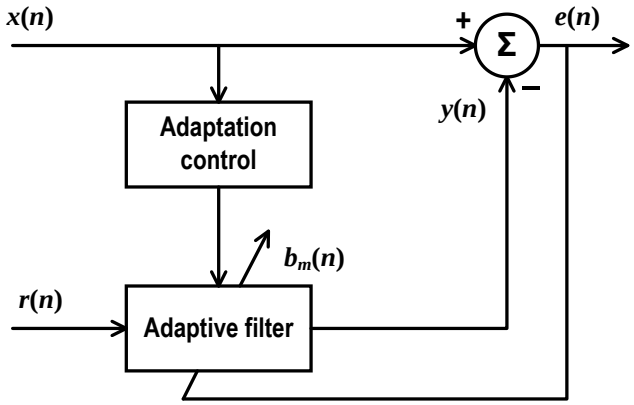


Fig. 2.2 Block diagram of the adaptive filter used by Epsy-Wilson *et. al.* [4].

based on least mean square [LMS] algorithm to reduce e(*n*). Figure 2.2 shows the block diagram of the implementation. The output of the filter is given by

$$y(n) = \sum_{m=0}^{N-1} b_m(n) r(n-m) \tag{2.1}$$

where b_m's are the coefficients of the filter. The error is given as

$$e(n) = x(n) - y(n) \tag{2.2}$$

The coefficients of the filters are updated as

$$b_m(n) = b_m(n-1) + \mu e(n) r(n-m),\ m = 0 \ldots N\text{-}1 \tag{2.3}$$

where μ is the convergence parameter and N is the filter length.

Correlation between $x(n)$ and $r(n)$ varies and when they are highly correlated $y(n)$ approximates to $x(n)$ resulting in nearly no signal. So an adaptation control was used based on the average energy of the current window. If the energy was greater than an empirically determined threshold, adaptation was suspended and a static filter with the latest adapted coefficients was used for filtering else adaptation was continued. It has been reported that a marked reduction in background noise was observed in low-energy intervals. The quality of the output was improved in the high-energy intervals but the background noise was not removed fully. The intelligibility of the input speech was not affected by the processing.

In electrolaryngeal speech, the speech signal and the background noise originate from the pulsatile vibrations of the diaphragm and hence they are strongly correlated. So adaptive filtering method is ineffective for the reduction of background noise in electrolaryngeal speech. A single input noise cancellation method based on pitch-synchronous application of generalized spectral subtraction [15], [16] for the suppression

of background noise in electrolaryngeal speech has been proposed in [5], [17]. The next section describes this method.

2.3 Spectral subtraction

In the spectral subtraction for enhancement of noisy speech, an estimate of the spectrum of the noise is subtracted from that of the noisy speech and the resulting magnitude spectrum is combined with the phase spectrum of the noisy speech for resynthesizing the clean speech [15], [16]. The method is based on the assumption that the speech and the noise are uncorrelated. But electrlolaryngeal speech is highly correlated with the background noise. However, it has been shown in [5], [17] that if the spectra are calculated pitch-synchronously, the speech and noise become uncorrelated and spectral subtraction can be employed.

A model of the generation of the background leakage noise in electrolaryngeal speech is shown in Fig. 2.3. The impulse response of the vocal tract filter and the impulse response of the leakage path are represented as $h_v(n)$ and $h_l(n)$, respectively. The speech signal $s(n)$ and the leakage noise $l(n)$ are generated by convolution of the pulsatile excitation $e(n)$ with the respective impulse responses

$$s(n) = e(n) * h_v(n) \tag{2.4}$$

$$l(n) = e(n) * h_l(n) \tag{2.5}$$

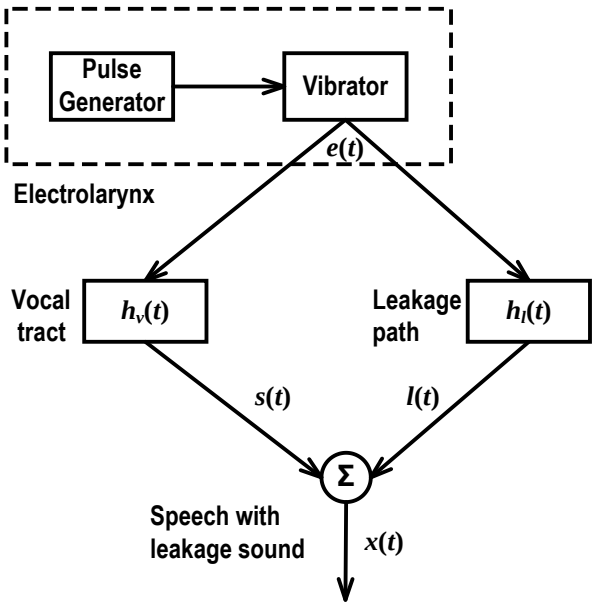


Fig. 2.3 A model of the background leakage noise generation in electrolaryngeal speech [5].

The noisy speech signal is given as

$$x(n) = s(n) + l(n) \tag{2.6}$$

The vocal tract acts as a time-varying filter during speech production, while the filter response of the leakage path varies slowly due to changes in the orientation and pressure in holding the vibrator against the neck tissue. Applying short-time Fourier transform on (2.6), we get

$$X_n(e^{j\omega}) = E_n(e^{j\omega})\,[H_{vn}(e^{j\omega}) + H_{ln}(e^{j\omega})] \tag{2.7}$$

The impulse responses of the vocal tract filter and the leakage path may be assumed to be uncorrelated, and hence

$$|X_n(e^{j\omega})|^2 = |E_n(e^{j\omega})|^2\,[|H_{vn}(e^{j\omega})|^2 + |H_{ln}(e^{j\omega})|^2] \tag{2.8}$$

If a pitch-synchronous window is used to evaluate short-time spectra, $|E_n(e^{j\omega})|^2$ may be considered as constant $|E(e^{j\omega})|^2$. During the non-speech intervals, $s(n)$ will be negligible and the noise spectrum is given as

$$|L_n(e^{j\omega})|^2 = |E(e^{j\omega})|^2\,|H_{ln}(e^{j\omega})|^2 \tag{2.9}$$

The noise spectrum can be estimated from the noise during explicit silences (lips closed), or dynamically using a voice activity detector, or using statistical techniques without a voice activity detector. A block diagram of the spectral subtraction technique is shown in Fig. 2.4. All the spectral estimates are computed using FFT. In the generalized spectral subtraction technique [16], the cleaned magnitude spectrum is obtained as

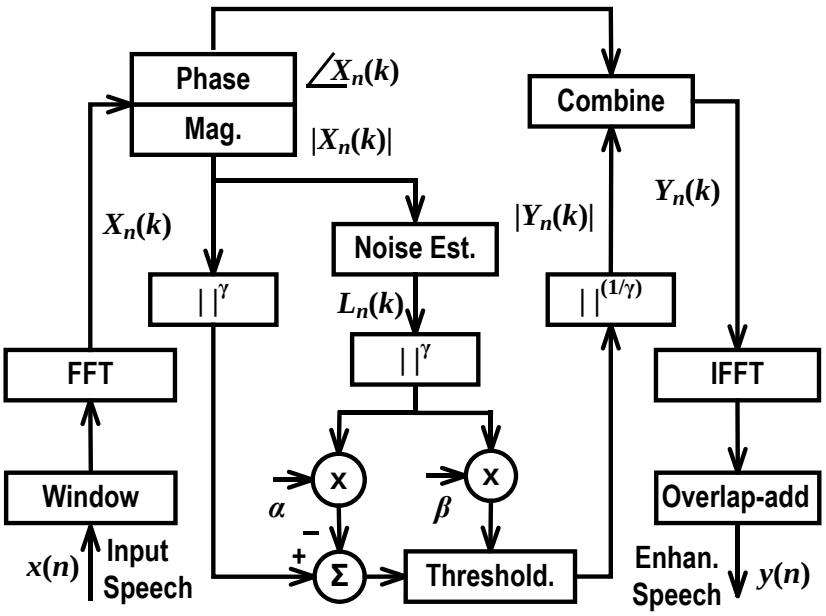


Fig. 2.4 Block diagram of generalized spectral subtraction [18].

$$E(k) = |X_n(k)|^{\gamma} - \alpha|L_n(k)|^{\gamma} \quad (2.10)$$

$$|Y_n(k)| = [E(k)]^{(1/\gamma)}, \text{ if } E(k) > [\beta|L_n(k)|]^{\gamma}$$

$$\beta|L_n(k)|, \text{ otherwise} \quad (2.11)$$

where α is an oversubtraction factor used to reduce the residual noise due to short-time variations in the noise. Oversubtraction may result in negative values in the spectrum, causing time-varying tonal sounds, known as "musical noise", which adversely affect the quality of the resynthesized speech. This noise is masked by a floor noise, controlled by the floor factor β. The subtraction power $\gamma = 2$ results in power subtraction and $\gamma = 1$ results in magnitude subtraction. The values of the three parameters need to be empirically obtained for each type of noise estimation and the device. The magnitude spectra after spectral subtraction are combined with the corresponding phase spectra of the noisy speech and the resulting complex spectra are used to resynthesize speech by using overlap-add method.

2.4 Estimation of noise spectrum for spectral subtraction

The characteristics of the background noise due to leakage of acoustic energy from the vibrator are different from those of the noise due to external sources. Its spectrum slowly varies due to the changes in the orientation of the electrolarynx against the neck tissue and the hand pressure in holding it during speech production. Its level and spectral characteristics are very similar to those of the speech, and hence voice activity detector is very difficult. Here some of the methods for the estimation of noise spectrum in electrolaryngeal speech are reviewed.

Estimation of the noise spectrum using averaging based noise estimation (ABNE) method was reported by Pandey *et al.* [5]. In this method, the electrolaryngeal speech with the speaker's lips closed for about 2 s was used to estimate the noise spectrum. Square magnitudes of the spectra of all the adjacent windowed frames in this non speech interval were averaged to get the estimated noise spectrum. Pratapwar [19] reported that the position of the window with respect to the excitation pulse did not affect the quality of the output speech in the pitch-synchronous application of spectral subtraction. As the background noise is dynamically varying, estimated noise spectrum has to be updated dynamically. Hence use of the noise spectrum estimated with the lips closed cannot be used for a continuous use of the device.

A statistical estimate of the noise without an explicit voice/silence detector has been reported to be effective in estimating the dynamically varying noise for noise reduction for speech recognition application[20]. Pandey *et al.* [6] used quantile-based noise estimation (QBNE) reported in [20] for enhancement of electrolaryngeal speech with spectral subtraction. In this method, the quantile values for different spectral components were selected by matching the noise spectrum estimated over a long speech record to match that obtained by averaging during the initial silence. Pratapwar [19] investigated different methods for selection of a particular quantile value to estimate the noise. These methods included those using a single quantile value, two-band quantile values, frequency dependent quantile values, and signal-level based dynamic selection of quantile values. It was reported that the spectral subtraction based on fixed quantile values was less effective during weak and non-speech segments compared to that using signal-level based dynamic selection of quantile values. Because of processing time and memory requirement, a real-time implementation of this method is difficult.

Another method for dynamically estimating the noise without speech-non speech discrimination was reported by Mitra and Pandey [9] and Kabir *et al.* [10] using the minimum statistics based noise estimation (MSBNE) based on the noise estimation and speech enhancement technique reported earlier by Martin [21], [22]. In this method, minimum of the magnitudes of each spectral sample in a set of past frames was considered to be the spectral sample of the estimated noise. Parameters for spectral subtraction were dynamically estimated in [10] based on the algorithm reported in [21], whereas fixed subtraction parameters were used in [9]. MSBNE is computationally less expensive and is suitable for real-time implementation if the subtraction parameters do not have to be dynamically estimated from the signal statistics.

2.5 Spectral subtraction based on auditory masking

Liu *et al.* [7] reported spectral subtraction with adaptation of parameters using frequency domain masking properties of the auditory system for suppression of the leakage noise as well as the noise from external sources. This method considers the frequency domain masking that the weak signal is inaudible if both strong and weak signals occur simultaneously. Figure 2.5 shows the block diagram of the implementation. The noise was estimated from the noisy speech using minimum statistic based recursively smoothed

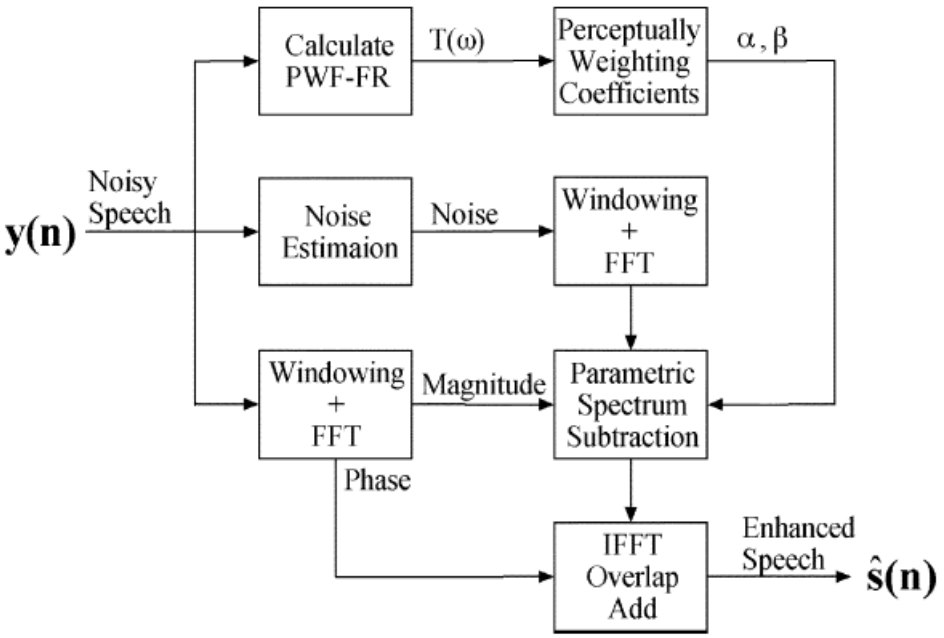


Fig. 2.5 Block diagram of spectral subtraction based on auditory masking reported in [7].

spectrum. A perceptual weighting filter frequency response (PWF-FR) which masks the noise in the formant regions was calculated from the noisy speech. The spectral subtraction parameters were updated based on the PWF-FR. If the PWF-FR was low, which is the case near the formant frequencies, subtraction parameters were increased. This results in masking the musical noise generated from over subtraction by the formant frequencies. If the PWR-FR was high, which is the case near the valleys, subtraction parameters were decreased. This avoids over subtraction, else musical noise will be introduced in the output speech as it cannot be masked near valleys. It was reported that this algorithm and power spectral subtraction (PSS) algorithm reduced the background noise in the electrolaryngeal speech effectively. But if PSS algorithm was used when white Gaussian noise and speech babble noise were added to the electrolaryngeal speech, noise reduction was poor in high frequency regions. Thus it may be concluded that spectral subtraction with auditory masking is more effective to enhance the electrolaryngeal speech with additive noise added to it, while PSS method is suitable in the absence of significant external noise.

Liu *et al.* [8] reported a supplementary auditory masking threshold (SAMT) algorithm to eliminate both additive noise and background noise in electrolaryngeal speech. In this algorithm an auditory masking threshold (AMT) is calculated using

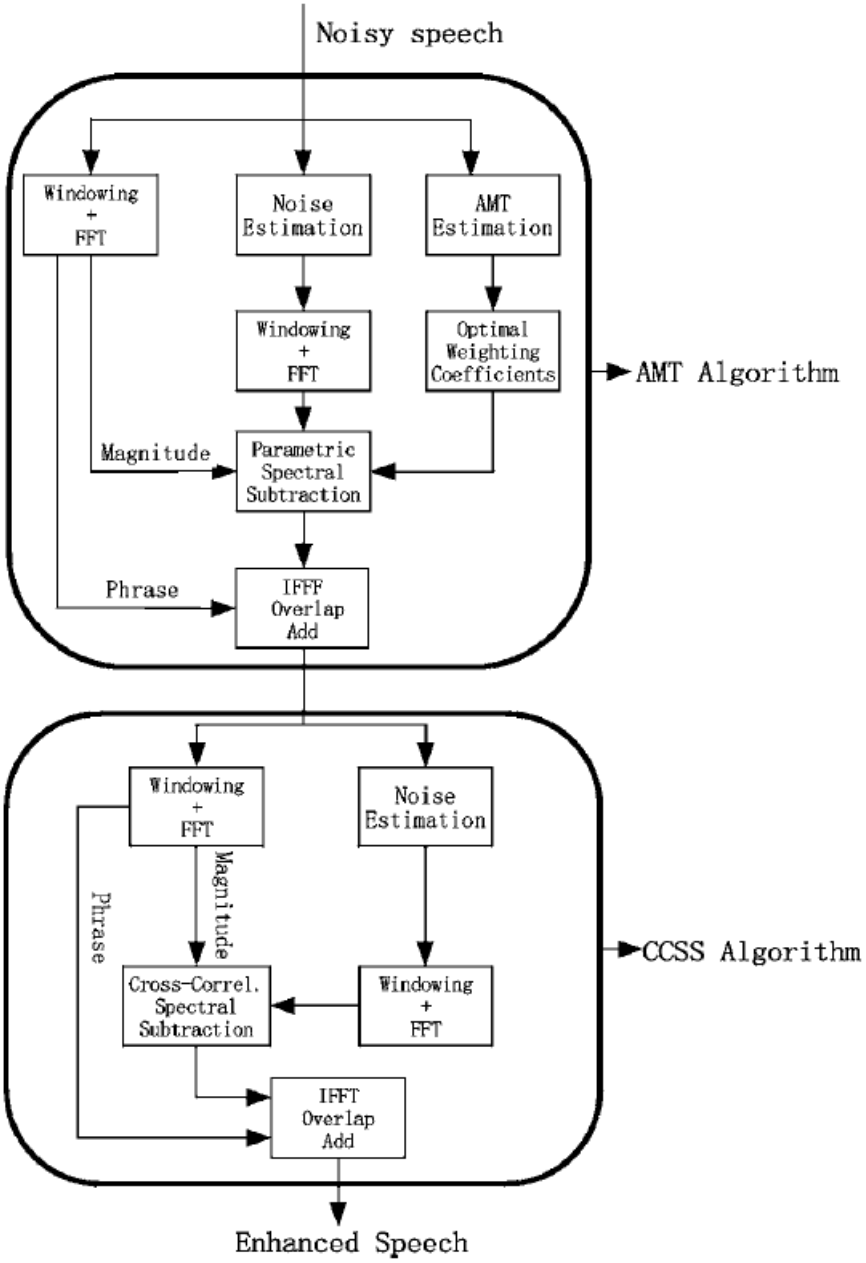


Fig. 2.6 Block diagram of SAMT algorithm reported in [8].

auditory masking model. Any of the noise components below AMT will not be perceptible for the human ear and there is no need to suppress them. So it will be sufficient to minimize the audible noise spectrum. The SAMT algorithm, shown in Fig. 2.6, has two stages. In the first stage estimated noise was spectrally subtracted from the noisy speech using AMT algorithm. In the second stage cross-correlation spectral subtraction (CCSS) was employed to reduce the correlated noise present in the enhanced speech from the first stage. In the AMT algorithm, noise was estimated using minimum statistic based recursively smoothed spectrum and subtraction parameters were updated

by the AMT. If the AMT was low, spectral parameters were increased to reduce the noise. If there was any musical noise introduced in the output speech due to the over subtraction, it would be masked by the background noise present in the electrolaryngeal speech due to high spectral floor. If the AMT was high, spectral subtraction parameters were decreased to their minimum values so that the residual noise would be below the AMT and it would be masked naturally. The authors reported that the perceptual tests showed effective reduction of background noise in electrolaryngeal speech with and without additive noise using both AMT and SAMT algorithms. Compared to these two algorithms, the PSS algorithm was not effective in reducing the high frequency noise in the case of electrolaryngeal speech without the additive noise. The two algorithms were better than PSS algorithm in case of electrolaryngeal speech with additive noise. The CCSS algorithm compensated the deficit of the AMT algorithm in the reduction of the low frequency noise. So SAMT algorithm can be preferred in the case where additive noise gets added to the electrolaryngeal speech. The choice of using the AMT algorithm instead of PSS or vice versa depends on the tradeoff between the quality of the output speech and the complexity of calculations.

2.6 Real-time implementation of spectral subtraction

Budiredla [23] has reported the real-time implementation of spectral subtraction algorithm using a DSP board based on the digital signal processor TMS320C6211 from Texas Instruments (TI). For real-time implementation, Code composer studio, the integrated development environment (IDE) for TI DSP processors, was used for configuring, building, interfacing and debugging purposes. The important features of the processor on the board include two multi-channel buffered serial ports (McBSPs), enhanced direct memory access (EDMA) controller which transfers the data between regions in the memory map without CPU intervention. The board has a codec AD535. The inbuilt ADC and DAC in the codec operate at a fixed sampling rate of 8 kSa/s. Spectral subtraction was implemented in real-time with over subtraction factor α = 2, spectral floor factor β = 0.001, and exponent factor γ = 1. The background noise was estimated using QBNE based on SNR. The resynthesizing was carried out by setting the phase spectrum as zero. It was reported that for effective noise estimation, 55 or more past frames were needed but because of the processing speed constraints, the number of

frames considered for noise estimation in real-time was 8. The output speech was reported to be poor in quality compared to the Matlab and C offline implementations.

Mitra [24] implemented spectral subtraction with ABNE and MSNBE (described in section 2.4) algorithms and spectral compensation for low frequency deficit in real-time using the DSP evaluation board EZ-Kit Lite from Analog Devices. The board is based on ADSP-BF 533 Blackfin processor. Some of the main features of the Blackfin processor core architecture are: dual MAC signal processing engine, an orthogonal RISC-like microprocessor instruction set, flexible single instruction multiple data (SIMD) capabilities. The kit has a codec AD1836A with three stereo DACs and two stereo ADCs using multibit sigma-delta architecture. The codec operates at a fixed sampling rate of 48 kHz or 96 kHz. So software decimation by a factor of 4 was done to bring down the sampling rate to 12 kHz. Spectral subtraction algorithm with ABNE and MSBNE was implemented in real-time. The resynthesis was carried out by using the original noisy phase spectrum. The spectral parameters used were, α = 10, β = 0.001, and γ = 1 in the case of implementation with MSBNE, and α = 2, β = 0.001, and γ = 1 in the case of implementation with ABNE . In MSBNE due to memory constraints, noise was estimated using the approach of cascading two minima estimations. Minima of every 10 successive windows were calculated and stored in a buffer of length 10. When this buffer gets filled with 10 such minima, the minimum of these 10 local minima was calculated which gives an effective minimum of 100 successive windows. The minimum of the second buffer was considered as the noise estimate. Results of real-time implementation of spectral subtraction with ABNE by keeping a 2 s no speech interval were reported to be comparable with that of Matlab offline implementation. But the estimation of noise using ABNE itself is not useful for long duration speeches as the noise is not getting updated. Mitra [24] reported that output speech of real-time implementation using MSBNE showed effective noise reduction, though not as good as offline processing, possibly because of the effects of finite precision arithmetic.

Chapter 3

INVESTIGATIONS USING OFFLINE IMPLEMENTATION

3.1. Introduction

Spectral subtraction algorithm (described in Sec. 2.3) can be used to reduce the background noise in the electrolaryngeal speech. In the algorithm, estimated magnitude spectrum of the noise is subtracted from the magnitude spectrum of noisy speech. The resultant magnitude spectrum is combined with the retained noisy phase to get the resynthesized clean speech. In this chapter, three investigations for enhancing electrolaryngeal speech are presented: (i) spectral subtraction with different noise estimation techniques reported in the literature (described in Sec. 2.4), (ii) effect of estimating the phase spectrum using different techniques, (iii) a LPC based analysis-synthesis method for introduction of jitter and shimmer in the electrolayngeal speech after spectral subtraction for suppressing its monotonic nature. These investigations are carried out using Matlab based offline implementation of the signal processing on recorded electrolaryngeal speech.

3.2. Spectral subtraction with different noise estimation techniques

Spectral subtraction was implemented using (i) average based noise estimation (ABNE) proposed by Pandey *et al.* [5], (ii) minimum statistics based noise estimation (MSBNE) with fixed subtraction parameters reported by Mitra and Pandey [9], and (iii) median based noise estimation (MBNE) (QBNE reported by Pandey *et al.* [6], with 0.5 quantile). Spectral subtraction is implemented as shown in Fig. 2.4, with the original noisy phase spectrum used for resynthesis. Investigations are carried out with several electrolaryngeal speech sentences uttered by three normal speakers. Recordings are done at a sampling rate of 11.025 kHz with a initial 2 s silence period, i.e. speaker closed his lips for initial 2 s while the device is on and coupled to the neck. So initial 2 s of the speech contains only background leakage noise. Rectangular window is used for the windowing and each windowed frame is of length two pitch periods and has 50 % overlap with the previous frame. Processing is carried out using 256-point FFT. The parameters for spectral

subtraction are selected empirically after observing the output for background noise reduction for the different set of parameters. Subtraction parameters resulting in the speech output judged to be of the best quality for the recorded electrolaryngeal speech for the sentence (S1), "*…Where were you a year ago? 1 2 3 4 5 6 7 8 9 10*", are given below.

In the implementation of ABNE, magnitude spectra of windowed frames of the initial 2 s silence period are stored row-wise in a two dimensional array. Average of each column is calculated and stored in another array, and it is considered as the estimated noise. In the output speech, after spectral subtraction with ABNE, a marked reduction in background noise was observed for shorter sentences less than 5 s length. But the noise reduction was not effective if the coupling of electrolarynx with neck was varied during the speech, even in shorter sentences. When spectral subtraction was implemented on S1, best output resulted for α = 10, β = 0.001, and γ = 1. Higher value of over subtraction factor, α, was required if input speech amplitude was higher. Effect of changing the length of overlapping samples in submultiples of frame length was investigated. It was found that quality of the output speech improved with higher overlapping.

Block diagram of spectral subtraction by MSBNE and MBNE is shown in Fig. 3.1. Magnitude spectra of a set of immediate past frames are stored row-wise in a two-dimensional array and they are used for estimating the noise. In MSBNE, the minimum of the magnitudes in each column of the array is stored in another array and it is considered as the estimated noise. Magnitude spectrum of each new windowed frame replaces the oldest row in the two-dimensional array, and the new estimate of the noise is calculated. It was found that a set of past frames corresponding to a length of approximately 3 s was required for better estimation of the noise. Output speech after implementation of spectral subtraction, was distorted in some segments of speech without frequent pauses. Subtraction parameters resulting in best output for Speech sentence S1 were α = 25, β = 0.005, and γ = 1. It was observed that for the input speech with higher amplitude, higher value of over subtraction parameter was required for better noise reduction.

In the implementation of MBNE, magnitude spectra of a set of past frames are stored row-wise in a two-dimensional array. Median of the magnitude values in each column after sorting in ascending order is stored in another array and it is considered as the estimated noise. There was an effective background noise reduction in the output speech after the implementation of spectral subtraction. If the coupling of electrolaynx with neck was varied during the speech, some background noise was present for a

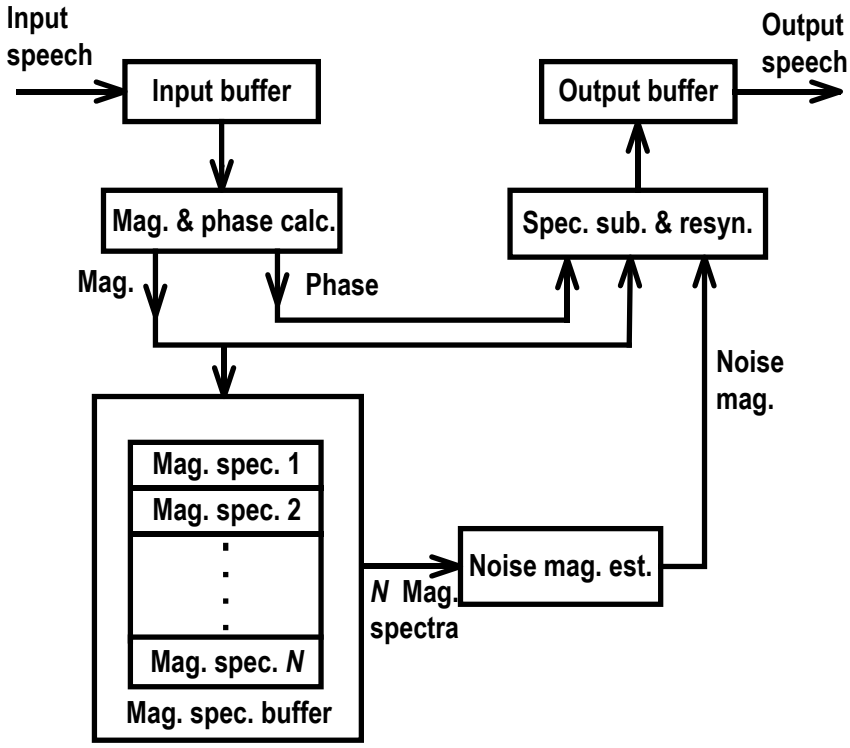


Fig. 3.1 Block diagram of spectral subtraction using MSBNE and MBNE.

duration corresponding to the number of past frames used to estimate the noise. It was found that better estimate of the noise was obtained with a set of frames corresponding to a speech segment of length 0.5 s. Subtraction parameters that resulted in better output for Speech sentence S1 were α = 1.5, β = 0.001, and γ = 1. It was found that for the input speech with higher amplitude, higher value of over subtraction parameter was required for better noise reduction.

On the basis of a comparison of the results obtained using MSBNE, and MBNE, it may be concluded that MBNE is better suited for continuous enhancement of electrolaryngeal speech as it requires a smaller number of frames for estimation of the noise. Further it needs a smaller over subtraction factor, resulting in a smaller possibility of subtraction of speech itself and hence a smaller possibility of distortion. Hence it was decided to implement the spectral subtraction with MBNE in real-time.

Memory of the selected processor (dsPIC33FJ128GP804 from Microchip [25]) was not sufficient to store the set of past frames corresponding to a speech segment of length 1.5 s. Hence a 3-point 4-stage cascaded median approach was used to estimate the noise. A block diagram of 3-point *m*-stage cascaded median approach is shown in

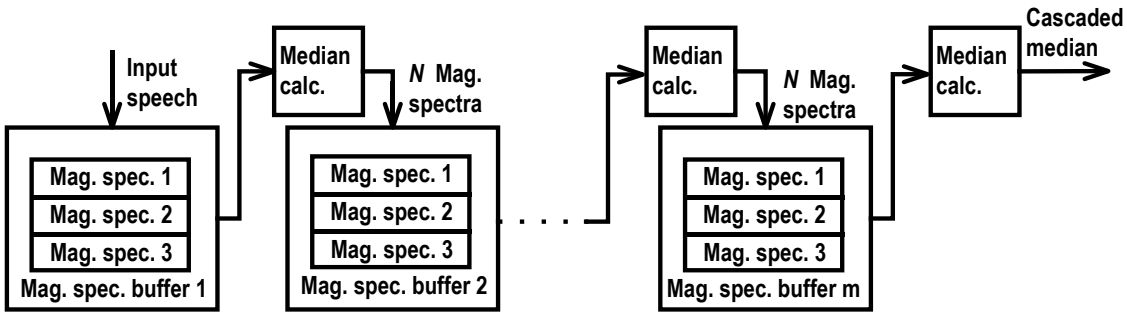


Fig. 3.2 Block diagram of 3-point m-stage cascaded median approach to find cascaded median of 3^m input frames.

Fig. 3.2. Each stage has two-dimensional integer arrays of size 3 x 128. Arrays in the first stage store the magnitude values of the FFTs of the three immediate past frames. After every three frames, an ensemble median is calculated and stored in the array of the next stage. Similar process is applied on each stage till the last stage. In order to limit the maximum computation time taken in each frame, at the most only one median is calculated every frame, giving priority to the higher stages. Therefore the calculation of the medians in a given stage may be missed at certain frames, but this is not likely to affect the estimated noise. With *m* stages, this approach gives an approximation of the median of 3^m past frames using a memory that can store the magnitude values of 3*m* frames. Spectral subtraction with 3-point 4-stage cascaded median approach using Matlab reduced noise effectively. The output was similar to that obtained using 3-point 5-stage approach. It was earlier found that MBNE was effective in noise reduction if the median was estimated over the past frames corresponding to approximately 0.5 s (corresponding to 150 frames). The output using the 3-point 4-stage cascaded median approach was similar to that obtained using 150-point median and much better than 12-point median. Hence it may be concluded that noise estimation using 3-point 4-stage cascaded median approach suits for real-time implementation of spectral subtraction with dynamic noise estimation.

3.3. Estimation of phase spectrum

In speech enhancement by spectral subtraction, the magnitude spectrum resulting from spectral subtraction is associated with the phase spectrum of the noisy speech and the resulting complex spectrum is used for resynthesis. Effect of associating phase spectra

obtained by different methods was investigated, with the objective of finding methods which can help in reducing the computation and improving the speech quality.

To assess the effect of the phase spectrum in the electrolaryngeal speech, speech was also resynthesized using (i) noisy phase, (ii) zero phase, (iii) randomly selected phase, (iv) continuous phase, and (v) minimum phase. For the third method, the uniformly distributed random phase in the interval (0, 2π) is considered as the estimated phase. For the fourth method, phase spectrum is estimated by assuming continuity of phase across frames. Noisy phase is taken as the initial phase and the phase is calculated as

$$\theta_n(k) = \theta_{n-1}(k) + (2\pi n_d k)/N \tag{3.1}$$

where n_d = window shift, N = FFT size, k is frequency bin index. A minimum phase signal can be recovered from the magnitude of its Fourier transform [25]. The vocal tract can be modeled as a minimum phase system because of its passive nature [26]. Considering the processed speech as minimum phase signal, its phase spectrum may be calculated from the magnitude spectrum after spectral subtraction using iterative technique [27], [28] or cepstrum-based non-iterative technique [11], [25], [29]. We have explored using the non-iterative technique to estimate the phase spectrum given the magnitude spectrum.

Input speech segment $x(n)$ of length equal to two-pitch periods ($2M$ samples) is processed by spectral subtraction using 256-point FFT. We get the enhanced magnitude spectrum $|Y(k)|$. Its cepstrum is calculated as

$$c(n) = \text{IFFT}\,[\log |Y(k)|] \tag{3.2}$$

For a minimum phase sequence, we can calculate the corresponding complex ceptrum as

$$\hat{y}(n) = \begin{cases} 2c(n), n > 0 \\ c(0), n = 0 \\ 0, n < 0 \end{cases} \tag{3.3}$$

From this complex cepstrum, we obtain the desired complex spectrum as

$$Y(k) = \exp(\text{FFT}[\hat{y}(n)]) \tag{3.4}$$

In the original method using noisy phase, we use $|Y(k)|\exp(j\angle X(k))$ as the enhanced spectrum. In the minimum-phase assumption based method, we use $Y(k)$ as obtained in (3.4) as the enhanced complex spectrum for resynthesis.

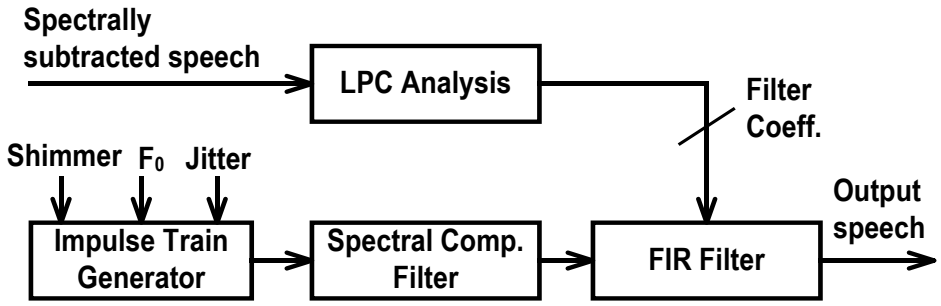


Fig. 3.3. Introduction of jitter, shimmer, and spectral compensation using LPC based analysis-synthesis [18].

3.4. Introduction of jitter, shimmer and spectral compensation

Random variations in the pitch and the level in speech are known as the jitter and the shimmer, respectively. Electrolaryngeal speech sounds monotonous and unnatural, as it has no jitter and shimmer. While a dynamic control of voicing, pitch, and level by the user of the device is very difficult, introduction of jitter and shimmer in the electro-laryngeal speech, either by introducing it in the vibrator itself or by processing of the signal after suppression of the background noise, may help in reducing its unnaturalness. For investigating the effect of jitter and shimmer in electrolaryngeal speech, a LPC based analysis-synthesis, as shown in Fig. 3.3, is used. The time-varying response of the vocal-tract filter is estimated by LPC analysis [11] and the coefficients of the prediction filter are used to realize a time-varying filter for resynthesizing the speech. The LPC analysis is carried out using 2-pitch period window and autocorrelation method for estimating 12 predictor coefficients. To closely track the vocal tract variation, 5-sample frame shifting is used. The time-varying resynthesis filter is excited by an impulse train with its frequency equal to that of the vibrator. Shimmer is introduced by varying the amplitude of the impulses as $a(1+sr_1)$, where a is the amplitude, r_1 is a random number uniformly distributed over ± 0.5, and s is the peak- to-peak shimmer. Jitter is introduced by varying the spacing of the successive impulses as $N(1+jr_2)$, where N is the pitch period in number of samples, r_2 is a random number uniformly distributed over ± 0.5, and j is the peak-to-peak jitter.

Electrolaryngeal speech is deficient in low frequency content due to a relatively higher attenuation of low frequency components during the transmission of the vibrations through the neck tissue. Use of an impulse train as the excitation source in the LPC based

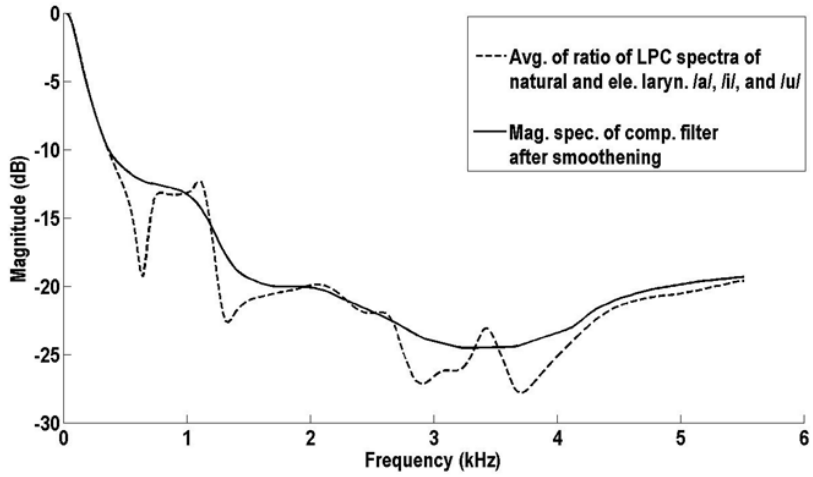
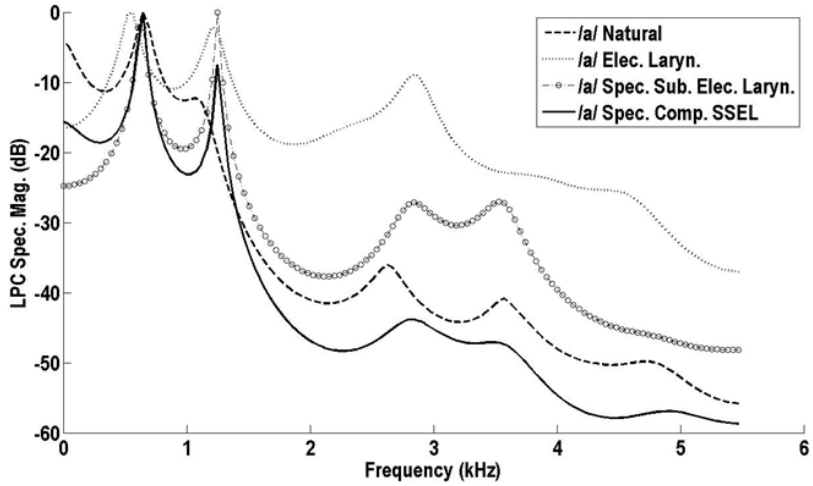
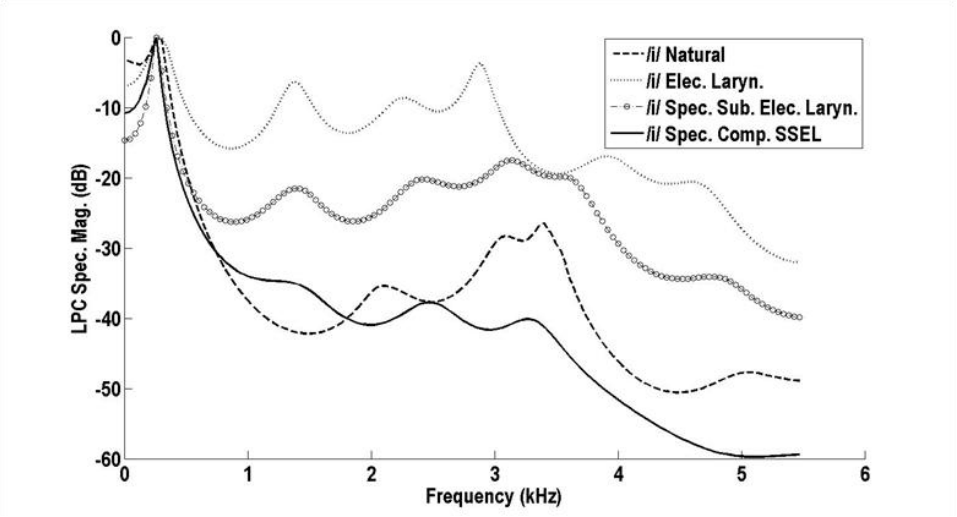


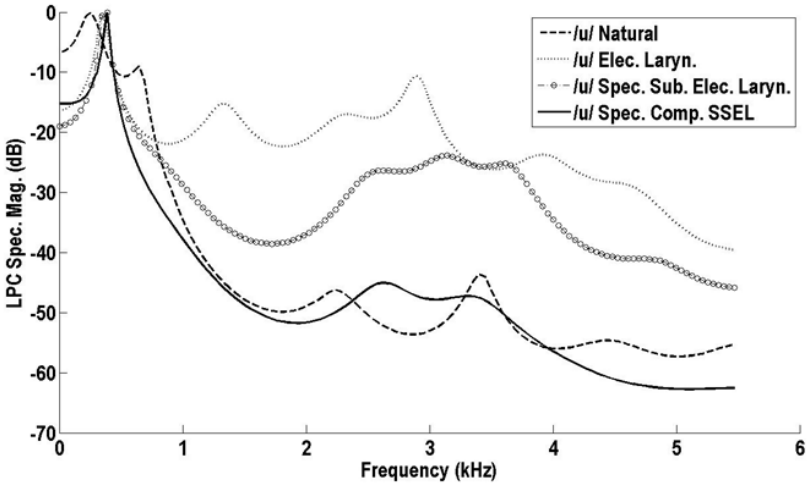
Fig. 3.4. Magnitude response of the designed compensation filter after smoothing the average of ratio of LPC spectra of natural and electrolaryngeal /a/, /i/, and /u/.



(a)



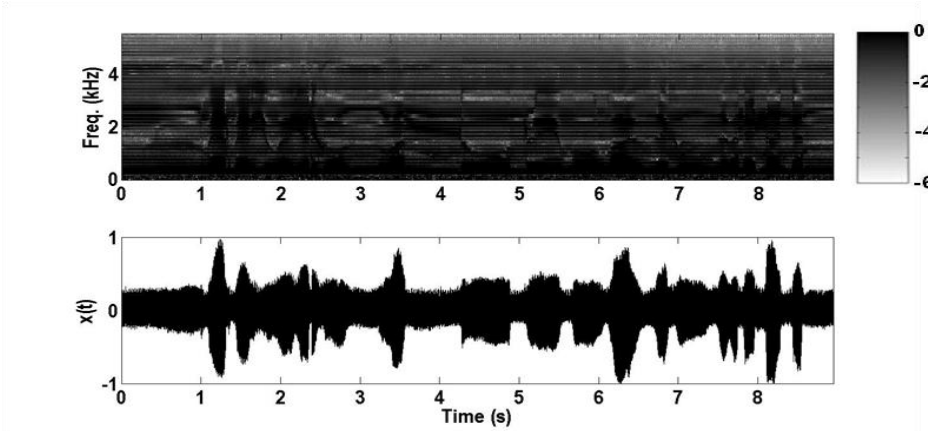
(b)



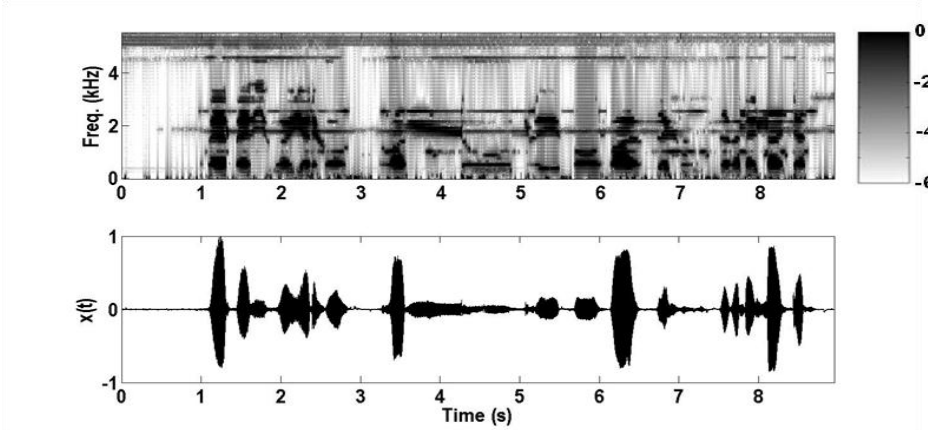
(c)

Fig. 3.5. LPC spectral magnitudes of natural, electrolaryngeal, spectrally subtracted electrolaryngeal (SSEL), spectral compensated SSEL (a) /a/, (b) /i/, and (c) /u/.

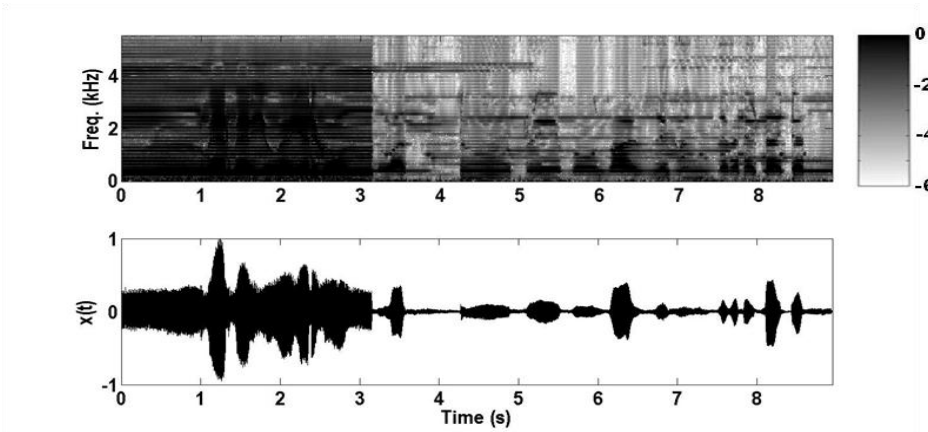
analysis-synthesis result in an emphasis of high frequency in the resynthesized speech. A spectral compensation filter is inserted in the excitation path to approximate the long-duration averaged spectrum of the resynthesized signal to that of the natural speech. Sustained vowels /a/, /i/, /u/ were recorded from a speaker speaking naturally and by using an electrolarynx. Ratio of the averaged LPC-smoothened spectra of the natural speech and the electrolaryngeal speech after spectral subtraction was used to obtain the magnitude spectrum of the compensation filter and the filter was designed as a linear-phase FIR filter. The magnitude response of the designed compensation filter is shown in Fig. 3.4. LPC spectra of natural, electrolaryngeal, spectrally subtracted electrolaryngeal (SSEL), spectral compensated SSEL /a/, /i/, /u/ are shown in Fig. 3.5. Effect of spectral compensation without using LPC based analysis-synthesis on the output speech after spectral subtraction is also investigated. Output speech after spectral subtraction with MBNE is passed through a linear phase FIR filter of magnitude response shown in Fig. 3.4. The resulting speech was better in quality compared to the output speech after spectral subtraction but not as good as the output speech obtained after LPC based analysis synthesis with 6 % jitter.



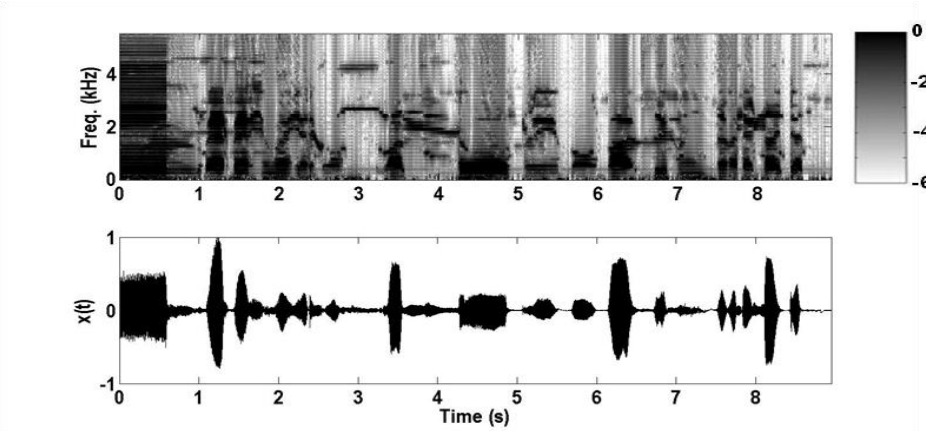
(a) Recorded speech waveform and its spectrogram.



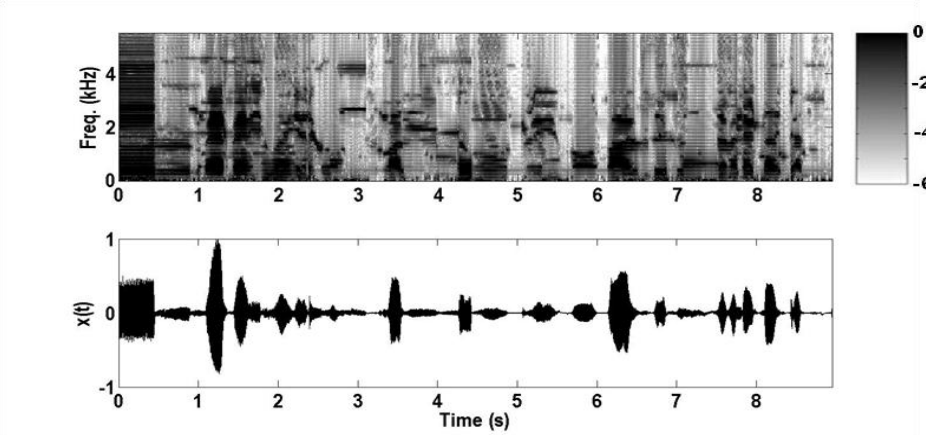
(b) Speech after spectral subtraction with α = 10, β = 0.001, and γ = 1 using ABNE.



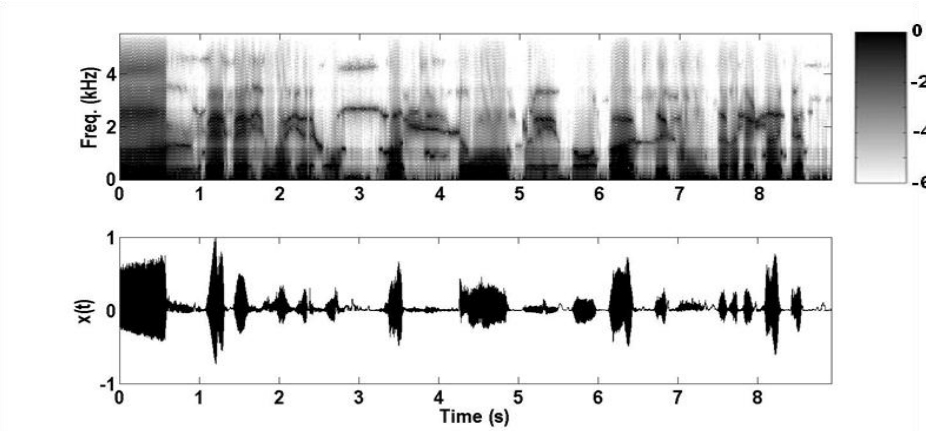
(c) Speech after spectral subtraction using MSBNE with 400 frames, α = 25, β = 0.005, and γ = 1.



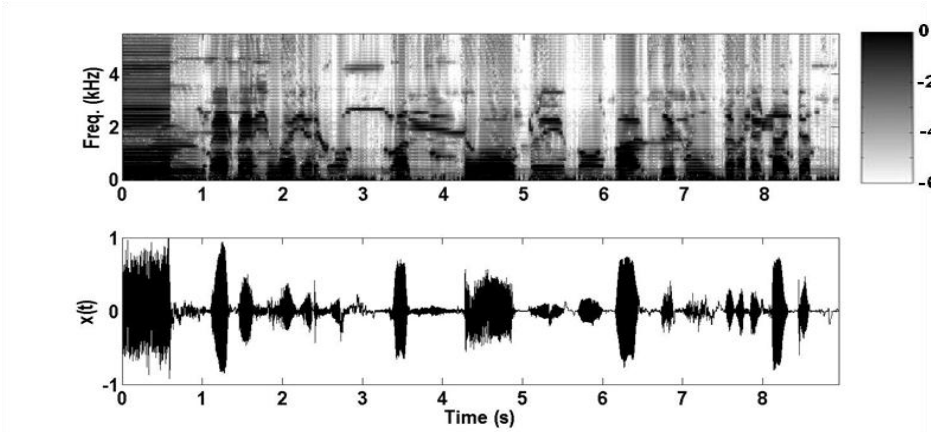
(d) Speech after spectral subtraction using MBNE with 150 frames, α = 1.5, β = 0.001, and γ = 1.



(e) Speech after spectral subtraction using 3-point 4-stage cascaded median approach, α = 1.3, β = 0.002, and γ = 1.



(f) Speech after spectral subtraction, spectral compensation, and introduction of jitter and shimmer using MBNE with 150 frames, α = 1.5, β = 0.001, γ = 1, j = 0.06, and s = 0.



(g) Speech after spectral subtraction using MBNE and spectral compensation without using LPC based analysis-synthesis.

Fig. 3.6 Recorded and output speech after spectral subtraction using ABNE, MSBNE, MBNE, 3-point 4-stage cascaded median approach, output speech after spectral subtraction, spectral compensation, and introduction of jitter and shimmer with MBNE, and output speech after spectral subtraction and spectral compensation. Speaker: PCP, material: "*…Where were you a year ago? 1 2 3 4 5 6 7 8 9 10*", generated using Solatone electrolarynx.

3.5. Results and discussion

Electrolaryngeal speech was recorded from two normal speakers, using electrolarynx models SolaTone (pitch frequency = 126.7 Hz) and NP-Voice (93.4 Hz), at a sampling rate of 11.025 kHz and 16-bit quantization. Spectral subtraction was performed using 2-pitch period frames with 50 % overlap. All the processing was carried out using Matlab. Effects of spectral subtraction, frequency compensation, and introduction of jitter and shimmer were assessed through informal listening tests.

The optimal values of the three factors in the generalized spectral subtraction were found to be dependent on the noise estimation method and input speech amplitude. It was found that use of power $\gamma = 1$ resulted in more tolerance to the variations in the values of the over-subtraction factor α and the floor factor β. For the noise estimated by averaging the noise during initial 2-s segment with lips closed, best results were obtained with $\alpha = 10$ and $\beta = 0.001$. However, the noise estimation was effective for spectral subtraction only up to about 5 s. With minimum statistics based noise estimation, best results were obtained for $\alpha = 25$ and $\beta = 0.005$. The method was found to need about 3 s of silence for

correctly estimating the noise. It was found that in the absence of frequent pauses in speech, the noise estimation was affected by speech segments and resulted in distortion of speech. Median based noise estimation was able to track the noise without requiring a long initial silence or frequent pauses in speech. Best results were obtained with α =1.5, β =0.001. It was observed that increase in the percentage of overlap of the windowed frames increased the output speech quality.

Estimation of noise using 3-point 4-stage cascaded median approach was investigated. In this method, an approximate median of magnitudes of 3^4, i.e 81, past frames was calculated using a memory that can store magnitude values of 12 frames. Best results were obtained with α = 1.3, β = 0.002, and γ = 1. The noise reduction in output speech was similar to that obtained with MBNE which requires a memory that can store magnitude values of 150 frames to estimate the noise. Estimation of noise using 3-point 5-stage cascaded median approach resulted in similar output obtained using 3-point 4-stage cascaded median approach.

The investigations on the effect of phase estimation methods were carried out on a 5 s segment of speech using ABNE. The speech quality for minimum-phase estimation was not better than that obtained by using the phase of the noisy speech. Use of zero phase, random phase, and phase estimated using phase continuity approach resulted in poor quality compared to that obtained with minimum-phase estimation.

An example of noise suppression with different types of noise estimation is shown using the waveforms and spectrograms in Fig. 3.6, for the original electrolaryngeal speech, the speech after spectral subtraction, resynthesis by LPC-based analysis-synthesis, and spectral compensation without using LPC-based analysis-synthesis. ABNE results in very good noise suppression immediately after the estimation but it degrades with lapse of time possibly due to changes in the noise characteristics. Output speech after spectral subtraction with MSBNE, MBNE, and 3-point 4-stage cascaded median approach, contains background noise until enough number of past frames have contributed to the estimation of noise. They were found to be effective in dynamically estimating the noise.

Use of compensation filter significantly improved the quality of the speech. Speech was resynthesized by introducing jitter and shimmer with the peak-to-peak values varied from 0 to 40 %. A peak-to-peak jitter of 6 % resulted in maximum improvement in naturalness, while the values above 20 % resulted in degradation of speech. Introduction

of shimmer up to 20 % did not result in an improvement in naturalness, while the larger values of shimmer degraded the speech. Output speech after spectral compensation without using LPC-based analysis-synthesis was found to be better in quality compared to the output speech after spectral subtraction but not as good as the output speech obtained after adding 6 % jitter using LPC-based analysis-synthesis.

Chapter 4

REAL-TIME SPECTRAL SUBTRACTION USING dsPIC33FJ128GP804

4.1. Introduction

For real-time implementation of spectral subtraction algorithm (described in Sec. 2.3), the main considerations in selecting the DSP processor are low power consumption, RAM space, direct memory access (DMA) controller for efficient block processing, processing speed for FFT and IFFT operations, analog interface modules. One of the available processors meeting most of these requirements is dsPIC33FJ128GP804 from Microchip [30]. This chapter presents the implementation of the real-time spectral subtraction algorithm using this processor.

4.2. Hardware

Microchip dsPIC33FJ128GP804 has a 16-bit fixed point digital signal processing unit, on-chip program memory of 128 KB and data memory of 16 KB including DMA memory of 2 KB, and on-chip analog I/O modules. A block diagram of the chip with the resources needed for spectral subtraction is shown in Fig. 4.1. The chip has several options of internal and external clock sources. To keep the component count low, the processor was used with its internal RC oscillator and PLL to operate at 40 MHz. On-chip successive approximation type ADC, having the input range of 0 - 3.3 V, can be configured in 4-channel 10-bit, or single channel 12-bit, modes. 10-bit ADC configuration has sampling frequency up to 1.1 MHz and it can sample 4 channels simultaneously. 12-bit ADC configuration has sampling frequency up to 500 kHz but simultaneous sampling of multi channels is not possible in this configuration. In the present application, ADC is configured in single channel 12-bit mode and the sampling rate is set to 10 kHz. The conversion results from the ADC are stored in a single-word result buffer ADC1BUF0. The chip has a 16-bit delta-sigma converter type DAC, having

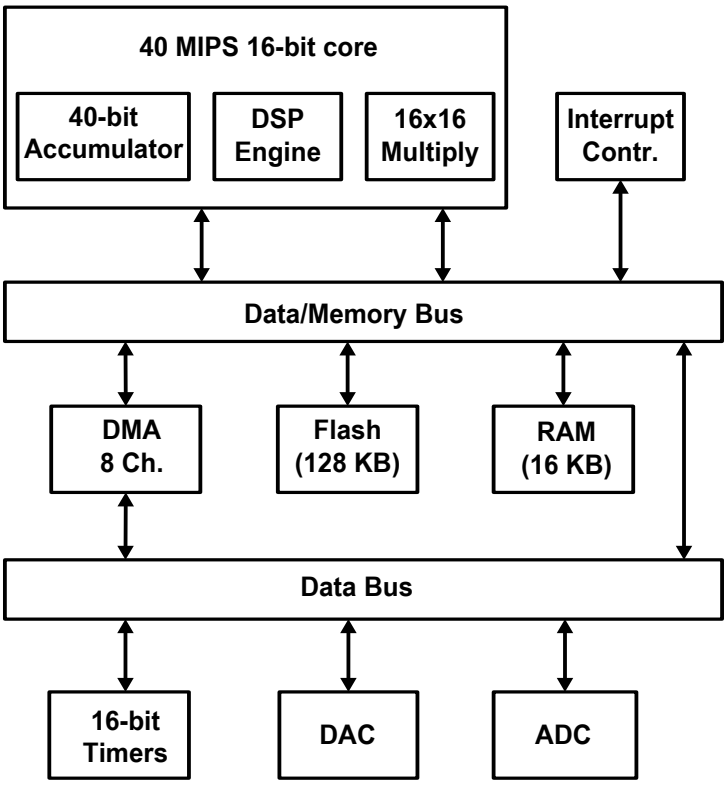


Fig. 4.1 Simplified schematic of the architecture of Microchip dsPIC33FJ128GP804, adapted from [30].

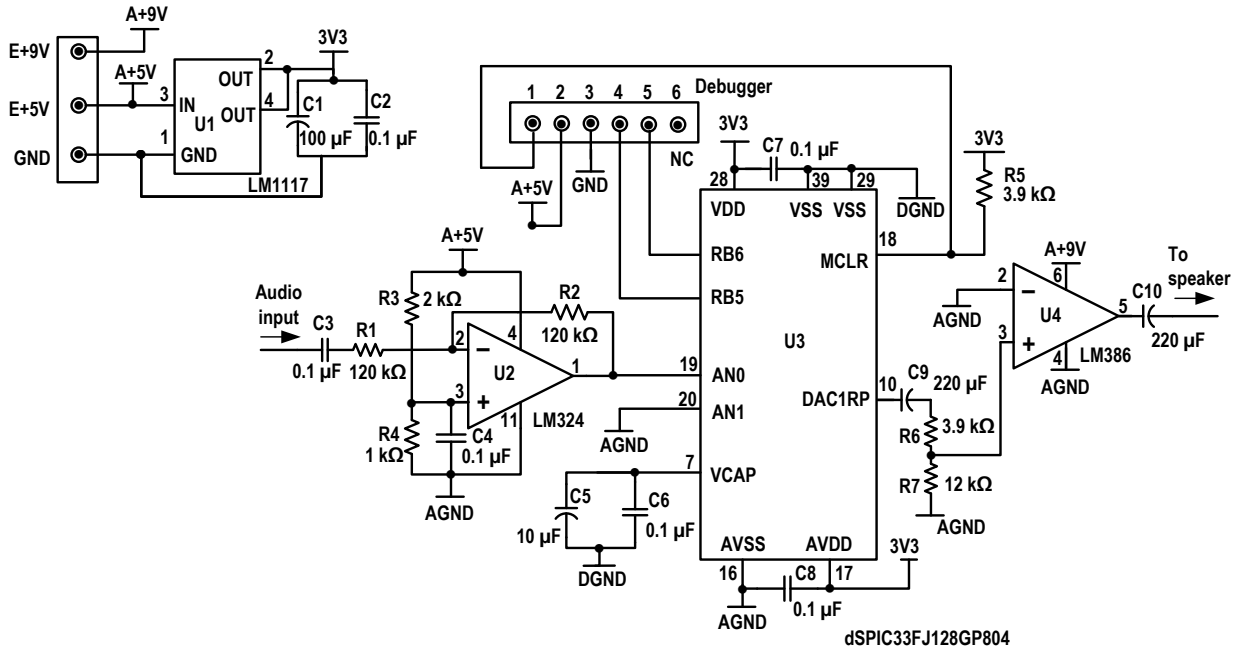


Fig. 4.2 Circuit diagram of the dSPIC33FJ128GP804 based hardware used for real-time spectral subtraction.

the output range of 1.125 − 2.235 V, with two output channels. Two four word FIFO arrays, DAC1LDAT and DAC1RDAT, buffer the data for the left and right channels respectively. The sampling rate of the DAC is set to 10 KHz. The maximum supported sampling rate by DAC is 100 kHz. The chip has nine 16-bit timers, and some of them can be cascaded to form 32 bit timers.

The circuit diagram of the hardware used for real-time implementation of spectral subtraction algorithm is shown in Fig. 4.2. The audio signal from the PC sound card, which is the audio input to ADC of the processor, U3 (dsPIC33FJ128GP804), has a range of -1 to 1 V. But the ADC input range is 0 - 3.3 V so a dc bias of 1.67 V is added without any ac gain to the input audio signal using the operational amplifier U2 (LM324, low power quad op amp). The processed data will be output by positive terminal of DAC right channel, DAC1RP. The DAC output is amplified by the power amplifier U4 (LM386, low voltage audio power amplifier) and the amplified output is given to the speaker. The power supply to the processor, 3.3 V, is provided from the output of the regulator U1 (LM1117). The figure also shows the 'Debugger' used to program the processor, U3.

4.3. Software

All the programs described below are written in C and loaded in the on-chip program memory of the processor with the help of development programmer/debugger Microchip PICkit 2. Student version compiler 'mplabc30-v3.25-comboLITE' is used for compiling the C programs.

Before implementing the spectral subtraction algorithm on electrolaryngeal speech in real-time the following programs are tested on the processor: (i) ADC-DAC loop back 'loop_back', (ii) FFT-IFFT of the input data 'FFT_IFFT', and (iii) implementation of spectral subtraction algorithm on a sinusoidal wave 'Spec_sub_sine'. All these programs configure the processor to operate at 40 MHz. In the first program 'loop_back', ADC samples the audio input and the sampled values are output to the DAC with specified delay and scaling. This program is used to verify the operation of ADC and DAC. In the second program 'FFT_IFFT', IFFT of the FFT of the input data is calculated and the values obtained are compared with the Matlab calculated values. FFT and IFFT are calculated using the predefined functions 'FFTComplex' and 'IFFTComplex' provided in

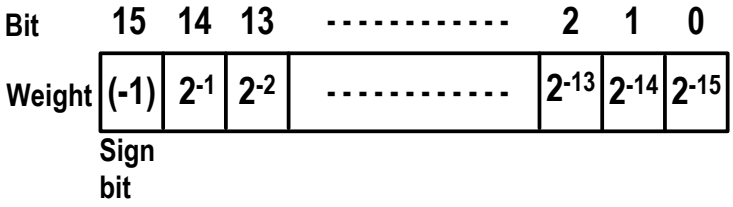


Fig. 4.3 Data representation of type ‘fractional’.

the header file ‘dsp.h’. These two functions operate out of place i.e. the result is stored in a predefined array which is different from the input array. Input to the function ‘FFTComplex’ must be an array of type ‘fract complex’ which is a structure having the members ‘real’ and ‘imag’ of type ‘fractional’. The ‘fractional’ data type represented in Fig. 4.3 is used to represent the data that has 1 sign bit, and 15 fractional bits. Data which uses this format is commonly referred to as “Q1.15” data. So 0.25 in ‘fractional’ format is represented as 0x2000 in hexadecimal format. Fractional arithmetic avoids the overflows in the FFT or IFFT calculations. Output of the function ‘FFTComplex’ is scaled by the length of the FFT. This program ensures the correct operation of FFT and IFFT functions. In the third program ‘Spec_sub_sine’, spectral subtraction algorithm with minimum statistics based noise estimation (MSBNE) is implemented on a sinusoidal wave. In MSBNE, the minimum of magnitudes of FFTs of a set of past frames is considered as an estimate of the noise. The maximum number of frames that could be stored in the on-chip data memory available is 15. The subtraction parameters used are $\alpha = 1$, $\beta = 0$, and $\gamma = 1$. The sinusoidal input is taken from a function generator and the output expected after few fractions of a second is zero as the estimated noise will be same as the signal itself. This program checks the correctness of the implementation. After testing these three programs successfully, real-time implementation of the spectral subtraction algorithm on electrolaryngeal speech was carried out.

Final implementation with a minimal number of components is shown in Fig. 4.4. The input electrolaryngeal speech from the preamplifier is sampled at 10 kHz using the on-chip ADC configured in the single channel 12-bit operating mode. The output data of the ADC is set to signed fractional format. The sample values are acquired and stored in the memory for block processing using DMA. Conversion by ADC triggers the DMA channel 0 to store the input samples from ADC1BUF0 to a circular memory formed by buffers A and B located in the DMA memory. Whenever one of the two buffers gets

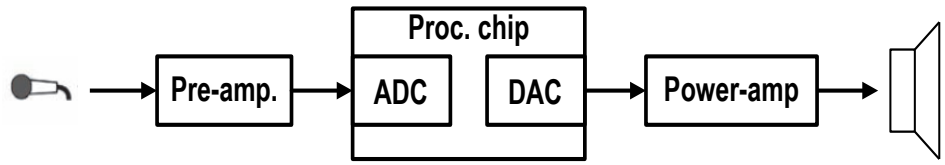


Fig. 4.4 Block diagram of the real-time speech enhancement system.

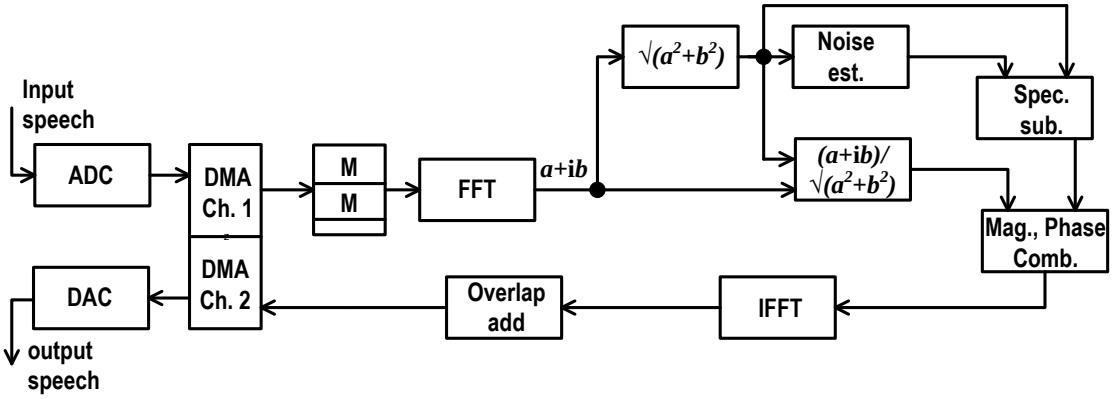


Fig. 4.5 Block diagram of the real-time implementation of spectral subtraction algorithm.

filled, DMA0 interrupt occurs and the values from the corresponding buffer are copied into the data memory for further processing. The processed values are output by using the buffers C and D located in DMA memory. DMA channel 1 is used to copy the data from copying of the values from one of the two buffers, an interrupt occurs and the new processed values are stored in the corresponding buffer. The implementation using program 'SS_EL_3p4s_CasMed' is shown in Fig. 4.5.

Analysis is carried out using a window length of two pitch periods. As the pitch period of the electrolaryngeal speech remains constant, pitch synchronous analysis is carried out by setting the length *M* of each of the two input buffers equal to one pitch period. Implementation is carried out with 256-point FFT, which permits pitch frequency higher than approximately 80 Hz. A separate array of 256 words in the memory serves as the FFT input buffer, which is initialized to zero values. It may be considered as three sub-arrays: first *M* words, next *M* words and the remaining words. After filling of either of the ADC input buffers, the values in the second sub-array are copied to the first sub-array and the values in the input buffer are copied to the second sub-array. For FFT of each frame, the samples corresponding to the analysis window get automatically zero padded.

The real and the imaginary parts of the complex spectral values are used to calculate the magnitude. Noise is estimated using median based estimation (QBNE with 0.5 quantile). To overcome the constraint on the number of past frames used for estimation of noise due to data memory size, a 3-point 4-stage cascaded median approach (described in Sec 3.2) is used.

The spectral subtraction is carried out with $\gamma = 1$ and settable values of α and β. The resulting magnitude is combined with the original phase, without explicit calculation, to get the complex spectral value as follows

$$[Y_n(k)]_{real} = |Y_n(k)| \, [X_n(k)]_{real} / |X_n(k)| \tag{4.1}$$

$$[Y_n(k)]_{imag} = |Y_n(k)| \, [X_n(k)]_{imag} / |X_n(k)| \tag{4.2}$$

Output speech is obtained by taking IFFT of Y_n and applying overlap-add on the resulting sequence, as the following

$$y_n(m) = \text{IFFT}\,(Y_n(k)) \tag{4.3}$$

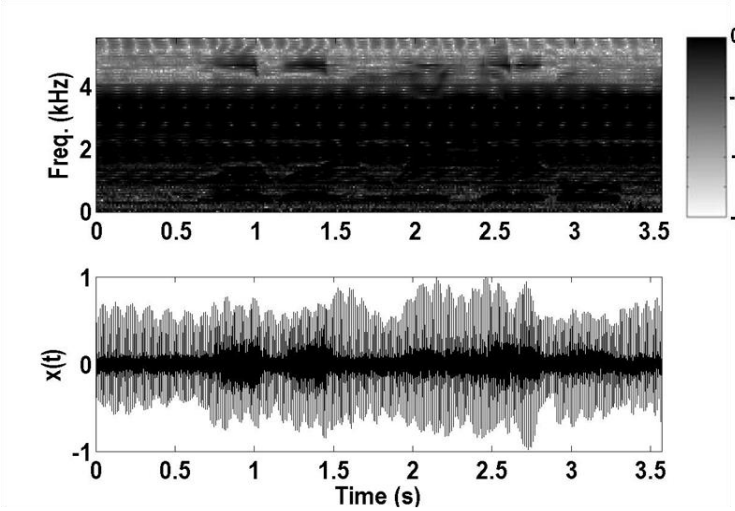
$$s_n(m) = 0.5\,[y_n(m) + y_{n-1}(m+M)],\ 0 \leq m \leq M\text{-}1 \tag{4.4}$$

The resulting values are stored in an array in memory and they are copied to buffer C or D when DMA channel 1 interrupt occurs. On-chip DAC outputs the data from DAC1RDAT which is continuously filled with the values from buffer C and D.

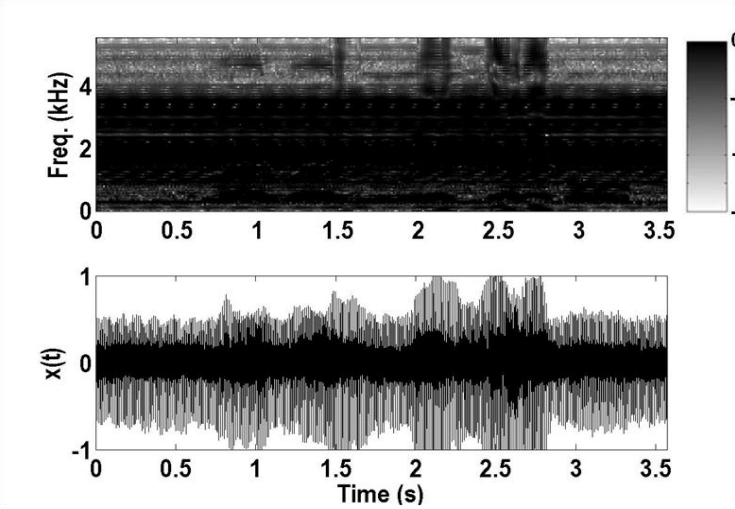
4.4. Results and discussion

Electrolaryngeal speech was recorded from two normal speakers, using electrolarynx models SolaTone (pitch frequency = 126.7 Hz) and NP-Voice (93.4 Hz), at a sampling rate of 11.025 kHz and 16-bit quantization. The recorded electrolayngeal speech from the PC sound card was processed by dsPIC33FJ128GP804 using program 'SS_EL_3p4s_CasMed' as described in the previous section. Sampling rate of ADC and DAC were set to 10 kHz. Spectral subtraction was performed using 2-pitch period frames with 50 % overlap. Median based noise estimation using 3-point 4-stage cascaded median approach was used to estimate the noise. The spectral subtraction parameters were empirically selected by checking the quality of the processed output. Best results were obtained with $\alpha = 2$ and results of the real-time implementation of spectral subtraction algorithm were found to be similar to those obtained using the Matlab based offline implementation. Similar output was obtained by using 3-point 5-stage cascaded median based noise estimation. Figure 4.6 shows the unprocessed electrolaryngeal speech and the enhanced electrolaryngeal speech for two sets of spectral subtraction parameters.

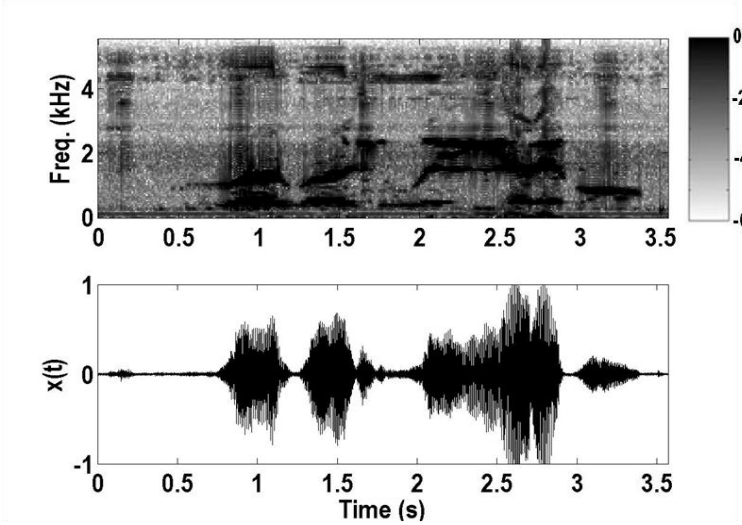
After real-time implementation of spectral subtraction algorithm, it was decided to implement LPC based analysis-synthesis algorithm (described in the section 3.3) for the introduction of jitter, shimmer, and spectral compensation on the enhanced electrolaryngeal speech after spectral subtraction. It was found that the on-chip data memory available and the clock speed of dsPIC33FJ128GP804 were not sufficient to implement the LPC based analysis-synthesis algorithm. So it was decided to use another processor having higher RAM and clock speed. TMS320C5515 eZdsp USB stick based on the digital signal processor TMS320C5515, having approximately three times the clock speed and twenty times the RAM compared to dsPIC33FJ128GP804 is selected to implement both the algorithms in real-time. The next chapter describes the real-time implementation of spectral subtraction using this processor.



(a) Recorded speech waveform and its spectrogram.



(b) Speech after spectral subtraction with $\alpha = 0$, $\beta = 0$, and $\gamma = 1$.



(c) Speech after spectral subtraction with $\alpha = 2$, $\beta = 0.002$, and $\gamma = 1$.

Fig. 4.6 Unprocessed and enhanced electrolaryngeal speech after real-time implementation of spectral subtraction with noise estimation using 3-point 4-stage cascaded median approach for two sets of subtraction parameters using dsPIC33FJ128GP804. Speaker: ARJ, material: "*…Where were you a year ago?*", generated using NP Voice electrolarynx.

Chapter 5

REAL-TIME SPECTRAL SUBTRACTION USING TMS320C5515

5.1 Introduction

This chapter presents real-time implementation of enhancement algorithm using TMS320C5515 eZdsp USB stick based on 16-bit fixed point processor TMS320C5515 (from Texas Instruments) [31].

5.2 Hardware

The block diagram of TMS320C5515 eZdsp USB Stick is shown in Fig. 5.1. The features of the TMS320C5515 eZdsp USB Stick evaluation tool [32] used in the real-time implementation are on-board DSP processor TMS320C5515, stereo codec TLV320AIC3204.

TMS320C5515 is an embedded controller having a 16-bit fixed-point digital signal processing unit with a maximum clock rate of 120 MHz and 16 MB of total memory. The memory map is as follows: on-chip RAM of 320 KB composing of 64 KB of Dual-Access RAM (DARAM) and 256 KB of Single-Access RAM (SARAM), 128 KB of on-chip ROM, and the remainder is for external memory interface. The DARAM is composed of eight blocks of 4K words each. Each DARAM block can perform two accesses per cycle (two reads, two writes, or a read and a write). The SARAM is composed of 32 blocks of 4K words each. Each SARAM block can perform one access per cycle (one read or one write). Both DARAM and SARAM can be accessed by the internal program, data, or DMA buses. The device has a DMA controller with four DMA having 4 channels each (16 channels total), and three 32 bit general purpose timers. The functional block diagram of the processor chip is shown in Fig. 5.2. Real-time clock (RTC) oscillator and PLL are used to set the system clock to 100 MHz. The device has 'FFT Hardware Accelerator' which supports 8 to 1024-point (in power of 2) real and complex-valued FFTs.

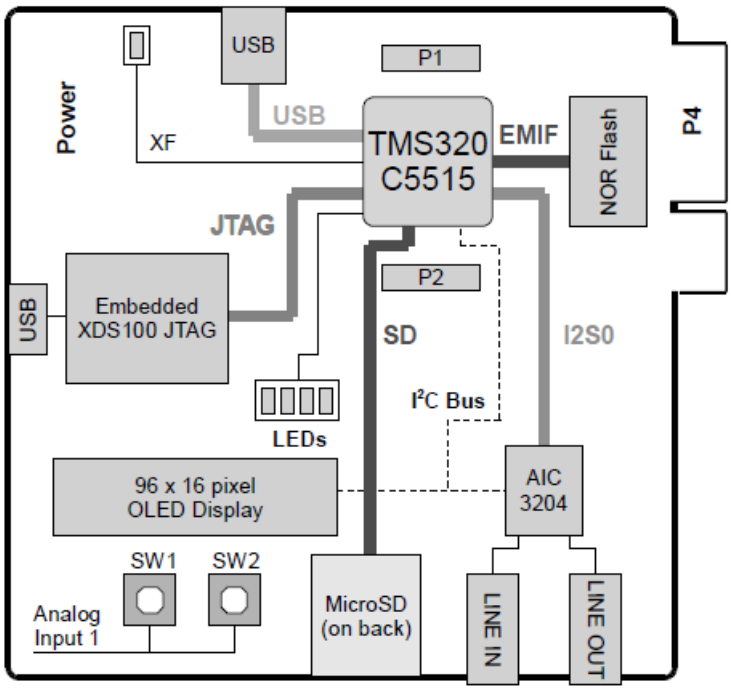


Fig. 5.1 Block diagram of TMS320C5515 eZdsp USB Stick reported in [32].

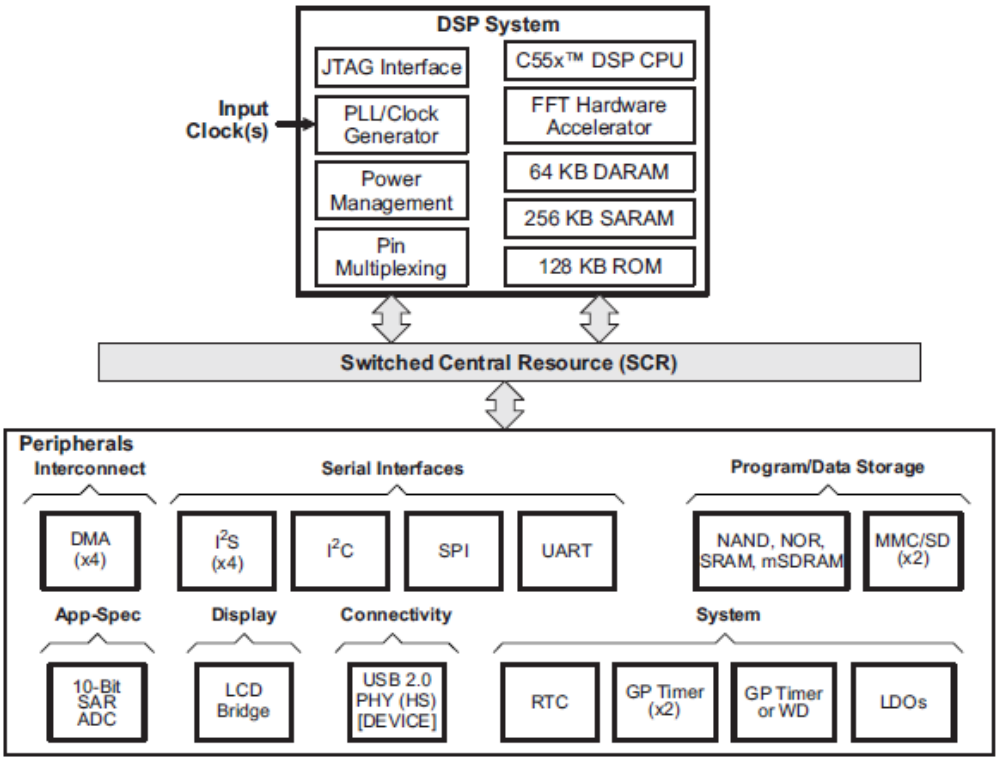


Fig. 5.2 Functional diagram of TMS320C5515 reported in [31].

TLV320AIC3204 (from Texas Instruments) [33] is a low-power stereo codec. The device has stereo ADC supporting sampling rates from 8 kHz to 192 kHz. In the present application ADC sampling rate is set to 12 kHz. The ADC uses a delta-sigma modulator and it can be powered up to a single channel, both channels, or no channels at a time. The ADC has six analog inputs which can be configured as either 3 stereo single-ended pairs or 3 fully-differential pairs. The device has stereo DAC supporting sampling rates from 8 kHz to 192 kHz.

5.3 Software

All the programs are written in C and loaded in the on-chip program memory of the processor with the help of ‘Code Composer Studio' (CCStudio) version 4.0, which is the integrated development environment for Texas Instruments' (TI) DSPs, microcontrollers and application processors. It includes compilers for each of TI's device families, source code editor, project build environment, debugger, profiler, simulators and many other features.

Before the real-time implementation of spectral subtraction algorithm the two programs, ‘loop_back_TI’ for ADC-DAC loop back, ‘FFT_IFFT_TI’ for FFT-IFFT of the input data, described in Sec. 4.3 are implemented for verifying the codec operation and FFT and IFFT calculations. Processor clock is set to 100 MHz and the sampling frequency of the ADC in codec is set to 12 kHz. FFT and IFFT are calculated using the predefined function 'hwafft_256pts' (FFT length considered in the program is 256) provided in the header file ‘hwafft.h’. Before passing the input data to this function, the data has to be bit reversed to facilitate radix-2 decimation in time computation. This bit reversing is done using the inbuilt function 'hwafft_br' which is also predefined in ‘hwafft.h’. FFT or IFFT operation is in-place if the return value of the function 'hwafft_256pts' is 0, out of place if the returned value is 1. The input and output values of these functions must be complex numbers represented by 32 bit integer data types in which most significant 16 bits are considered to be real part and least significant 16 bits are considered to be imaginary part of the complex number. Fractional arithmetic is used in the calculations of FFT and IFFT by considering real and imaginary parts of the complex number to be in Q1.15 format.

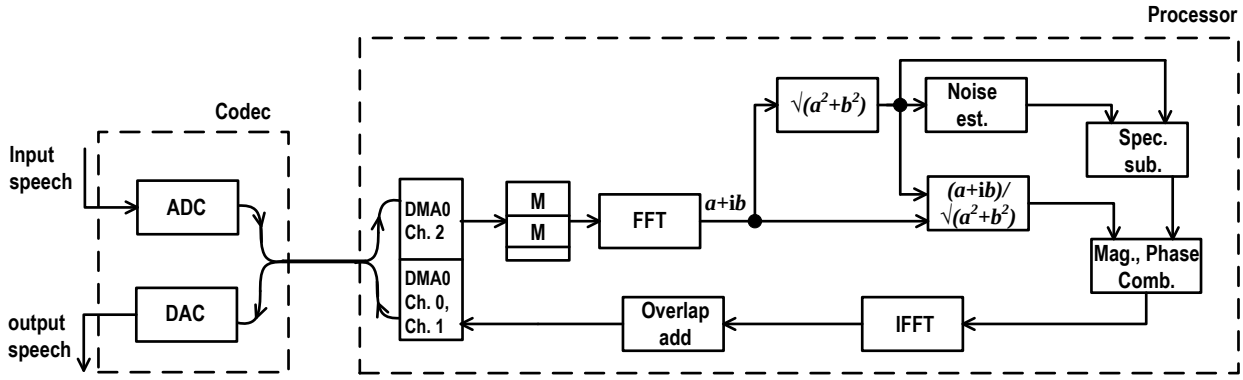
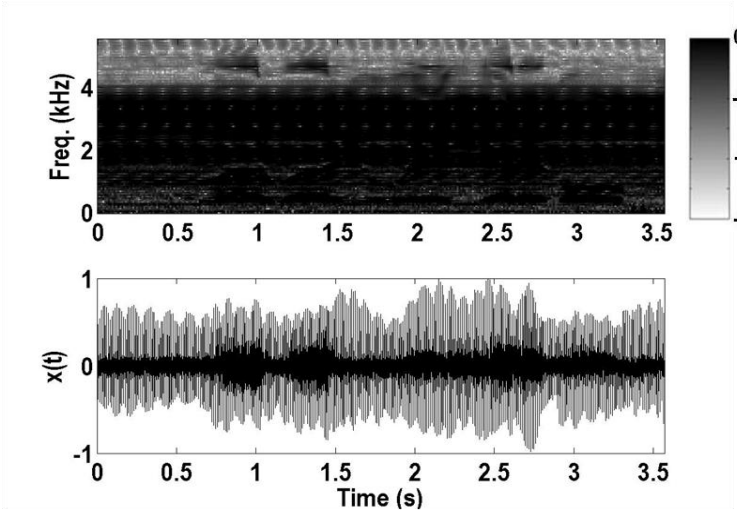


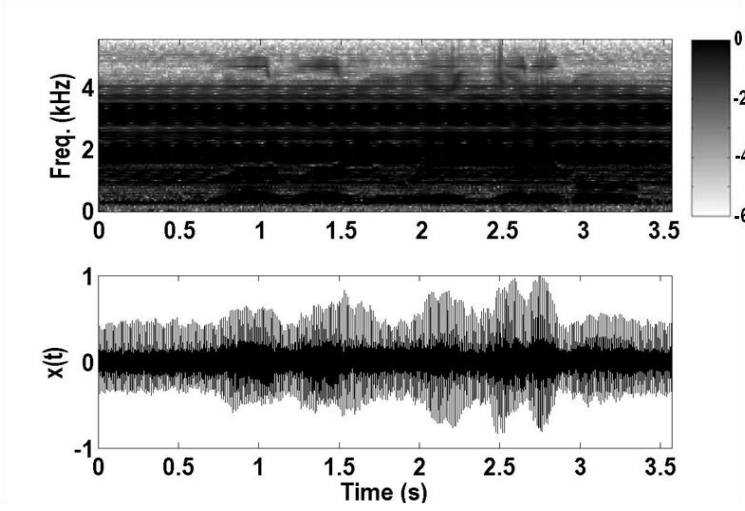
Fig. 5.3 Block diagram of the real-time implementation of spectral subtraction algorithm.

The implementation was carried out using the program 'SS_EL_3p4s_CasMed_TI' as shown in the Fig. 5.3. Input data samples from codec (data is taken only from the left channel as the input signal is mono) are stored in arrays 'RCVL1' and 'RCVL2' of length equal to a pitch-period using DMA. DMA0 channel2 is used for storing input data. The two registers 'DMA0_CH2_DST_MSW' and 'DMA0_CH2_DST_LSW' contain the most and least 16 bits of the 32 bit address respectively where the input data from codec is written using DMA. Initially these two registers contain the starting address of the array 'RCVL1' and after filling each element in the array the address in the registers is incremented. Whenever 'RCVL1' is filled, DMA interrupt occurs and starting address of 'RCVL2' is loaded in to the two registers and the value of a flag 'CurrentRxL_DMAChannel' (initialized to 1 at the starting of the program) is changed to 2. Whenever 'RCVL2' is filled, DMA interrupt occurs and starting address of 'RCVL1' is loaded in to the two registers and the value of the flag 'CurrentRxL_DMAChannel' is changed back to 1 and this process is repeated. A new array, 'temp_input', of length equal to FFT length i.e. 256 in the current application, is initialized to zeros. This array can be considered as 3 parts: 0 to *M*, *M*+1 to 2*M*, 2*M*+1 to 255 (*M* is the length of the pitch period). In function 'main' the value of the flag 'CurrentRxL_DMAChannel' is monitored continuously and whenever its value is changed from 1 to 2 the elements of the array 'RCVL1' (if the change is from 2 to 1 elements of the array 'RCVL2') replace the $(M+1)^{th}$ to $2M^{th}$ elements of the array 'temp_input'. This array serves as the input to the bit reverse function 'hwafft_br' and the output of this function is given as input to the function 'hwafft_256pts' to calculate the FFT. Spectral magnitudes are calculated from the complex FFT values. Noise is estimated using 3-point

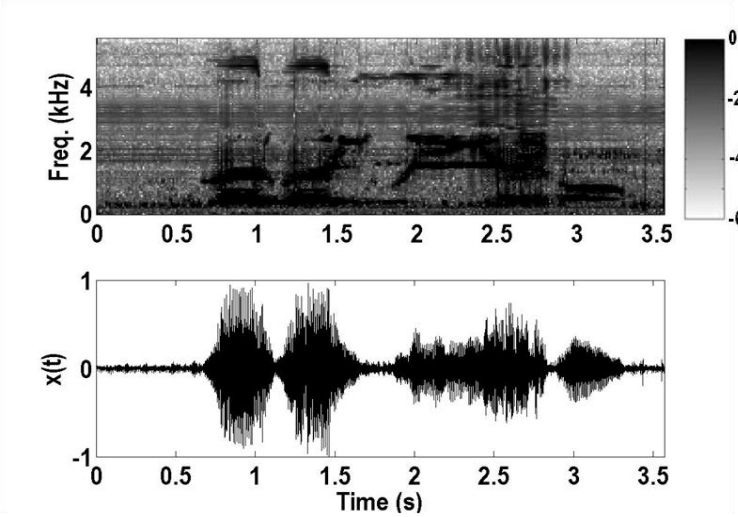
4-stage cascaded median approach (described in Sec 3.2). The spectral subtraction is carried out using $\gamma = 1$ and settable values of α and $\beta = 0$. The resulting magnitude is combined with the original phase, and IFFT of the resulting complex spectral values is calculated and the output speech is obtained after overlap-add as per the equations 4.1 to 4.4. The resulting values are stored in array 'XmitL1' (or 'XmitL2') if the input is taken from 'RCVL1'(or 'RCVL2'). DMA0 channel0 and DMA0 channel1 are used to copy the elements in 'XmitL1' or 'XmitL2' to the locations corresponding to left and right channels of line out in the codec respectively. After the processing is over, 0 to M elements of the array 'temp_input' are replaced by $(M+1)^{th}$ to $2M^{th}$ elements.



(a) Recorded speech waveform and its spectrogram.



(b) Speech after spectral subtraction with $\alpha = 0$, $\beta = 0$, and $\gamma = 1$.



(c) Speech after spectral subtraction with α = 1.625, β = 0, and γ = 1.

Fig. 5.4 Unprocessed and enhanced electrolaryngeal speech after real-time implementation of spectral subtraction with noise estimation using 3-point 4-stage cascaded median approach for two sets of subtraction parameters using TMS320C5515 eZdsp stick. Speaker: ARJ, material: "*…Where were you a year ago?*", generated using NP Voice electrolarynx.

5.4. Results and discussion

Electrolaryngeal speech was recorded from two normal speakers, using electrolarynx models SolaTone, and NP-Voice at a sampling rate of 11.025 kHz and 16-bit quantization. The recorded electrolayngeal speech from the PC sound card was processed using TMS320C5515 eZdsp USB stick based on the DSP processor TMS320C5515. Sampling rate of ADC and DAC were set to 12 kHz. Spectral subtraction was performed using 2-pitch period frames with 50 % overlap. Median based noise estimation using 3-point 4-stage cascaded median approach was used to estimate the noise. The spectral subtraction parameters were empirically selected for the better noise reduction. As mentioned earlier, the spectral subtraction was carried out using γ = 1. Best results were obtained with α = 1.625 and β = 0.002. Results of the real-time implementation of spectral subtraction algorithm were found to be similar to those obtained using the Matlab based offline implementation. Figure 5.4 shows the unprocessed and enhanced electrolaryngeal speech for two sets of spectral subtraction parameters.

Chapter 6

SUMMARY AND SUGGESTIONS FOR FUTURE WORK

6.1 Summary

Electrolaryngeal speech suffers from background leakage noise, monotonic nature, and low-frequency spectrum deficit. Quality of the electrolaryngeal speech is affected by the deficiency in low-frequency content and monotonic nature of the speech while background noise also affects intelligibility. Several signal processing techniques [4]-[10] have been reported in the literature. It has been shown in [5], [17] that if spectra are calculated pitch synchronously, speech and the background leakage noise become uncorrelated and spectral subtraction can be used to remove the background leakage noise. Average based noise estimation (ABNE) reported in [5], estimates the noise from the initial silence segment of about 2-s duration when speaker has closed his lips. This method does not update the noise dynamically. So it is not effective for continuous speech. Statistical techniques [6], [9], [10], [19]-[22] have been reported to dynamically update the estimated noise. In minimum statistics based noise estimation (MSBNE) [9], [10] for enhancement of electrolaryngeal speech, noise of a spectral sample is estimated as the minimum of magnitudes of that spectral sample in a set of past frames. In quantile based noise estimation (QBNE) [6], the quantile values for different spectral components were selected by matching the noise spectrum estimated over a long speech record to match that obtained by averaging during the initial silence. Budiredla [23], and Mitra [24] have earlier implemented spectral subtraction algorithm in real-time. Budiredla [23] reported that the background noise reduction in the output speech was not effective compared to the off-line C implementation. Mitra [24] reported that though the implementation showed effective background noise reduction, output speech was not as good as that obtained from offline processing.

In the present work, offline implementation of background noise reduction in electrolaryngeal speech was carried out using generalized spectral subtraction algorithm reported in [15], [16] and with ABNE, MSBNE, and median based noise estimation (QBNE with 0.5 quantile reported by Pandey *et al.* [6]). In the implementation of the

spectral subtraction with ABNE, background noise was reduced effectively in shorter sentences of duration less than 5 s length. But the noise reduction was not effective if coupling of the electrolarynx with neck was varied during the speech, even in shorter sentences. Spectral subtraction with dynamical estimation of noise using MSBNE did not result in effective noise reduction if there are no frequent pauses in the speech. A set of past frames corresponding to a length of 3 s was required for better noise estimation. There was an effective noise reduction in output speech after the implementation of spectral subtraction with MBNE. Better estimate of noise was obtained with a set of past frames corresponding to a speech segment of length 0.5 s. If the coupling of electrolarynx with the neck is varied, some background noise was present in the output speech for the duration corresponding to the number of past frames used to estimate the noise in both MSBNE and MBNE based spectral subtraction. For real-time implementation In spectral subtraction with all the three noise estimation techniques, it was observed that higher values of over subtraction parameter α were required if the input speech amplitude was higher. Also, it was observed that quality of the output speech was improved with higher overlapping samples.

It was decided to implement spectral subtraction using MBNE in real-time, but the memory of the selected processor, dsPIC33FJ128GP804 (from Microchip), was not sufficient to store the magnitude values of past frames corresponding to a length of 0.5 s. Hence a 3-point 4-stage cascaded median approach (described in Sec. 3.2) was used to estimate the noise. Offline implementation of this method in Matlab resulted in effective noise reduction similar to that with MBNE. There was no improvement in the quality of output speech by using 3-point 5-stage cascaded median approach.

In addition to the resynthesis using the original noisy phase, resynthesis using different techniques of estimating the phase spectrum was investigated. Speech quality for both the types of minimum-phase estimation was similar and not better than that obtained by using the phase of the noisy speech. Use of zero and random phases resulted in poor quality.

An introduction of jitter and shimmer in the speech signal, using LPC based analysis-synthesis, was investigated for improving its naturalness. A peak-to-peak jitter of up to 6 % increased the naturalness, while introduction of shimmer up to 20 % did not improve the quality and larger values of shimmer degraded the speech. Spectral compensation of the output speech after spectral subtraction with MBNE without using LPC analysis-synthesis was also carried out. The resulted speech was better in quality

compared to output speech after spectral subtraction but not as good as that obtained using LPC analysis-synthesis with 6 % jitter.

Real-time implementation of generalized spectral subtraction algorithm was carried out using dsPIC33FJ128GP804. After the analysis of offline implementations with different noise estimation techniques, it was decided to implement the spectral subtraction with MBNE and to resynthesize the magnitude spectrum after spectral subtrction with retained noisy phase. To overcome the constraint on the number of past frames used for estimation of noise due to data memory size, a 3-point 4-stage cascaded median approach (described in Sec 3.2) was used. Output speech was comparable to that obtained from offline implementation with Matlab. Due to constraints of on-chip available memory and processing clock speed, LPC based analysis-synthesis algorithm could not be implemented after spectral subtraction in dsPIC33FJ128GP804. It was found that implementation of the LPC based analysis-synthesis algorithm alone cannot be implemented on dsPIC33FJ128GP804 due to its memory and processing speed limitations. Use of the assembly language programming for real-time implementation needs to be investigated. Later TMS320C5515 eZdsp USB stick based on TMS320C5515 (from Texas Instruments) DSP processor, having approximately three times clock speed and twenty times RAM was selected to implement both the algorithms. Spectral subtraction with 3-point 4-stage cascaded median approach for noise estimation was implemented on TMS320C5515 eZdsp USB stick based on TMS320C5515. Real-time implementation of LPC based analysis-synthesis algorithm on this processor was not completed. Results of real-time implementation of spectral subtraction using TMS320C5515 eZdsp USB stick are similar to those obtained from the offline implementation using Matlab.

6.2 Suggestions for future work

Real-time implementation of LPC based analysis-synthesis algorithm to improve the quality of the output speech after spectral subtraction needs to be investigated. Subjective evaluation of intelligibility and quality of output speech need to be carried out. Instructions for evaluating the quality of the output speech using ‘Preference test’ are given in appendix A.

Appendix A

PREFERENCE TEST

A.1 Instructions for preference test

This is a listening test involving presentation of pairs of speech sounds. In each pair, the two sounds marked as A and B correspond to the same sentence or word(s) but they may differ in quality. After presentation of each pair of sounds, you have to respond by indicating the sound you found to be of better quality. You will be seated in front of a computer monitor and the test will be conducted in an automated manner, by presenting the sounds and recording your responses. The sounds will be presented using a speaker or a pair of headphones, with the volume (level of the sounds) adjusted to the most comfortable level for you.

During the test, the screen shows the current presentation number and the total number of presentations. There are five buttons marked as PLAY, A, B, NEXT, END. Below the button marked PLAY, there are two boxes marked A and B. After PLAY is clicked, the sounds A and B are presented with a gap of 0.5 s and the sound being presented is indicated by highlighting the corresponding box. The response buttons appear inactive until the sounds have been presented. The first three buttons appear active after the presentation. You can indicate your response by clicking on A or B depending on which one is perceived to be of better quality, or you can listen to the sounds again by clicking on PLAY. After the response, NEXT and END buttons become active. You can change the response by clicking on the other response button, or you may listen to the pair of sounds again. Once you are sure of your response, click on the NEXT for the next presentation. Clicking on END will terminate the test.

The sequence of presentations will be continued until the display shows “Test is over. Thank you for your participation”.

REFERENCES

[1] M. Weiss, G. Y. Komshian, and J. Heinz, "Acoustical and perceptual characteristics of speech produced with an electronic artificial larynx," *J. Acoust. Soc. Am.*, vol. 65, No. 5, pp. 1298-1308, 1979.

[2] Y. Lebrun, "History and development of laryngeal prosthetic devices," in *The Artificial Larynx*, Amsterdam: Swets and Zeitlinger, pp. 19-76, 1973.

[3] L. P. Goldstein, "History and development of laryngeal prosthetic devices," in *Electrostatic Analysis and Enhancement of Alaryngeal Speech*, Springfield, Ill: Charles C. Thomas, pp. 137-165, 1982.

[4] C. Y. Espy-Wilson, V. R. Chari, and C. B. Huang, "Enhancement of alaryngeal speech by adaptive filtering," in *Proc. ICSLP*, 1996, pp. 764-771.

[5] P. C. Pandey, S. M. Bhandarkar, G. K. Bachher, and P. K. Lehana, "Enhancement of alaryngeal speech using spectral subtraction" in *Proc. 14th Int. Conf. Digital Signal Processing (DSP 2002)*, Santorini, Greece, 2002, pp. 591-594.

[6] P. C. Pandey, S. S. Pratapwar, and P. K. Lehana, "Enhancement of electrolaryngeal speech by reducing leakage noise using spectral subtraction with quantile based dynamic estimation of noise", in *Proc. 18th International Congress on Acoustics*, (*ICA2004*), Kyoto, Japan, 2004, pp. 3029-3032.

[7] H. Liu, Q. Zhao, M. Wan and S. Wang, "Enhancement of electrolarynx speech based on auditory masking," *IEEE Trans. Biomed. Eng.*, vol. 53, pp. 865-874, 2006.

[8] H. Liu, Q. Zhao, M. Wan, and S. Wang, "Application of spectral subtraction method on enhancement of electrolaryngeal speech," *J. Acoust. Soc. Am.*, vol. 120, No. 1, pp. 398-406, July 2006.

[9] P. Mitra and P.C. Pandey, "Enhancement of electrolaryngeal speech by spectral subtraction with minimum statistics-based noise estimation," (abstract), *J. Acoust. Soc. Amer.*, vol. 120, p. 3039, 2006.

[10] R. Kabir, A. Greenblatt, K. Panetta, and S. Agaian, "Enhancement of alaryngeal speech utilizing spectral subtraction and minimum statistics," in *Proc. 7th Int. Conf. Machine Learning and Cybernetics*, Kunming, 2008, 12-15 July.

[11] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Englewood Cliffs, New Jersey: Prentice Hall, 1978.

[12] H. L. Barney, F. E. Haworth, and H. K. Dunn, "An experimental transistorized artificial larynx," *Bell Systems Tech. J.*, vol. 38, No. 6, pp. 1337-1356, 1959.

[13] Q. Yingyong and B. Weinberg, "Low-frequency energy deficit in electrolaryngeal speech," *J. Speech and Hearing Research*, vol. 34, pp. 1250- 1256, 1991.

[14] R. L. Norton, and R. S. Bernstein, "Improved LAboratory Prototype ELectrolarynx (LAPEL): Using inverse filtering of the frequency response function of the human throat," in *Annals of Biomedical Engineering*, 1993, Vol. 21, pp 163-174.

[15] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, Vol. 27, No. 2, pp. 113-120, 1979.

[16] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE ICASSP*, 1979, pp. 208-211.

[17] S. M. Bhandarkar / P. C. Pandey (Supervisor), "Reduction of background noise in artificial larynx" *M.Tech. Dissertation*, Dept. of Electrical Engineering, IIT Bombay, January 2002.

[18] P. C. Pandey and S. K. Basha, "Enhancement of electrolaryngeal speech by spectral subtraction, spectral compensation, and introduction of jitter and shimmer," in *Proc. 20th International Congress on Acoustics (ICA 2010)*, Sydney, Australia, 2010, Paper no. 670.

[19] S. S. Pratapwar / P. C. Pandey (Supervisor), "Reduction of background noise in artificial larynx", *M.Tech. Dissertation*, Dept. of Electrical Engineering Department, IIT Bombay, Feb. 2004.

[20] V. Stahl, A. Fisher, and R. Bipus, "Quantile based noise estimation for spectral subtraction and wiener filtering," in *Proc. IEEE ICASSP*, 2000, Vol. 3, pp. 1875-1878.

[21] R. Martin, "Spectral subtraction based on minimum statistic," in *Proc. 7th European Signal Processing Conf., EUSIPCO-94*, Edinburgh, Scotland, 13-16 September 1994, pp. 1182-1185.

[22] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech and Audio Processing*, Vol. 9, No 5, pp. 504-512, July 2001.

[23] B. R. Budiredla / P. C. Pandey (Supervisor), "Real-time implementation of spectral subtraction for enhancement of electrolaryngeal speech", *M.Tech. Dissertation,* Dept. of Electrical Engineering Department, IIT Bombay, July 2005.

[24] P. Mitra / P. C. Pandey (Supervisor), "Enhancement of electrolaryngeal speech by background noise reduction and spectral compensation", *M.Tech. Dissertation,* Dept. of Electrical Engineering Department, IIT Bombay, July 2006.

[25] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*, Englewood Cliffs, New Jersey: Prentice Hall, 1975.

[26] B. Bozkurt, and T. Dutoit, "Mixed-phase speech modeling and formant estimation, using differential phase spectrums," in *Proc. ISCA Voice Quality Conf.*, Geneva, Switzerland, CD-ROM, 2003.

[27] T. F. Quatieri and A. V. Oppenheim,"Iterative techniques for minimum phase signal reconstruction from phase or magnitude," *IEEE Trans. Acoust., Speech, Signal Process.,* vol 29, pp. 1187-1193,1981.

[28] S. H. Nawab, T. F. Quatieri, and J. S. Lim "Signal reconstruction from short-time Fourier transform magnitude," *IEEE Trans. Acoust., Speech, Signal Process.,* vol 31, pp. 986-998,1983.

[29] B. Yegnanarayana and A. Dhayalan, "Noniterative techniques for minimum phase signal reconstruction from phase or magnitude," in *Proc. IEEE ICASSP,* 1983, pp. 639-642.

[30] Microchip Technology Inc., "dsPIC33FJ32GP302/304, dsPIC33FJ64GPX02/X04, and dsPIC33FJ128GPX02/X04 Data Sheet: High-Performance 16-bit Digital Signal Controllers," 2009, [online] Available: ww1.microchip.com/downloads/en/DeviceDoc/70292D.pdf.

[31] Texas Instruments Inc., "TMS320C5515 Fixed-Point Digital Signal Processor," 2011, [online] Available: http://focus.ti.com/lit/ds/symlink/tms320c5515.pdf.

[32] Spectrum Digital Inc., "TMS320C5515 eZdsp USB Stick: Technical Reference," 2010, [online] Available:

http://support.spectrumdigital.com/boards/usbstk5515/reva/files/usbstk5515_TechRef_RevA.pdf.

[33] Texas Instruments Inc., “TLV320AIC3204 Ultra Low Power Stereo Audio Codec,” 2008, [online] Available: http://focus.ti.com/lit/ds/symlink/tlv320aic3204.pdf

Acknowledgements

I would like to express my gratitude to my respected guide Prof. P.C. Pandey, for his invaluable guidance, support, and encouragement throughout the course of this project. I am also thankful to him for sparing his invaluable time in correcting my reports, and staying till late nights with me in the lab during the work related to the paper presented at ICA 2010. Also I am thankful to him for sharing knowledgeable discussions.

I am thankful to Nandakumar Pai for his great support during the real-time implementation work. I would like to thank Vidyadhar Kamble for helping me in all the lab related issues, and Nataraj, Rajath, and Jagbandhu for their helpful advices in the different issues of the project and sparing their time in correcting my reports. I am thankful to Dr. Panduranga Kulkarni, for his help in reviewing my paper draft and presentation for ICA2010. I would like to thank Jayan, Parveen Lehana, and Santosh for their help with programming tips and sharing interesting discussions with me. I am also thankful to all my friends especially in the SPI lab, EI Lab, and in WEL for their whole-hearted support during the tenure of this project.

Shaik Khadar Basha

June 2011

Author's Resume

Shaik Khadar Basha: He received the B. Tech. degree in electrical and electronics engineering from the College of Engineering GITAM, Visakhapatnam (Andhra Pradesh) in 2005. He worked as an Assistant Professor at Geethanjali College of Engineering and Technology, Hyderabad from August 2006 to May 2007. Presently, he is pursuing the M.Tech. degree in electrical engineering at the Indian Institute of Technology Bombay. His research interests include embedded system design, digital signal processing, and speech processing.

Thesis related publication

P. C. Pandey and S. K. Basha, "Enhancement of electrolaryngeal speech by spectral subtraction, spectral compensation, and introduction of jitter and shimmer," in *Proc. 20th International Congress on Acoustics* (*ICA 2010*), Sydney, Australia, 2010, Paper no. 670.