Implementation of Multi-band Frequency Compression for Listeners with Moderate Sensorineural Impairment

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

Master of Technology
by

Nitya Tiwari
(10307037)

under the supervision of

Prof. P. C. Pandey

Department of Electrical Engineering
Indian Institute of Technology Bombay
June 2012

Indian Institute of Technology, Bombay

M. Tech. Dissertation Approval

This dissertation entitled **"Implementation of Multi-band Frequency Compression for Listeners with Moderate Sensorineural Impairment"** by **Nitya Tiwari** (Roll No. **10307037**) 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)
	.....	(Dr. K. Samudravijaya)
Chairperson	.....	(Prof. K.P. Karunakaran)

Date: 22 June 2012

Place: Mumbai

Declaration

I declare that this dissertation represents my ideas in my words and where ideas or words are taken from others, I have adequately cited and referenced the original sources. I declare that I have adhered to all principles of academic honesty and integrity and I 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.

(Nitya Tiwari)

Date: 22 June 2012

Place: Mumbai

Nitya Tiwari / Prof. P. C. Pandey (Supervisor): “Implementation of multi-band frequency compression for listeners with moderate sensorineural impairment”, *M.Tech. dissertation*, Department of Electrical Engineering, Indian Institute of Technology Bombay, June 2012.

ABSTRACT

Widening of auditory filters in persons with sensorineural hearing impairment leads to increased spectral masking and degraded speech perception. Multi-band frequency compression of the complex spectral samples using pitch-synchronous processing has been reported to increase speech perception by persons with moderate sensorineural loss. It is shown that implementation of multi-band frequency compression using fixed-frame processing along with least-squares error based signal estimation reduces the processing delay and the speech output is indistinguishable from pitch-synchronous processing.

For real-time operation, the processing is implemented on a DSP board based on the 16-bit fixed point processor TMS320C5515. Codec and DMA are used to continuously acquire the input signal and output the processed signal at a sampling rate of 10 kHz. The data transfer and buffering operations are devised for an efficient realization of analysis-synthesis with 75 % overlap. The spectral modification operations are facilitated by the on-chip FFT hardware. The real-time processing with analysis window length of 26 ms and 512-point FFT is implemented, using about one-tenth of the computing capacity of the processor. The processing delay is approximately 35 ms, making it suitable for hearing aid applications.

CONTENTS

List of abbreviations

Abbreviation	Explanation
ACB	auditory critical band
ADC	analog-to-digital converter
CDF	cumulative distribution function
CPU	central processing unit
DAC	digital-to-analog converter
DFT	discrete Fourier transform
DMA	direct memory access
DSP	digital signal processor
FFT	fast Fourier transform
FF	fixed frame
GCI	glottal closure instant
IDFT	inverse discrete Fourier transform
IFFT	inverse fast Fourier transform
LSEE	least square error estimation
MOS	mean opinion score
MRT	Modified Rhyme Test
PC	personal computer
PDF	probability density function
PESQ	Perceptual Evaluation of Speech Quality
PGA	programmable gain amplifier
PS	pitch synchronous
RMS	root mean square
SNR	signal-to-noise ratio
STFT	short-time Fourier transform
TI	Texas Instruments
USB	Universal Serial Bus

List of symbols

Symbols	Explanation
a	starting point of the spectral segment contributing to a sample on the compressed scale
b	end point of the spectral segment contributing to a sample on the compressed scale
f_1	lower edge of auditory critical band
f_2	upper edge of auditory critical band
f_c	center frequency of auditory critical band
F0	fundamental frequency of glottal excitation, pitch
k	spectral sample index on the unprocessed spectrum
k'	spectral sample index on the frequency compressed spectrum
k_{ic}	center frequency of the ith analysis band
k_{ie}	ending index of the ith analysis band
k_{is}	starting index of the ith analysis band
L	window length
m	lowest integer higher than a
n	highest integer lower than b
N	size of the DFT
p	constant determining analysis-synthesis window type
q	constant determining analysis-synthesis window type
S	window shift
$w(n)$	analysis-synthesis window
$x(n)$	real-valued discrete-time signal
$x'(n)$	resynthesized real-valued discrete-time signal
X	spectrum of the unprocessed speech signal
Y	spectrum of the frequency compressed speech signal
α	compression factor

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1. Overview

Sensorineural hearing impairment is caused by abnormalities in functioning of the hair cells in cochlea and the auditory nerve. It occurs due to aging, excessive exposure to noise, infection, or abnormalities at the time of birth. Sensorineural hearing loss is generally associated with widening of auditory filters, elevated hearing thresholds, loudness recruitment, reduced dynamic range, increased temporal and spectral masking leading to degraded speech perception [1]-[5] Speech audibility can be improved by frequency selective amplification, but it may not improve speech perception. Hence speech processing techniques are needed to overcome the intraspeech spectral masking and to improve speech perception.

1.2. Objective

Several signal processing techniques have been reported for reducing the effect of increased intraspeech spectral masking caused by widening of auditory filters. Binaural dichotic presentation has been used to reduce the effect of spectral masking for persons with moderate bilateral loss [6]-[9]. In case of monaural hearing, spectral contrast enhancement and multi-band frequency compression may reduce the effects of widened auditory filters [10]-[19]. Multi-band frequency compression of the complex spectral samples using pitch-synchronous processing has been reported to improve speech perception by persons with moderate sensorineural loss. The objective is to investigate this technique for its real-time implementation for use in hearing aids. This will involve devising methods for reducing the computational requirement for implementing it on a DSP chip, without degradation in the quality of the output speech. A processing technique with processed speech quality similar to that obtained using pitch-synchronous processing but with lesser computational overhead is described and is implemented for real-time processing on a DSP board

based on the 16-bit fixed point processor TMS320C5515, using a fraction of its computing capacity.

1.3. Outline

The signal processing techniques for reducing the effect of increased intraspeech spectral masking, and multi-band frequency compression are described in Chapter 2. Chapter 3 covers the investigations on Matlab based offline implementation of multi-band frequency compression. The investigations based on real-time implementation of multi-band frequency compression on a DSP board using 16-bit fixed point processor TMS320C5515 are described in Chapter 4. The last chapter provides summary and conclusion.

Chapter 2

MULTI-BAND FREQUENCY COMPRESSION

2.1. Introduction

Sensorineural hearing loss is generally associated with elevated hearing thresholds, loudness recruitment, reduced dynamic range, and increased temporal and spectral masking, leading to degraded speech perception [1]-[5]. Most of the hearing aids provide frequency-selective gain and automatic gain control [20]. Advanced hearing aids also have the facility for multi-channel dynamic range compression with settable parameters like attack time, release time, number of bands, and compression ratios in different bands [20],[21]. Several signal processing techniques have been reported for reducing the effect of increased intraspeech spectral masking caused by widening of auditory filters. Binaural dichotic presentation, using a pair of comb filters with complementary magnitude responses for spectral splitting, has been used to reduce the effect of spectral masking for persons with moderate bilateral loss [6]-[9]. In case of monaural hearing, spectral contrast enhancement and multiband frequency compression may reduce the effects of widened auditory filters. Spectral contrast enhancement is aimed at increasing the intelligibility of speech in noise for normal-hearing subjects and for subjects with sensorineural hearing loss [10]-[14]. It involves enhancing the perceptually important spectral peaks, thus increasing the contrast between spectral peaks and valleys. However, errors in identifying these peaks may subdue the advantage of spectral contrast enhancement. Further, the processing may result in an increase in the dynamic range of the speech signal, and thus may adversely affect speech perception by persons with recruitment (abnormal growth of loudness commonly associated with sensorineural loss).

Multi-band frequency compression [15]-[19] is another technique for reducing the effects of increased intraspeech spectral masking. In this technique, speech energy is presented in relatively narrow bands to avoid masking by adjacent spectral components. The processing involves dividing speech spectrum into analysis bands and compressing the spectral samples in each band towards the band center. Arai et al.

[15] applied multi-band frequency compression using auditory critical bandwidths on the magnitude spectrum and obtained the complex spectrum by associating it with the original phase spectrum. The spectrum was compressed using compression rates of 10% to 90%. Listening tests on subjects with hearing impairment showed best results with compression factor of 50% and the recognition score improved from 35.4 % for unprocessed speech to 38.3 % for the processed speech.

For decreasing the computation and reducing the processing related artifact, Kulkarni et al. [17] applied the compression on the complex spectrum without calculating the magnitude and phase spectra. Multi-band frequency compression, as implemented by Kulkarni et al. [17], concentrates the complex spectral samples in relatively narrower bands and is aimed at avoiding the intraspeech spectral masking. The time domain signal is segmented and complex spectrum obtained by discrete Fourier transform (DFT) is divided into a fixed number of analysis bands. The complex spectral samples within each band are compressed towards the band center depending on the compression factor. The output speech is resynthesized using IDFT and overlap-add.

2.2. Signal Processing

Multi-band frequency compression uses three steps: (A) segmentation and spectral analysis, (B) spectral modification, and (C) resynthesis. The results of multi-band frequency compression for different types of segmentation, bandwidths, and frequency mapping methods were investigated by Kulkarni et al. [17].

A. Segmentation and Spectral Analysis

The speech signal was segmented using two types of segmentation schemes: (i) fixed-frame segmentation, and (ii) pitch-synchronous segmentation. For the speech signal sampled at 10 kHz, 20 ms window with 50% overlap was chosen for fixed-frame segmentation. Pitch-synchronous segmentation used the window length equal to twice the local pitch period with an overlap of one pitch period. The pitch period was determined by finding the glottal closure instant (GCI) using Childers and Hu's algorithm [22]. When the speech signal was voiced in a segment, the window length was chosen so as to span three successive glottal closure instants. For the unvoiced segments the window length was chosen to be equal to that of the last voiced segment. Each segment is zero padded to form a sequence of length say N and N-point DFT is

used to get the complex spectrum. The complex spectrum is divided into a fixed number of analysis bands. The effect of three types of bandwidths, (i) constant bandwidth with number of bands varying from 2 to 18 in the 0 – 5 kHz frequency range, (ii) 1/3-octave bands, and (iii) bands based on auditory critical bandwidth (ACB) [23], were investigated.

B. Spectral Modification

Three frequency mapping techniques were investigated: (i) sample-to-sample mapping, (ii) spectral sample superimposition, and (iii) spectral segment mapping.

1) Sample-to-Sample Mapping: In this frequency mapping method, as explained in [17], the frequency index k of the original spectrum X is mapped to the frequency index k' of the compressed spectrum Y as

$$k' = k_{ic} + \text{round}(\alpha\,(k - k_{ic})) \tag{2.1}$$

where α is the compression factor (0 – 1). The center frequency k_{ic}, of the i th analysis band is given as

$$k_{ic} = 0.5(k_{is} + k_{ie}) \tag{2.2}$$

where k_{is} and k_{ie} are starting and ending indices for the i th band. The spectral samples are obtained as $Y(k') = X(k)$. Figure 2.1 shows sample to sample mapping method for 1.25 – 1.45 kHz band with center frequency of 1.35 kHz. In this type of mapping, there is a possibility of two or more input spectral samples getting mapped to same output sample. In such a situation, only the input sample with largest index amongst the overlapping samples is retained. This elimination of some samples leads to irregular variations in the spectrum and reduction in energy of signal.

2) Spectral Sample Superimposition: The problem of missing samples in sample-to-sample mapping is solved by adding the spectral samples which map to the same frequency index. As the number of input spectral samples which contribute to an output sample may vary, some irregular variations occur in the compressed spectrum. Figure 2.2 illustrates spectral sample superimposition mapping.

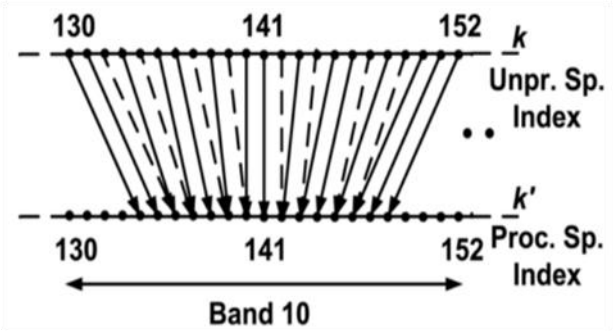


Figure 2.1. Sample-to-sample mapping for frequency band 1.25−1.45 kHz with 10 kHz sampling [17].

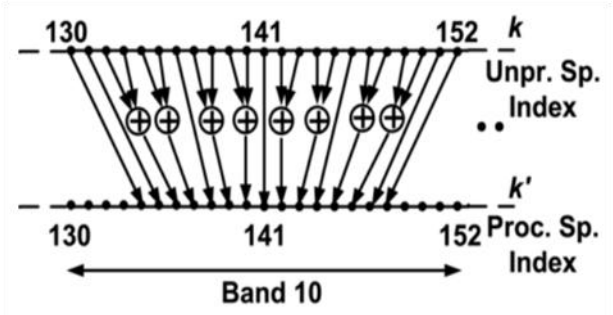


Figure 2.2. Spectral sample superimposition for frequency band 1.25−1.45 kHz with 10 kHz sampling [17].

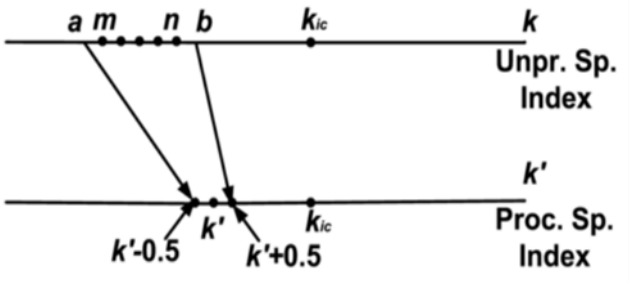


Figure. 2.3. Spectral segment mapping [17].

3) Spectral Segment Mapping: In this mapping, as shown in Figure 2.3, a one-sample interval centered on the output frequency sample is mapped to a corresponding segment of the frequency axis of the input spectrum. The edges *a* and *b* of the input frequency segment for the output spectral sample with frequency index *k'* in the *i* th analysis band with center frequency k_{ic} and compression factor α are given as

$$a = k_{ic} - [(k_{ic} - (k' - 0.5))/\alpha] \tag{2.3}$$

$$b = a + 1/\alpha \tag{2.4}$$

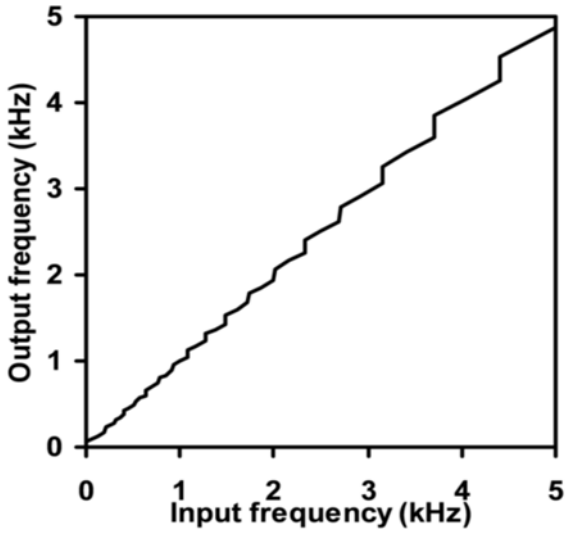


Figure 2.4. Frequency mapping for multi-band frequency compression, with auditory critical bandwidths and α = 0.6 [17].

The frequency sample in the compressed spectrum is calculated from samples of the complex spectrum as

$$Y(k') = (m\text{-}a)\,X(m) + \sum_{j=m+1}^{n-1} X(j) + (b-n)\,X(n) \qquad (2.5)$$

where m and n are the indices of the first and the last spectral samples in $[a,b]$, respectively. In this mapping, all the samples of input spectrum contribute uniformly to the compressed spectrum, without any irregular variations in the spectral energy.

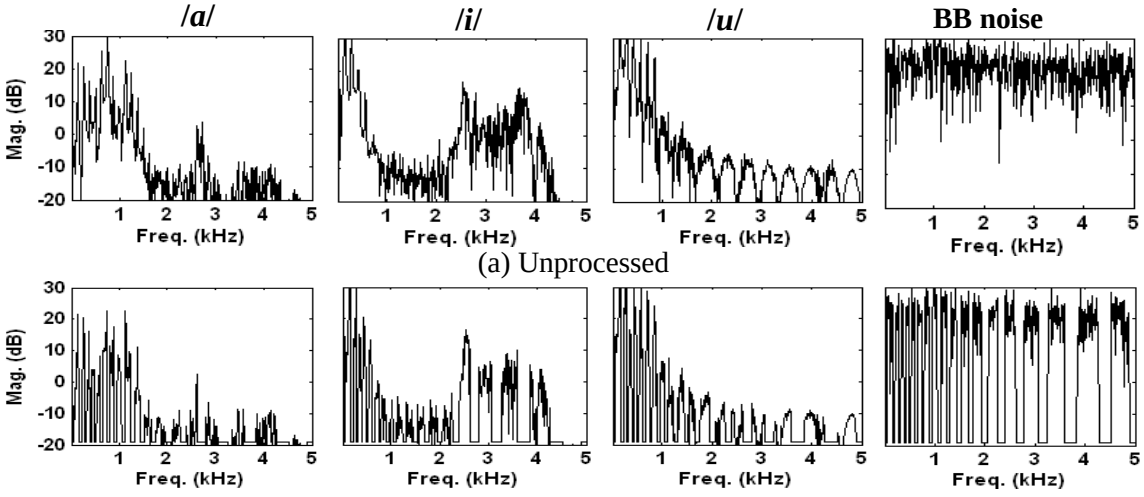
C. Resynthesis

The N-point IDFT of the modified complex spectrum was used for resynthesizing the speech signal by overlap-add method [24],[25].

Best results were obtained for auditory critical bandwidth based compression using spectral segment mapping and pitch-synchronous analysis-synthesis. Frequency mapping for multi-band frequency compression, with auditory critical bandwidths and compression factor of 0.6 is shown in Figure 2.4. Table 2.1 shows auditory critical bands with center frequency f_c , lower edge f_1 , and upper edge f_2. Spectra of 100 ms segments of vowels /*a*/, /*i*/, /*u*/, and broad-band noise for the unprocessed and the processed signals using fixed-frame analysis are shown in Figure 2.5. It is seen that the processing concentrates the spectral energy in narrow bands and it does not introduce any spectral tilt nor does it lead to compression of broadband spectrum.

Table 2.1. Auditory critical bands with center frequency f_c, lower edge f_1, and upper edge f_2 [19].

Band no.	f_c (kHz)	f_1 (kHz)	f_2 (kHz)
1	0.130	0.010	0.200
2	0.250	0.200	0.300
3	0.350	0.300	0.400
4	0.450	0.400	0.510
5	0.570	0.510	0.630
6	0.700	0.630	0.770
7	0.840	0.770	0.920
8	1.000	0.920	1.080
9	1.170	1.080	1.270
10	1.370	1.270	1.480
11	1.600	1.480	1.720
12	1.860	1.720	2.000
13	2.160	2.000	2.320
14	2.510	2.320	2.700
15	2.920	2.700	3.150
16	3.420	3.150	3.700
17	4.050	3.700	4.400
18	4.700	4.400	5.300



(a) Unprocessed

(b) Processed using spectral segment

Figure 2.5. Spectra of vowels /*a*/, /*i*/, /*u*/, and broad-band noise (100 ms segment): unprocessed and processed using spectral segment mapping, bandwidth: ACB, segmentation: fixed-frame, α = 0.6 [19].

The effectiveness of the processing in improving recognition of consonants was tested by conducting listening tests using modified rhyme test (MRT) on normal-hearing subjects in the presence of noise and on hearing-impaired subjects in quiet

[17]-[18]. Compression factor of 0.6 resulted in best performance. For normal-hearing subjects, there was an improvement of 17% in the recognition scores at lower SNR values, equivalent to SNR advantage of 6 dB. There was a mean decrease of 0.88 s in response time indicating a reduced perceptual load. For hearing-impaired subjects, there was 9−21% improvement in recognition scores (mean improvement = 16.5%) with a mean reduction of 0.89 s in response time, indicating the potential of the technique for its use in hearing aids for improving speech perception by persons with moderate sensorineural loss.

2.3. Multi-band Frequency Compression with Pitch-synchronous Segmentation

Speech output from fixed-frame analysis-synthesis has perceptible distortions. Listening test with normal-hearing listeners showed that the scores for the processed speech were higher than that for the unprocessed speech only in the presence of masking noise, indicating that perceptible distortions adversely affected the advantage of frequency compression. Use of pitch-synchronous processing removed the distortion [17] and its evaluation on hearing-impaired subjects showed a significant improvement in speech perception [18]. In a processing involving modification of short-time Fourier transform (STFT), the resulting spectrum, in general, may not be a valid STFT in the sense that it cannot be associated with a time-domain sequence. In practical terms, it happens because of discontinuities between the segments corresponding to the consecutive modified complex spectra. Use of 50% overlap-add in the fixed-frame processing helps in masking the discontinuity. But in case of voiced speech it leads to a processing artifact in the form of another superimposed pitch related with the shift interval used for the analysis window. Pitch-synchronous processing with window length of two local pitch periods and overlap of one pitch period avoids this problem.

However, pitch-synchronous processing is not very suitable for real-time operation due to the algorithmic and computational delays associated with it. The frame length and the number of overlapping samples vary according to the estimated pitch period during the processing. Thus the number of frames contributing to processed output samples varies leading to a sudden amplitude changes in the output. When non-speech sounds are processed using pitch-synchronous processing,

amplitude modulations are observed due to errors in the estimated pitch-period duration.

2.4. Modified Multi-band Frequency Compression

To avoid the artifact associated with the fixed-frame processing with 50% overlap and the additional algorithmic and computational delay associated with pitch estimation for the pitch-synchronous processing, we have used the Griffin-Lim method [26] of signal estimation from modified short-time complex spectrum. The method is based on least squared error estimation (LSEE), i.e. minimizing the mean squared error between STFT of the estimated signal and the modified STFT. The output signal is resynthesized by overlap-add of the partial sequences obtained as IDFT of the modified complex spectrum after multiplication with the analysis window. The window used should meet the requirement that sum of the squares of all the windows is unity, i.e.

$$\sum_{m=-\infty}^{\infty} w^2(mS - n) = 1 \tag{2.6}$$

For window length L and window shift $S = L/4$ corresponding to 75% overlap, this requirement is met by modified Hamming window, given as

$$w(n) = [\,1/\sqrt{4p^2 + 2q^2}\,][\,p + q\cos(2\pi(n + 0.5)/L)\,] \tag{2.7}$$

with $p = 0.54$ and $q = -0.46$. Griffin and Lim extended the method to reconstruct a signal from the modified short-time magnitude spectrum by using an iterative technique. Subsequently, other methods for signal estimation from modified short-time spectrum and suited for real-time processing have been reported [27]-[29]. Since the spectral modification using multi-band frequency compression results in complex spectra, it can be easily implemented using LSEE method of Griffin and Lim using modified Hamming window and 75% overlap. This implementation is subsequently referred to as LSEE processing. The methods for a comparison of the pitch-synchronous and LSEE processing were implemented using Matlab. Speech outputs from the two methods were found to be perceptually indistinguishable. It was observed that LSEE processing works equally well for speech, speech with noise, and music signals.

Wide-band spectrograms of the processed and the unprocessed signal for the sentence “where were you a year ago?” are shown in Figure 2.6. The spectrograms of

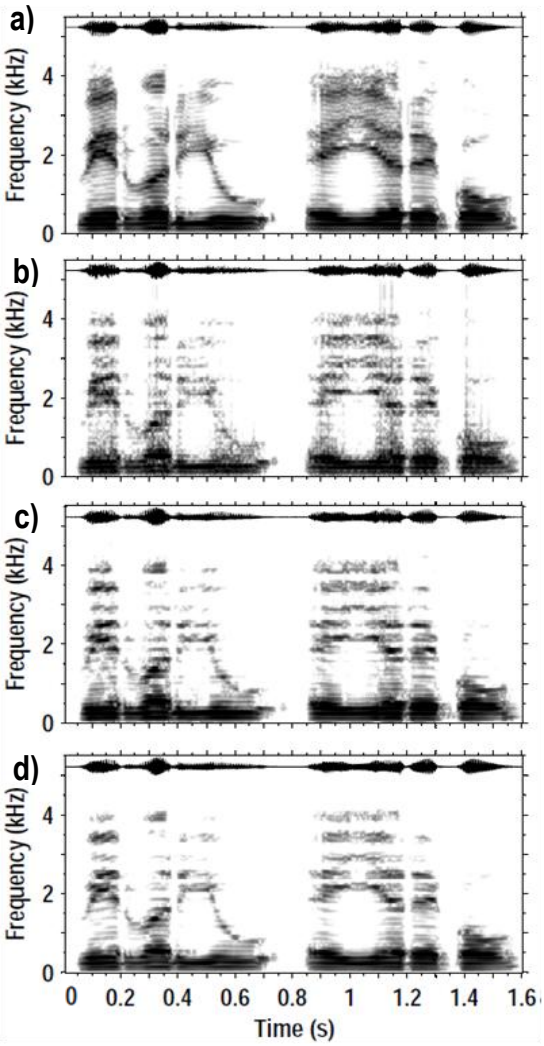


Figure 2.6*:* Comparison of analysis-synthesis methods: wide-band spectrograms of the sentence "where were you a year ago?": (a) unprocessed, (b) FF, (c) PS, and (d) LSEE. Processing with spectral segment mapping, auditory critical bandwidth, $\alpha = 0.6$.

processed speech show the concentration of spectral energy into narrower bands. The vertical striations indicate that the harmonic structure is approximately preserved. It is observed that processing retains the formant transitions with only slight shift in the formant locations. Because of compressing of the spectral components towards the band centers, the harmonics in different bands may not necessarily remain harmonically related. The effect is not noticeable for compression factor of 0.6, but it becomes perceptually noticeable at 0.4 and became more pronounced at lower values of the compression factor [17]. Processing artifact in the form of discontinuities is observed in the spectrogram of the output from fixed-frame processing, but not observed in the pitch-synchronous and LSEE outputs.

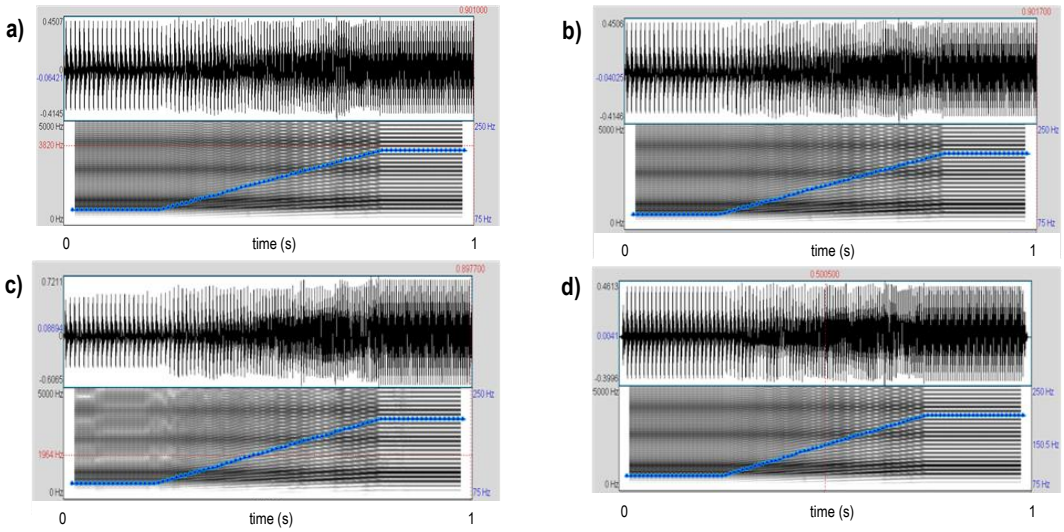


Figure 2.7. Comparison of analysis-synthesis methods: time domain waveforms and wide-band spectrograms of vowel synthesized with constant amplitude and 100 – 200 Hz pitch: (a) unprocessed, (b) FF, (c) PS, and (d) LSEE. Processing with spectral segment mapping, auditory critical bandwidth, $\alpha = 1$.

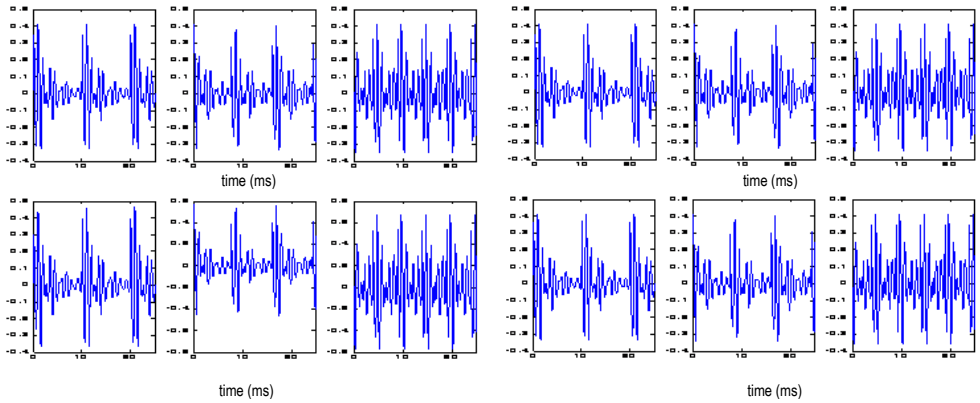


Figure 2.8. Comparison of analysis-synthesis methods: zoomed view of 25 ms segments of time domain waveforms of synthesized vowel /*a*/ in Figure 2.7. (a) unprocessed, (b) FF, (c) PS, and (d) LSEE. Processing with spectral segment mapping, auditory critical bandwidth, $\alpha = 1$. Vowel segments taken for F0 = 100 Hz, varying F0, and F0 = 200 Hz.

Vowel /*a*/ was synthesized with constant amplitude and 100 Hz – 200 Hz pitch and processed using different processing schemes for comparison of different analysis-synthesis methods. The results are shown in Figure 2.7. The effect of different types of processing on vowel segments with steady pitch of 100 Hz, segment with varying pitch, and steady state pitch of 200 Hz can be seen more clearly from zoomed view of 25 ms segments of time domain waveforms of synthesized vowel /*a*/ as shown in Figure 2.8. It is observed that the processed output obtained from LSEE method, when input pitch is constant, is similar to that obtained from PS method. However, when input pitch is varying the output obtained from PS method shows

amplitude variations unlike FF and LSEE methods. These results show that LSEE method is suitable for the segments with constant pitch as well as for the segments with variable pitch.

Chapter 3

INVESTIGATIONS ON OFFLINE IMPLEMENTATION

3.1 Introduction

The chapter presents a comparison between results obtained from multi-band frequency compression scheme implemented in Matlab using fixed-frame segmentation, pitch-synchronous segmentation, and fixed-frame segmentation based LSEE processing. The perceptual quality of processed speech is compared by informal listening tests. Investigations to compare the root mean square (RMS) values of the results obtained from fixed-frame segmentation based LSEE method for different types of input are presented in this chapter.

3.2 Fixed-frame segmentation based multi-band frequency compression

Speech output from fixed-frame analysis-synthesis has perceptible distortions which adversely affect the advantage of multi-band frequency compression. The distortions occur due to discontinuities which remain unmasked even after overlap-add operation for reconstruction. In case of voiced speech, it leads to a processing artifact in the form of another superimposed pitch related to the shift interval used for the analysis window. The spectrograms of the outputs obtained from fixed-frame segmentation based multi-band frequency compression for broad-band Gaussian noise input, swept tone input, and vowels */a/*, */i/*, and */u/* are shown in Figure 3.1 for compression factor of 1, 0.8, and 0.6. With compression factor set as one, the output obtained is same as the input. For lower compression factors, the spectral energy is concentrated into narrower bands as seen in the spectrograms of processed output.

Separation between the bands increases with increase in amount of compression. The frequency of the swept tone input varies between 500 Hz and 5 kHz in time duration of 2 s. When the compression factor is set as one, the spectrogram of the processed output is same as that of input swept tone, indicating absence of spectral

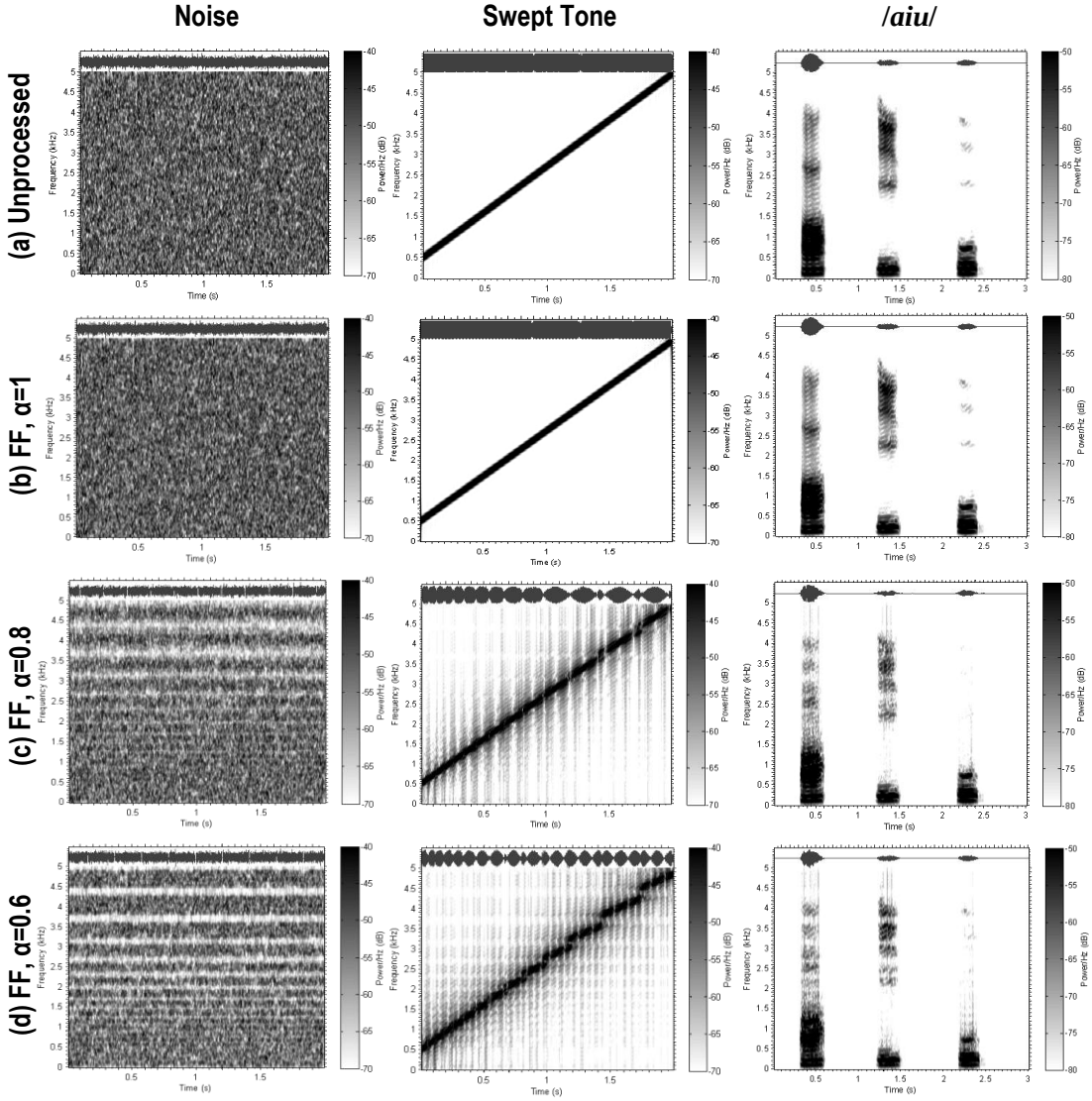


Figure 3.1. Processing using FF: S\spectrograms of noise, swept tone, and vowels */aiu/*: a) unprocessed, b) processed with α = 1, c) processed with α = 0.8 and d) processed with α = 0.6. Window length = 20 ms, overlap = 50 %.

compression. A staircase like structure is seen in the output spectrograms for lower compression factors. This occurs because each frequency in the input swept tone is mapped to a frequency closer to the center frequency of a particular auditory critical band. The processed output can have frequency in one of these bands and thus the processed spectrogram has a discrete staircase like structure. The spectrograms of the processed output obtained for sentence “Where were you a year ago?” are shown in Figure 3.2. With a compression factor of one, the output spectrum remains unchanged. For lower compression factors, it is concentrated into narrower bands.

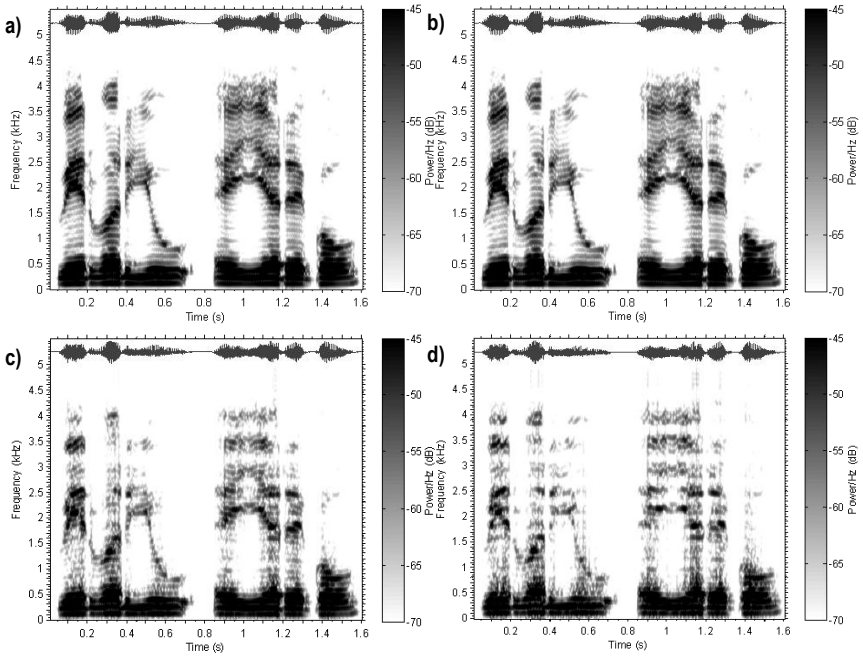


Figure 3.2. Processing using FF: spectrograms of sentence "where were you a year ago": a) unprocessed, b) processed with α = 1, c) processed with α = 0.8 and d) processed with α = 0.6. Window length = 20 ms, overlap = 50 %.

3.3 Pitch-synchronous segmentation based multi-band frequency compression

In fixed-frame processing, reconstruction using overlap-add helps in masking the discontinuities. But in case of voiced speech, it leads to a processing artifact in the form of another superimposed pitch related to the shift interval used for the analysis window. Pitch-synchronous processing with a window length of two local pitch-periods and overlap of one pitch period avoids this problem as the discontinuities do not occur at the frame boundaries.

The spectrograms of the outputs obtained from pitch-synchronous segmentation based multi-band frequency compression for broad-band Gaussian noise input, vowels */a/*, */i/*, and */u/*, and the sentence "Where were you a year ago ?" are shown in Figure 3.3 for compression factor of 1, 0.8, and 0.6. With compression factor set as one, the output obtained is same as the input showing that the output spectrum is uncompressed. The spectrograms of processed output show the concentration of spectral energy into narrower bands. As the amount of compression increases, the separation between the bands also increases. The noise input is preceded by a small segment of voiced speech, so that pitch period estimated during

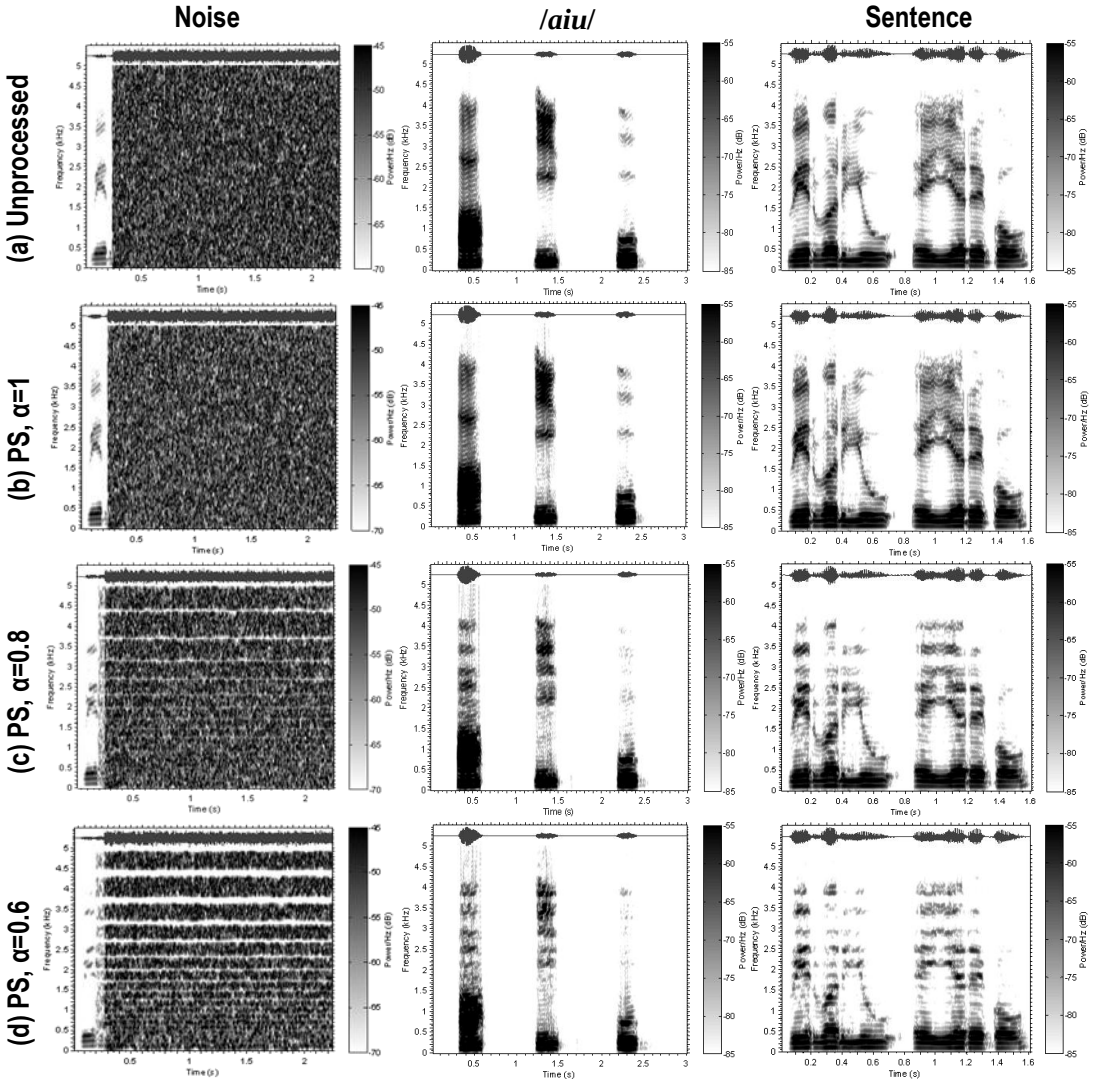


Figure 3.3. Processing using PS: spectrograms of noise, vowels */aiu/*, and sentence "Where were you a year ago ?": a) unprocessed, b) processed with α = 1, c) processed with α = 0.8 and d) processed with α = 0.6.

this segment may be used in the processing for multi-band frequency compression of noise. The spectrograms obtained as a result of processing the sentence "Where were you a year ago?" and vowels */a/*, */i/*, and */u/* show that the output remains uncompressed for a compression factor of one and the spectral energy is concentrated into narrower bands for lower compression factors. Figure 3.4 shows spectrograms for music a clip. It is observed that processing introduces distortions in the output even with compression factor of one due to inability to estimate pitch correctly. In musical sounds there are multiple sources with different pitches which are not correctly tracked by the GCI detection used in PS method.

The pitch estimation by Childers-Hu algorithm for GCI detection [22] used in PS processing was compared with that obtained by Praat []. Figure 3.5 shows the

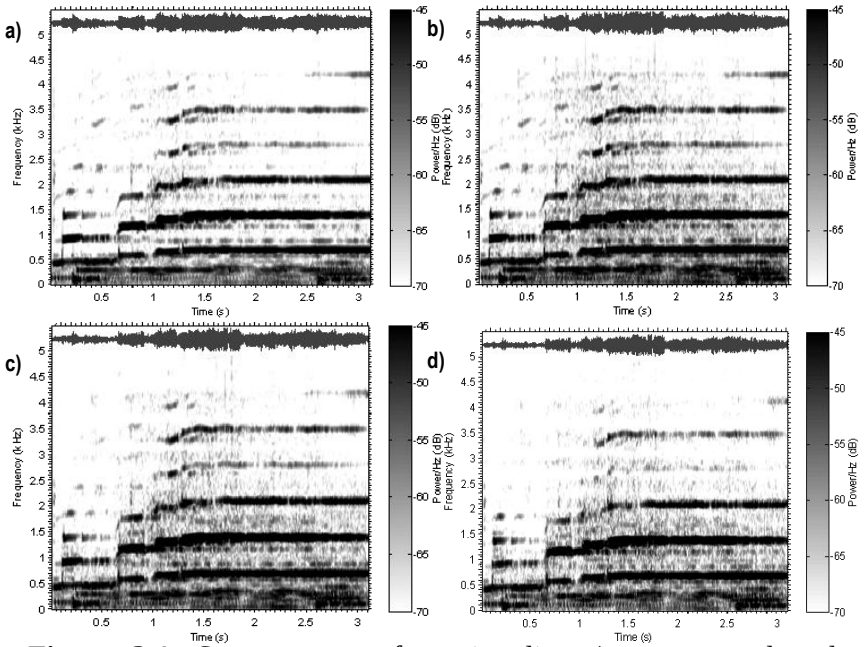


Figure 3.4. Spectrograms of music clip: a) unprocessed and processed using PS with b) α = 1, c) α = 0.8 and d) α = 0.6.

results for music clip and the sentence "Where were you a year ago?" at different values of SNR (with additive broad band noise). It is observed that the pitch values estimated by the two methods are approximately same for noise-free signal. However, when noise is added to the speech signal, the pitch values estimated by the two methods are different and the difference becomes more noticeable at lower SNR values. It is seen that Childers-Hu algorithm does not estimate the pitch values correctly for music clip due to presence of multiple excitation sources. For Gaussian white noise preceded by a small voiced speech segment as the input, the pitch values estimated by Childers- Hu algorithm for non-speech segments are equal the to pitch value estimated in the last voiced segment.

3.4 Multi-band frequency compression using LSEE method

To avoid the artifact associated with the fixed-frame processing with 50 % overlap and the additional algorithmic and computational delay associated with pitch estimation for the pitch-synchronous processing, Griffin-Lim method [26] of signal estimation from modified short-time complex spectrum was used.

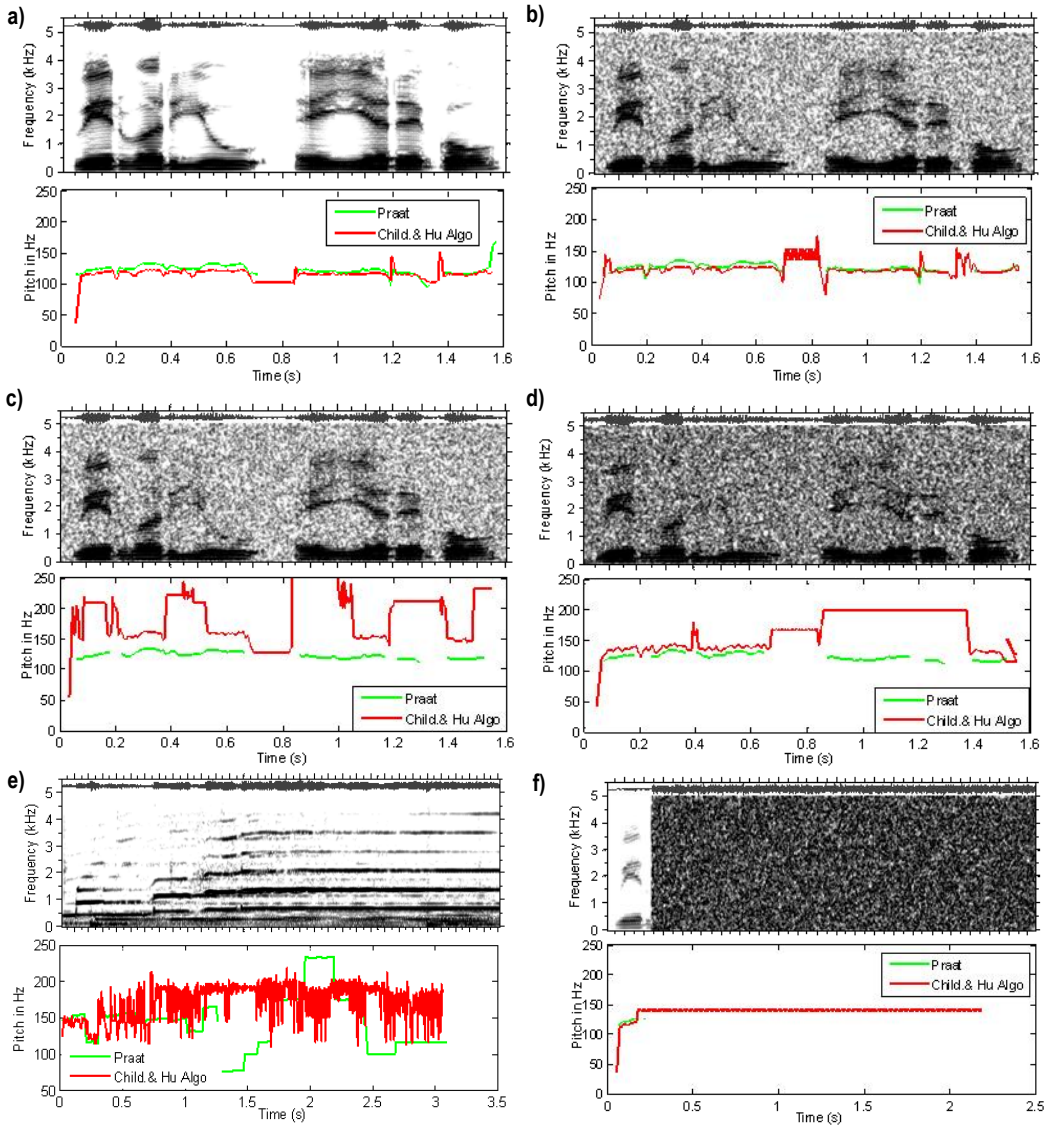


Figure 3.5. A comparison of pitch estimated using Childers-Hu algorithm and Praat for a) sentence "Where were you a year ago?", b) sentence with SNR = 20 dB, c) sentence with SNR = 10 dB, d) sentence with SNR = 5 dB, e) music clip, and f) white noise .

The spectrograms of the outputs obtained from LSEE for broad-band Gaussian noise input, vowels */a i u/*, and sentence "Where were you a year ago?"are shown in Figure 3.6 for compression factor of 1, 0.8, and 0.6. The spectrograms of processed output show concentration of spectral energy into narrower bands. As the amount of compression increases, the separation between the bands also increases. PESQ-MOS was calculated for quantifying the similarity between speech output from PS and LSEE processing as given in Table 3.1, the scores were 3.7−4.3 for compression factors of 0.6−1.0. Figure 3.7 shows spectrograms of processed and unprocessed music clip obtained from LSEE processing. The output has a better perceptual quality

Table 3.1. PESQ-MOS scores between offline and real-time outputs for vowel /*a i u*/ and sentence "Where were you a year ago?".

Input signal	Compression Factor α=1	α=0.8	α=0.6
/*a i u*/	4.3	3.9	3.7
Sentence	4.3	3.8	3.7

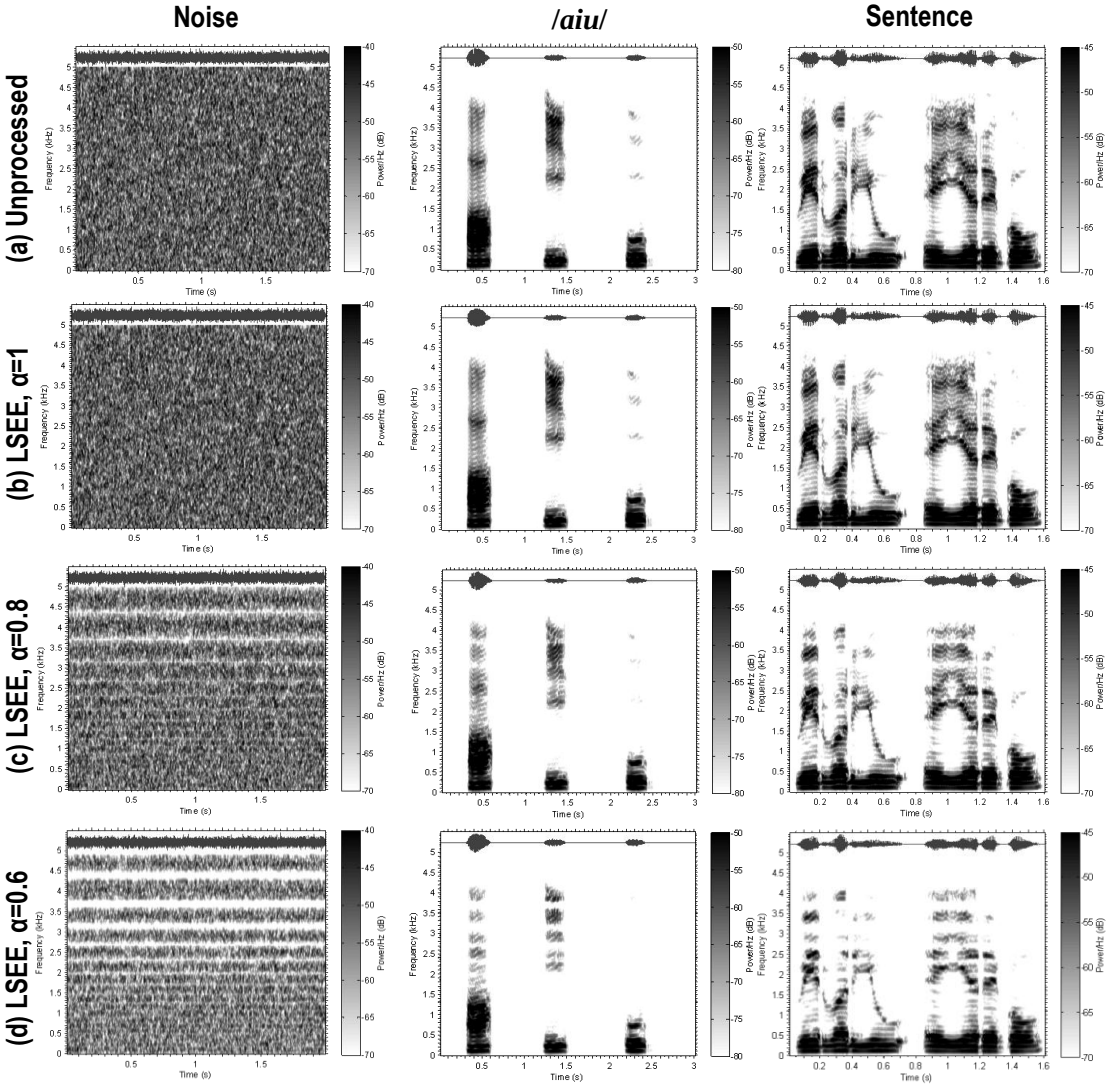


Figure 3.6. LSEE processing with window length of 26 ms: spectrograms of noise, vowels */aiu/*, and sentence "Where were you a year ago?": a) unprocessed b) processed with α = 1, c) processed with α = 0.8 and d) processed with α = 0.6.

than that obtained from pitch-synchronous implementation.

Multi-band frequency compression results in a change in RMS values of the signal and the change depends on the length of analysis window. A plot of the root mean square values of the output obtained from LSEE method for different window

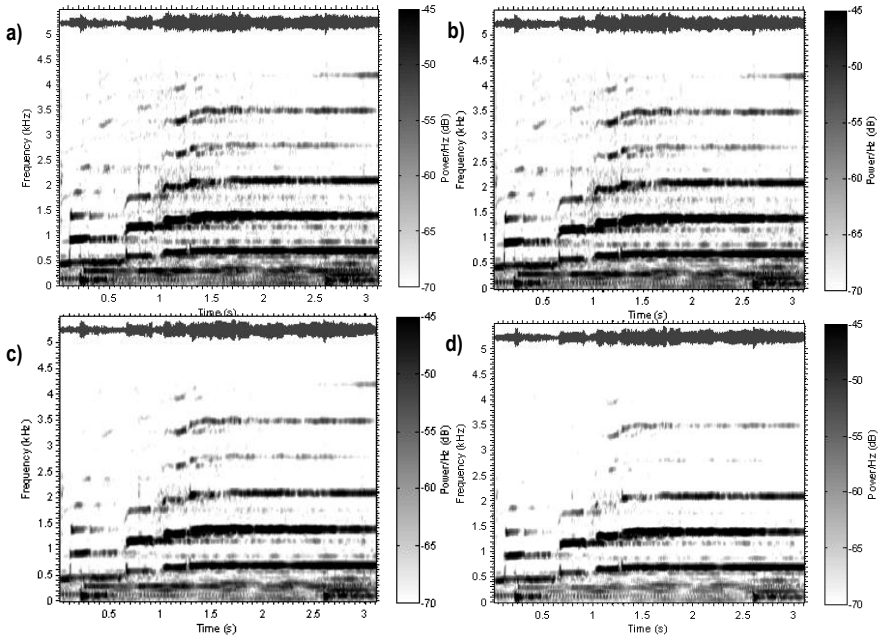


Figure 3.7. LSEE processing using window length of 26 ms: spectrograms of music clip: a) unprocessed, b) processed with α = 1, c) processed with α = 0.8 and d) processed with α = 0.6 using.

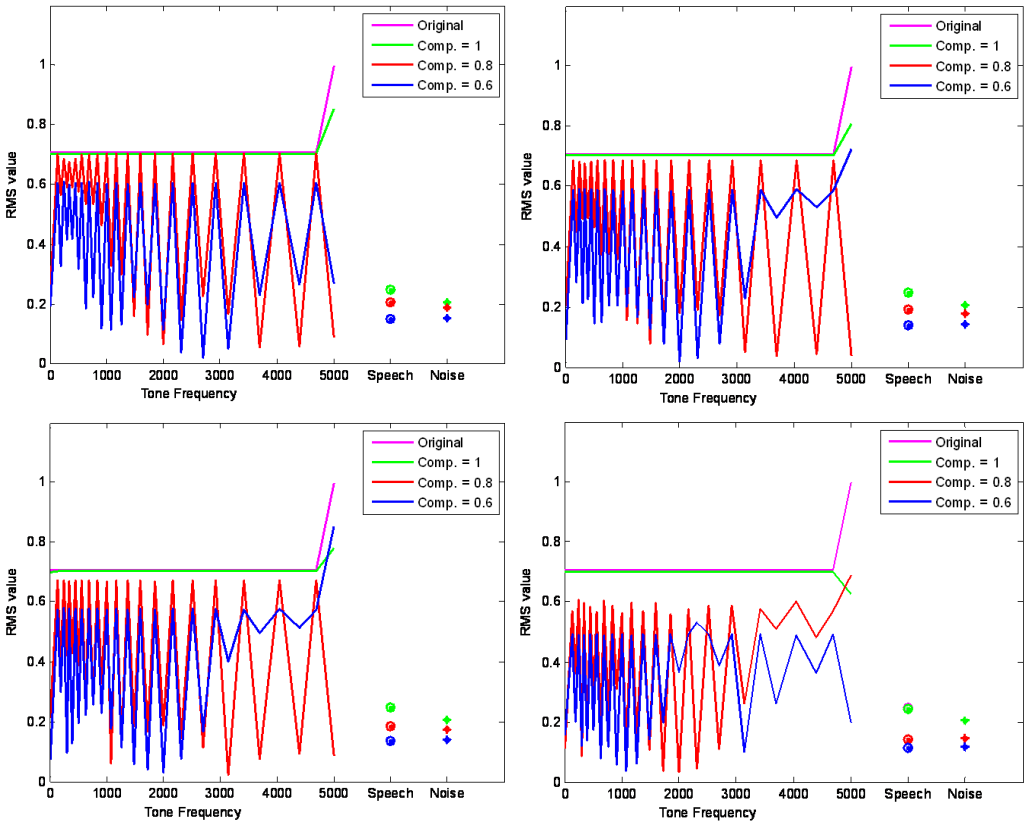


Figure 3.8. Plots of the RMS values of the input and output with compression factor α=1, 0.8, 0.6 and window lengths of a) 20 ms, b) 26 ms, c) 30 ms, d) 51.2 ms.

lengths is given in Figure 3.8. The plots are obtained for sinusoidal input with frequencies equal to the edge and the center frequencies of the auditory critical bands, broad-band Gaussian noise, and the sentence “Where were you a year ago?” as input, for compression factor of 1, 0.8, and 0.6. It is observed that RMS value of the center frequencies of auditory critical bands remains comparable to their original RMS values whereas the RMS value of the edge frequencies of auditory critical bands decreases considerably. There is a decrease in the RMS values of the output with decrease in compression factor for a given window length. For a compression factor less than one, the RMS value of the output decreases with increase in the window length. The window length chosen for implementation should be short enough to capture the dynamic changes in vocal tract shape, thus it should be less than 30 ms. However if the window length is too short, the spectrum will not have adequate resolution and hence the window length should be at least two to three pitch-periods long. Thus we need a minimum of 20 ms window for pitch of 100 Hz. Hence a window length of 26 ms (260 samples at 10 kHz sampling frequency) was selected for real-time implementation. The RMS values of the output corresponding to compression factors of 0.8 and 0.6 are 0.8 and 0.7 times the original RMS values respectively.

Effect of processing was also studied by examining the probability density function (PDF) and cumulative distribution function (CDF) of the time domain signal. Figure 3.9 shows the plots for different compression factors for speech, Gaussian noise, and sine waves of different frequencies. The PDF plots show the distribution of amplitude values as the compression factor changes. As the compression factor decreases the PDF plots become narrower and peaky indicating higher probability of lower amplitudes and thus a decrease in RMS value. The CDFs show that with decrease in compression factor the saturation occurs at a much lower amplitude value.

To reduce the computational load, the FFT length N was reduced from 1024 to 512 and its effect was observed on compression, RMS values, and amplitude distribution function of the processed output obtained from Matlab based offline implementation. The output obtained using $N = 512$ was indistinguishable from that obtained using $N = 1024$. Figure 3.10 shows spectrograms of the output obtained from LSEE method of multi-band frequency compression. The RMS values of the output for sine waves with frequencies equal to the edge and the center frequencies of the

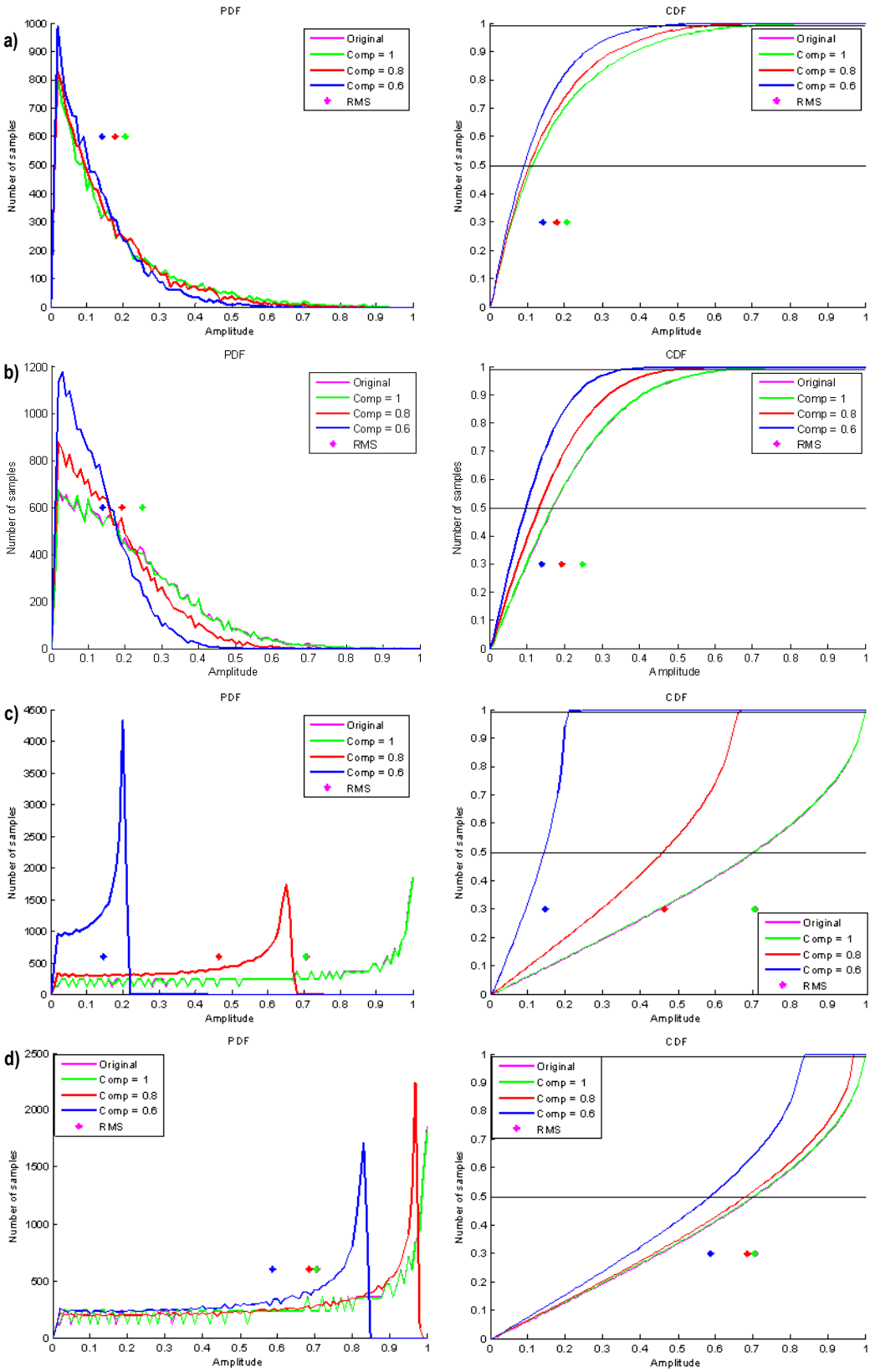


Figure 3.9. LSEE processing with window length = 260 and FFT length = 1024: PDF and CDF plots for compression factors of 1, 0.8, and 0.6 for a) sentence "Where were you a year ago?", b) broad-band Gaussian noise, c) sine wave of 510 Hz, and d) sine wave of 570 Hz.

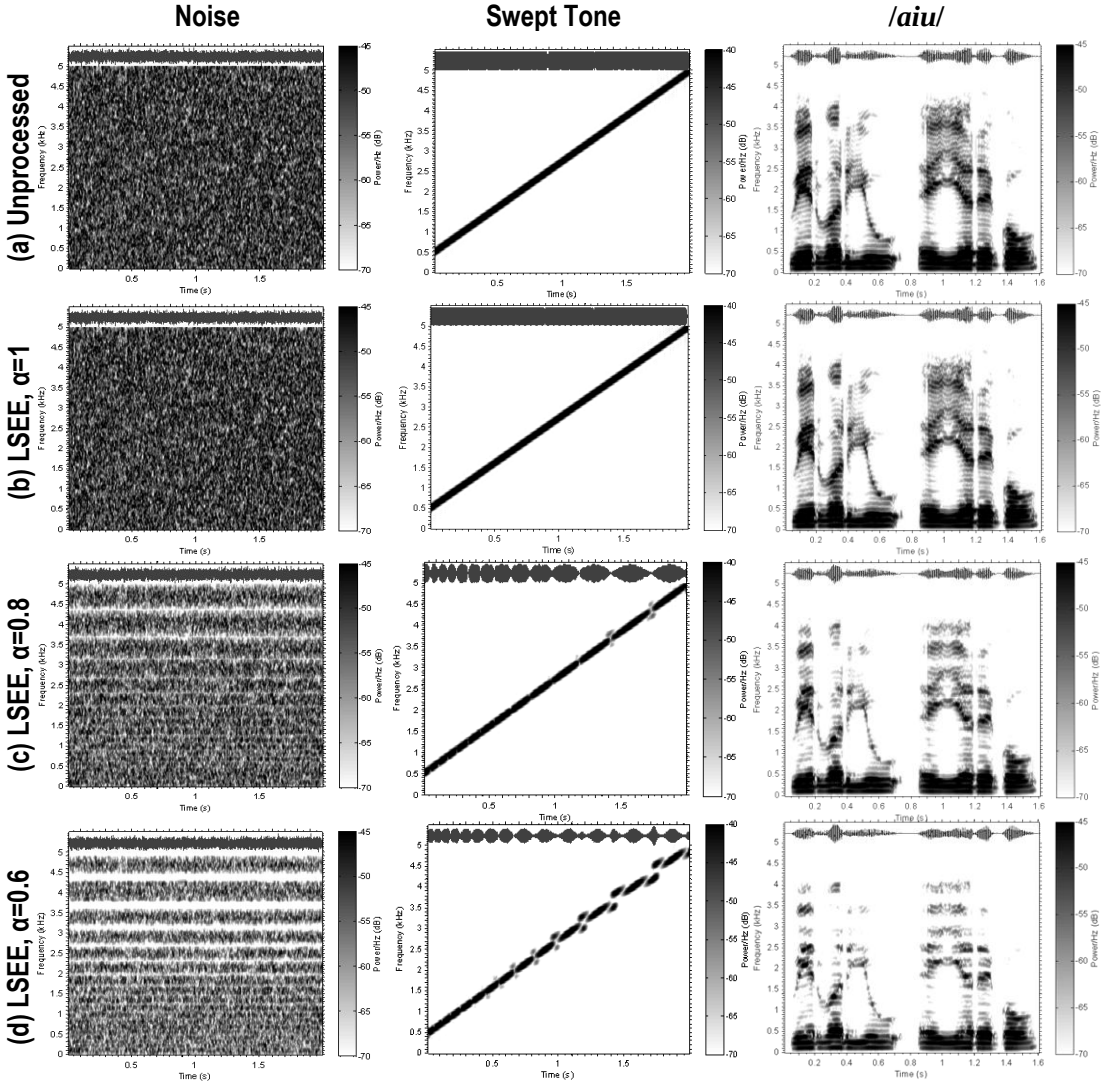


Figure 3.10. LSEE processing with window length of 26 ms and FFT length = 1024: spectrograms of noise, swept tone, and sentence "Where were you a year ago": a) unprocessed b) processed with α = 1, c) processed with α = 0.8 and d) processed with α = 0.6.

auditory critical bands, broad-band Gaussian noise, and the sentence "Where were you a year ago?" as input are shown for compression factor of 1, 0.8, and 0.6 along with the original RMS values are shown in Figure 3.11. The histograms or PDF plots of amplitudes levels of speech signal, Gaussian noise and the corresponding CDF plots for different compression factors are shown in Figure 3.12.

3.5 Summary

Investigation showed that multi-band frequency compression implemented using LSEE based analysis-synthesis produces better perceptual quality of the processed output than fixed-frame analysis-synthesis and has lower additional algorithmic and

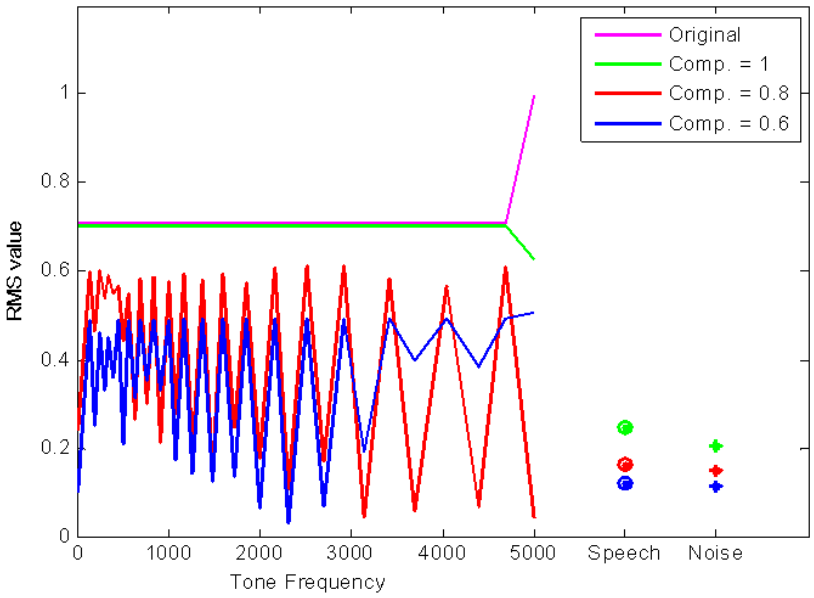


Figure 3.11. Comparison of RMS values for window length = 260 samples and FFT length = 512 samples.

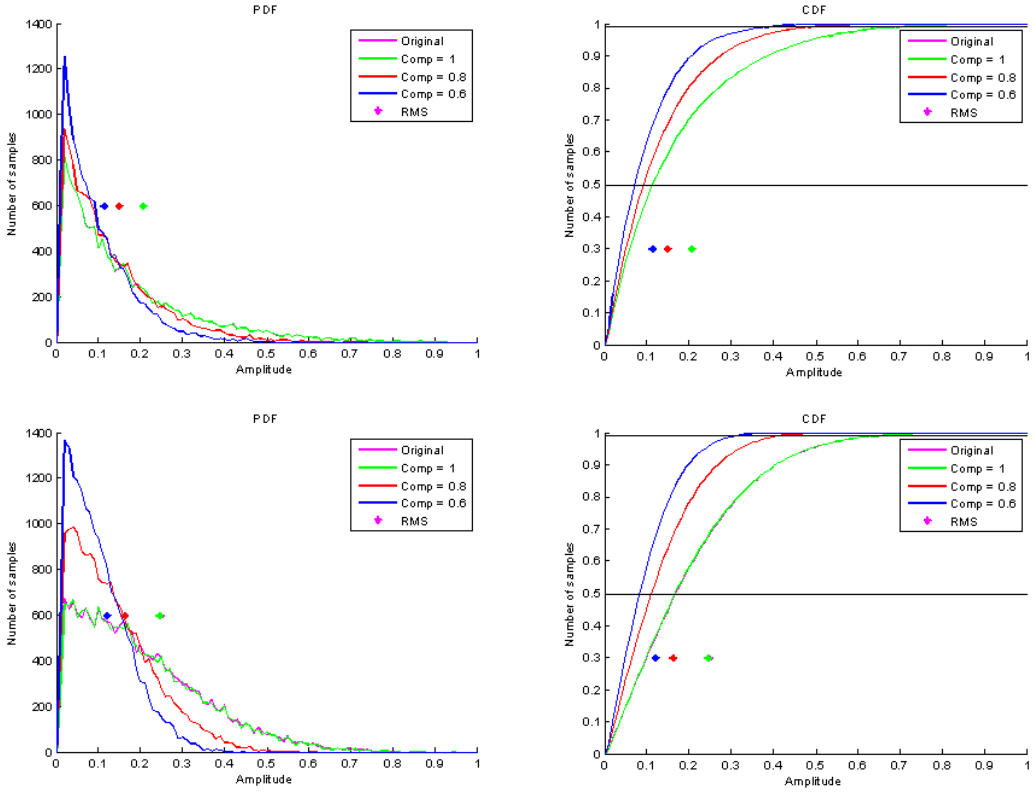


Figure 3.12. LSEE processing with window length of 26 ms ans FFT length = 512: PDF and CDF plots for compression factors of 1, 0.8, and 0.6 for a) sentence "Where were you a year ago?", b) broad-band Gaussian noise.

computational delays as compared to pitch-synchronous analysis-synthesis. For a comparison of the pitch-synchronous and LSEE processing, both were implemented using MATLAB. The speech outputs from the two implementations were perceptually

similar and the PESQ-MOS was found to be 3.7. LSEE processing is found to be well suited for non-speech audio also, as it does not require pitch estimation. Therefore it is decided to use it for real-time implementation. Decrease in processing time and computational load by using $N = 512$ instead of $N = 1024$ as used in [18] may help in combining multi-band frequency compression with other processing techniques for use in hearing aids.

Chapter 4

REAL-TIME IMPLEMENTATION

4.1 Introduction

The scheme of multi-band frequency compression is implemented for real-time processing on DSP board "eZdsp", based on 16-bit fixed-point processor TI/TMS320C5515 [30]. The processor can be operated at a clock frequency of up to 120 MHz, with option of dynamic switching between internal and external sources. It has a unified memory map with total address space of 16 MB. The main on-chip features include 320 KB RAM (with 64 KB dual access data RAM), 128 KB ROM, four 4-channel DMA controllers, three 32-bit timers, FFT hardware accelerator tightly coupled to CPU for efficiently computing 8 to 1024-point complex as well as real-valued FFT. The complex numbers are stored using 4-byte words in data memory, with each word holding the 16-bit real and 16-bit imaginary parts. The board has 4 MB on-board NOR flash for user program. Its on-board codec TLV320AIC3204 [31] has stereo ADC and DAC, with 16/20/24/32-bit quantization and sampling rate of 8 – 192 kHz. The block diagram of TMS320C5515 eZdsp USB Stick is shown in Figure 4.1.

4.2 Implementation Details

The block diagram of the implementation of the multi-band frequency compression scheme on the DSP board is shown in Figure 4.2. Codec and DMA are used to continuously acquire and output the speech signal. For reducing conversion overheads, the input samples, spectral values, and the processed samples are all stored as 4-byte words, with 16-bit real and 16-bit imaginary parts. The imaginary part of input sample is set to zero.

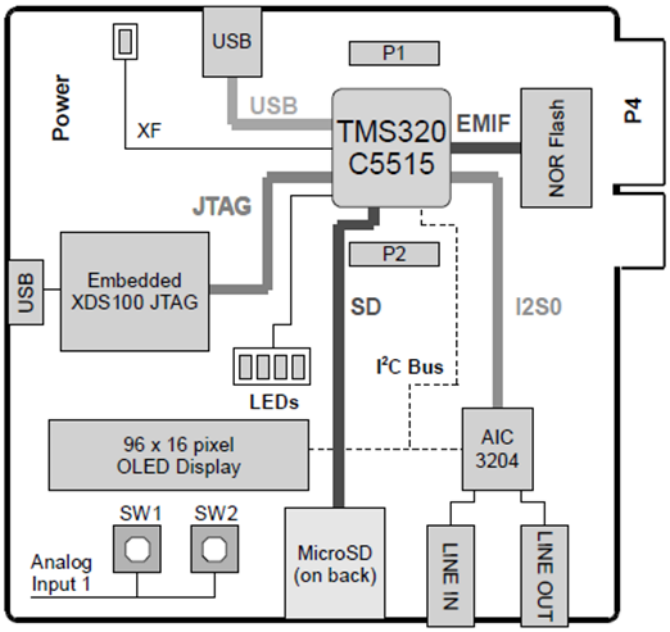


Figure.4.1. Block diagram of TMS320C5515 eZdsp USB Stick [30].

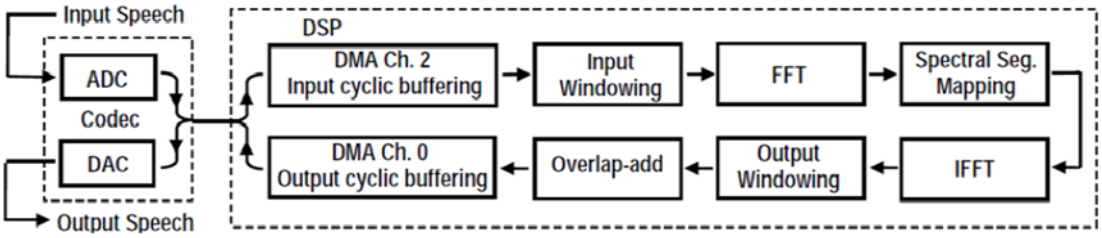


Figure 4.2. Block diagram of implementation of multi-band frequency compression on the DSP board

The data transfer and buffering operations are shown in Figure 4.3. These are devised for an efficient realization of analysis-synthesis with 75% overlap, with window length *L* and FFT size N. Signal acquisition uses a 5-block DMA input cyclic buffer, with each block of *S* words. An input data buffer of *N* words, initialized with zero values, serves as the input array for FFT computation. At the regular intervals set by the sampling rate, DMA channel-2 reads the input sample values from ADC and writes them in the DMA input cyclic buffer, with the 16-bit input as the real-part of the 32-bit word The output is handled using a 2-block cyclic buffer, with each block of *S* words. DMA channel-0 is used to cyclically output 16-bit real-part of the 32-bit word. The output is handled using a 2-block cyclic buffer, with each block of *S* words. DMA channel-0 is used to cyclically output to DAC. Pointers are used to keep track of the current input, just-filled input, current output, and write-to output blocks, and these are initialized to 0, 4, 0, and 1, respectively. When a block gets filled, a DMA interrupt is generated. All the four block pointers are incremented cyclically. The

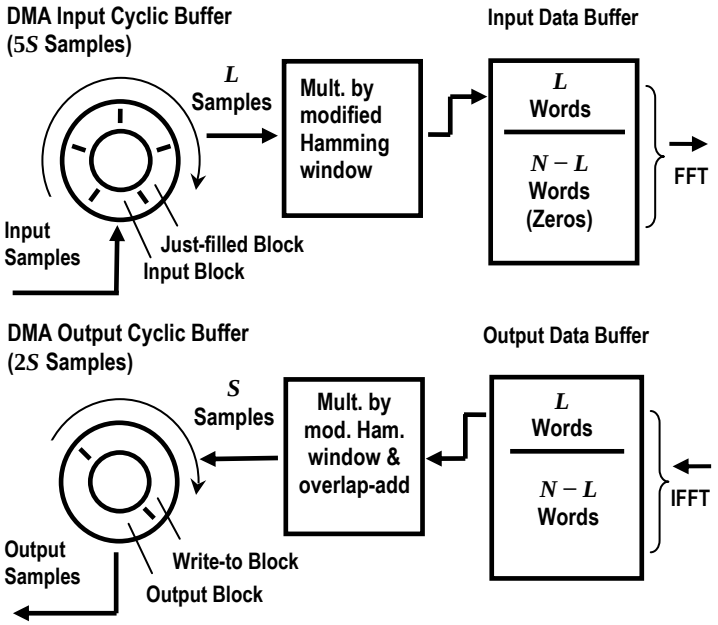


Figure 4.3. Data transfer and buffering operations ($S = L/4$).

DMA mediated reading from ADC into the current input block and writing from the current output block to DAC are continued. The samples of the just-filled block and the previous three blocks are copied into the input data buffer. These samples are then multiplied by modified Hamming window of length L as given in equation (2.7). They are padded with $N-L$ zero-valued samples to serve as the input array to N-point FFT. This method of copying from the DMA input cyclic buffer to the input data buffer results in an efficient realization of 75% overlap and zero padding.

The DFT of a real-valued discrete-time signal is conjugate-symmetric. This property can be used in reducing the computations involved in processing the input signal. The first $N/2$ output complex spectral samples are obtained from processing and the last $N/2$ complex spectral samples can be taken as complex-conjugate of mirror image. Another approach may be used to further reduce the computations. Let $x(n)$ be a discrete-time real-valued sequence with length N and $X(k)$ be its N-point DFT. Let $x'(n)$ be the discrete-time complex sequence obtained by taking N-point IDFT of sequence formed by padding first $N/2$ complex spectral samples with $N/2$ zero-valued samples. Then it can be shown that

$$x(n) = 2\,\mathrm{Re}\,[x'(n)] \qquad n = 0, 1, \ldots N\text{-}1 \qquad (4.1)$$

Thus a discrete time real-valued signal can be reconstructed from its first $N/2$ complex spectral samples padded with $N/2$ zero-valued samples. The FFT hardware accelerator of the processor is used to calculate N-point complex DFT of the input array and the result is stored in an N-word buffer. The first $N/2$ spectral samples of the compressed

spectrum are calculated using spectral segment mapping as given in (2.5), using a look-up table of pre-calculated values of m, n, $m-a$, and $b-n$ for each output spectral index. The other samples are kept zero-valued. The FFT hardware accelerator is used to calculate N-point complex IDFT of the modified complex spectrum. Real part of the first L samples of the resulting sequence is multiplied by twice the modified Hamming window and stored in the output data buffer as partial outputs. Overlap-add operation uses a buffer of 3S samples. The first S samples of the output data buffer are added to the first S samples of the overlap buffer containing the partial results from the previous operation. The resulting samples are written as the processed output to the write-to output block. The next 2S samples of the output data buffer and the overlap buffer are added together and copied as the first 2S samples of the overlap buffer. The last S samples of the output data buffer are copied as the last S samples of the overlap buffer. For real-time processing, all the operations on the samples in the input data buffer should get completed before generation of the next DMA interrupt, i.e. in less than the time corresponding to S samples. The processing has an algorithmic delay of L samples (4S samples) and computational delay of less than L/4 samples (S samples). Thus the total delay between input and the processed output is the time required to input 5S samples or time required to completely fill the DMA input cyclic buffer.

4.3 Software

The program was written in C, using TI's 'CCStudio, ver. 4.0' as the development environment. The sampling rate is selected as 10 kHz, and only one channel of the stereo codec is used with 16-bit quantization. The codec ADC has a programmable gain amplifier whose gain is set to 0 dB. The input data buffer length N and window length L were set as 1024 words and 260 words respectively. The FFT length was set equal to 1024-points. The processor was set to run on internal clock of 120 MHz. For reducing conversion overheads, the input samples, spectral values, and the processed samples are all stored as 4-byte words, with 16-bit real and 16-bit imaginary parts. The imaginary part of input sample is set to zero.

The input samples acquired by codec ADC are copied to blocks "`RcvL1`", "`RcvL2`", "`RcvL3`", "`RcvL4`", and "`RcvL5`" of DMA cyclic buffer using channel-2 of DMA0. The samples in four consecutive blocks from DMA cyclic buffer are

multiplied by modified Hamming window and are stored in the input data buffer "`FilterIn`", as 16-bit real part of 32-bit number. The input data buffer is initialized with zero valued samples. FFT hardware accelerator is used to calculate 1024-point FFT of the samples stored in "`FilterIn`". The assembly language FFT routines are provided by TI in the file hwafft.asm [32] which can be used to find 8-point to 1024-point real as well as complex valued FFT. The input samples in "`FilterIn`" are stored in bit reversed order in a buffer named "`data_br_buf`". The complex spectral samples obtained as FFT of "`data_br_buf`" are stored in a 32-bit buffer called "`scratch_buf`". The multi-band frequency compression is applied on the first *N*/2 samples stored in "`scratch_buf`" and the *N*/2 compressed spectral samples padded with *N*/2 zero valued samples are stored in a buffer called "`convolved_buf`". The data stored in "`convolved_buf`" is bit reversed and stored in "`data_br`". The IFFT of the complex spectral samples stored in bit reversed data buffer is found and stored in buffer called "`FilterOut`". The real parts of the first *L* samples of "`FilterOut`" are multiplied with modified Hamming window and are used for resynthesis of output using overlap-add. The resulting samples are stored in buffers called "`Xmit1`" and "`Xmit2`". Since the spectral modifications are done on only first *N*/2 complex spectral samples and the last *N*/2 spectral samples are made zeros, the output samples in "`Xmit1`" and "`Xmit2`" should be doubled in magnitude. This can be done by either multiplying the first *L* samples of "`FilterOut`" with twice the modified Hamming window or by setting the programmable gain amplifier (PGA) gain of CODEC DAC to 2 (i.e. 6 dB). The second option is used in our implementation.

An underflow during the processing using lower value of compression factors may introduce distortion and reduce the output RMS. This can be avoided by pre-multiplying the complex spectral samples with a scale factor before applying the compression algorithm. This however will also amplify the processing noise. To avoid noise amplification the PGA gain of ADC may be set such that the dynamic range of ADC is used optimally, without causing input saturation or computation overflow.

4.4 Test results

For testing the implementation on the DSP board, the input signal for processing was generated from a PC sound card and given to the DSP board through one channel of

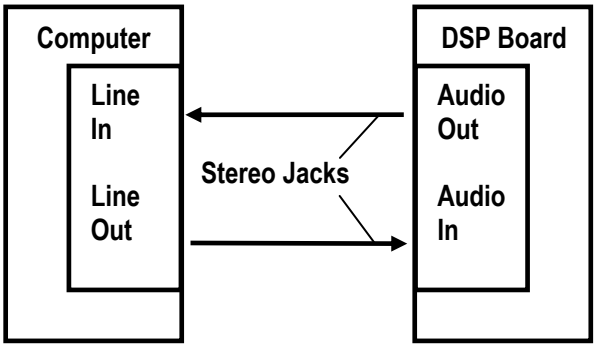


Figure 4.4. Setup for giving input and recording output from DSP board.

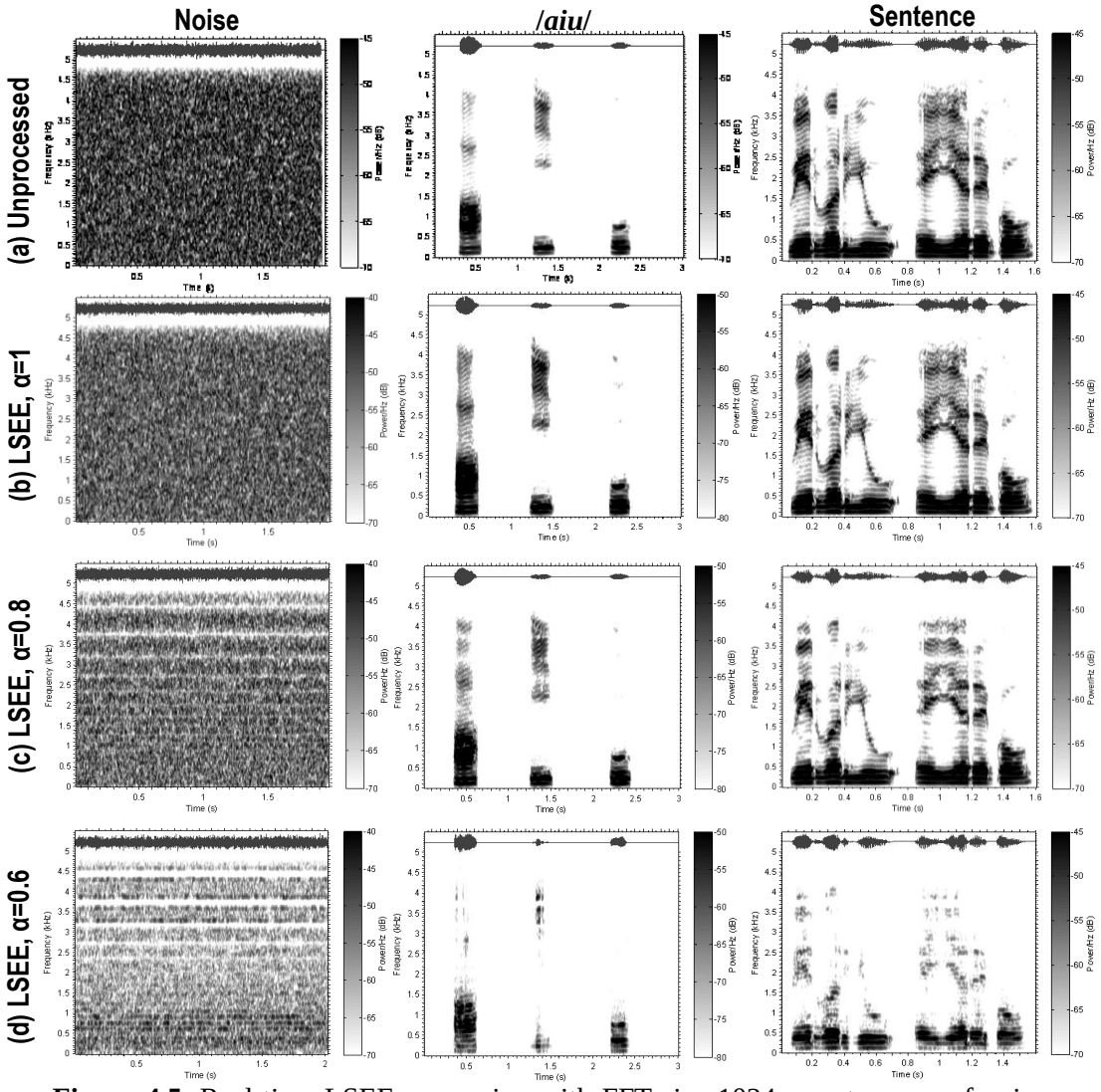


Figure 4.5. Real-time LSEE processing with FFT size 1024: spectrogram of noise vowels */aiu/*, and sentence "Where were you a year ago?": a) unprocessed, b) processed with $\alpha = 1$, c) processed with $\alpha = 0.8$ and d) processed with $\alpha = 0.6$.

its stereo-in audio connector. The processed output signal from one channel of stereo-out audio connector was acquired through the PC sound card as shown in Figure 4.4.

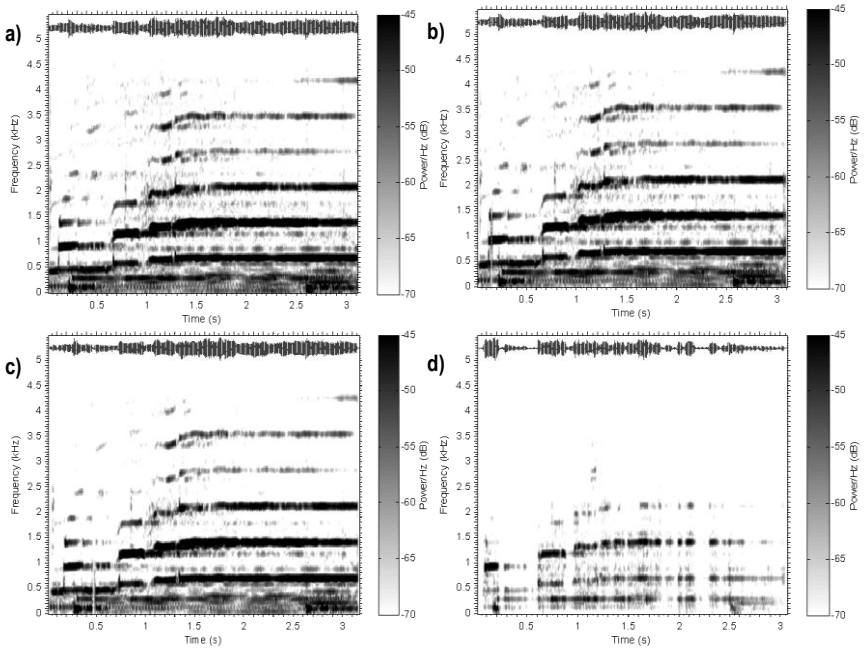


Figure 4.6. Real-time LSEE processing with FFT size = 1024: spectrograms of a music clip: a) unprocessed b) processed with α = 1, c) processed with α = 0.8 and d) processed with α = 0.6.

The signals were also acquired bypassing the processing for studying the effect of ADC and DAC of the board. The spectrogram of the outputs obtained from real-time implementation using LSEE method for broad-band Gaussian noise input, vowels */a i u/*, and the sentence "Where were you a year ago?" for compression factors of 1, 0.8, and 0.6 are shown in Figure 4.5. The spectrograms show that multi-band frequency compression concentrates the spectral energy into narrower bands. As the amount of compression increases, the separation between the bands also increases. With compression factor set as one, the output obtained is same as the input showing that the output spectrum is uncompressed. The spectrograms of the output obtained when a music clip is given as the input are shown in Figure 4.6. The processed output is perceptually indistinguishable from that obtained by the offline implementation. PESQ-MOS was calculated for quantifying the similarity between offline and real-time speech output as given in Table 4.1, the scores were 2.5–3.4 for compression factors of 0.6–1.0.

To reduce the computational load, the FFT length N was reduced to 512. The output was indistinguishable from that obtained using N = 1024. The results obtained using FFT length of 512 are similar to those obtained using FFT length of 1024. Spectrograms of broad-band Gaussian noise input, vowels */a i u/*, and the sentence "Where were you a year ago?" for compression factors of 1, 0.8, and 0.6 and FFT

Table 4.1. PESQ-MOS scores between offline and real-time outputs for vowel /*a i u*/ and sentence "Where were you a year ago?".

Input signal	Compression Factor α=1	α=0.8	α=0.6
/*a i u*/	3.4	3.6	3.0
Sentence	3.4	3.8	2.5

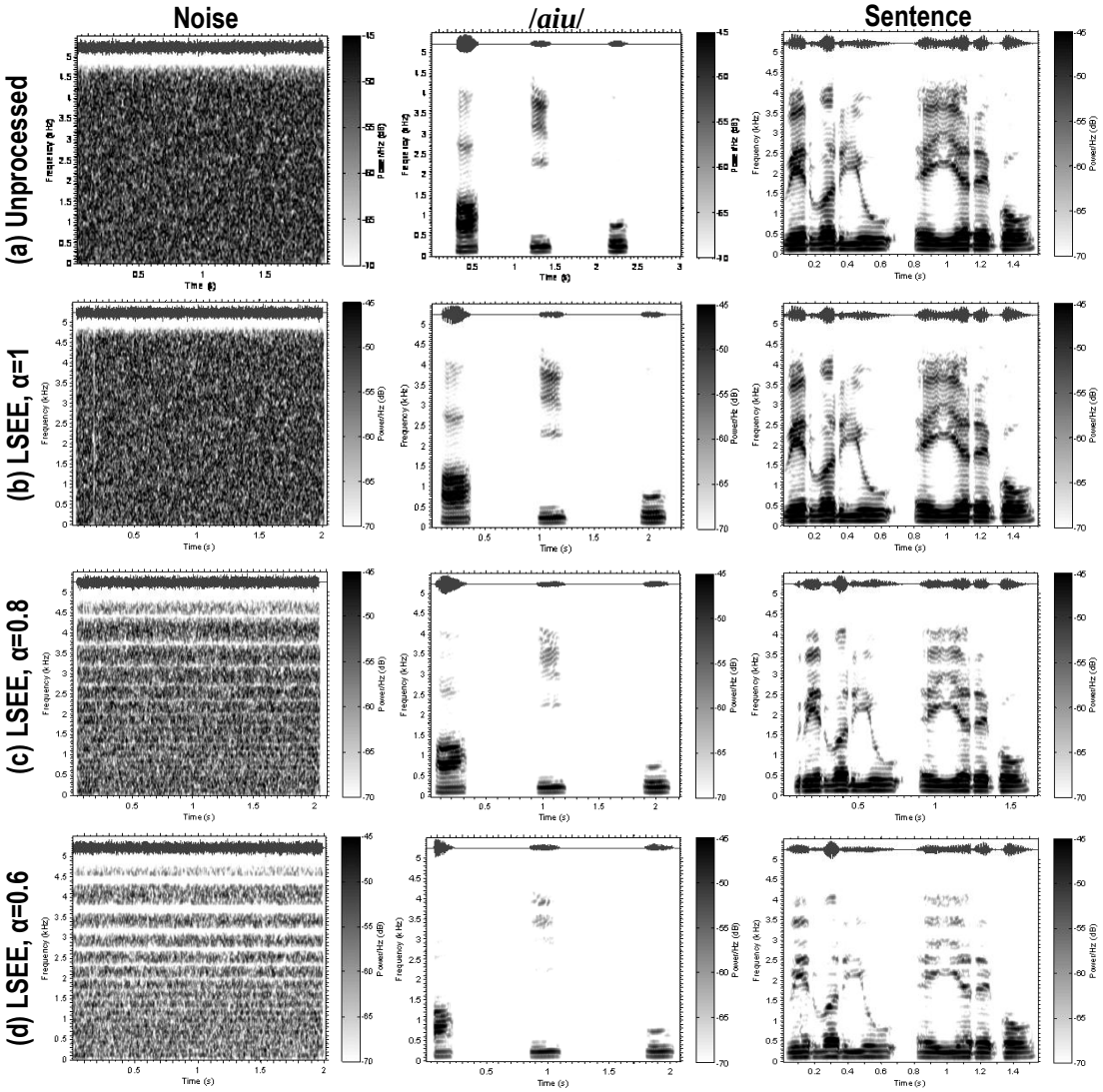


Figure 4.7. Real-time LSEE processing with FFT size = 512: spectrograms of noise, vowel /*aiu*/, and sentence "Where were you a year ago?": a) unprocessed, b) processed with α = 1, c) processed with α = 0.8 and d) processed with α = 0.6.

length 512 samples are shown in Figure 4.7. The spectrogram of the outputs obtained from real-time implementation using LSEE method for music clip input with compression factors of 1, 0.8, and 0.6 and FFT length 512 samples are shown in Figure 4.8. The expected processing delay for window length L = 260 samples and

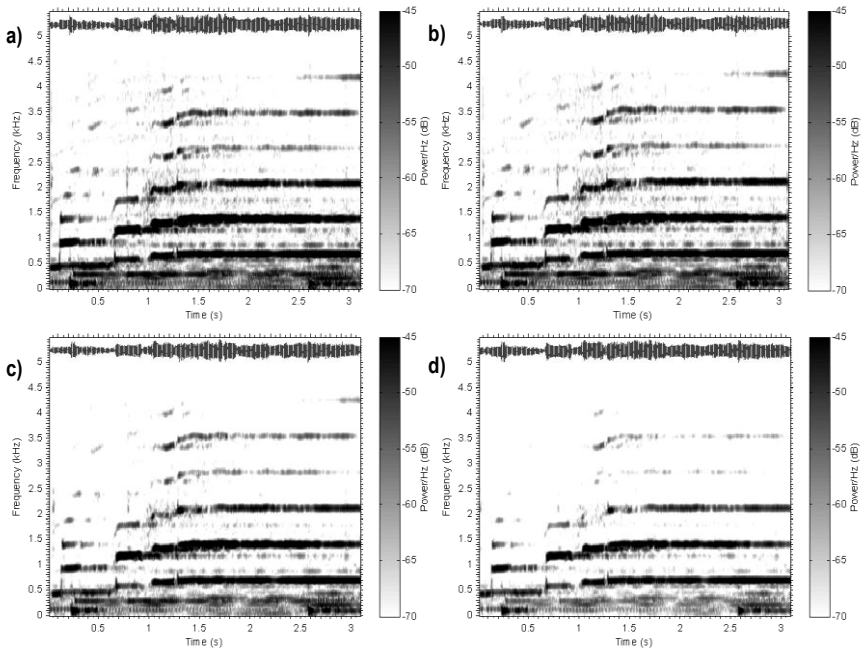


Figure 4.8. Real-time LSEE processing with FFT length 512: spectrograms of a music clip a) unprocessed, b) processed with $\alpha = 1$, c) processed with $\alpha = 0.8$ and d) processed with $\alpha = 0.6$.

shift $S = 65$ samples is $5S = 325$ samples or 32.5 ms. The processing delay observed by giving a burst sine wave of 1 kHz input with burst duration of 80 ms observed to be approximately 35 ms. This delay can be considered as acceptable for use in the hearing aids along with lipreading.

To estimate the computational capacity of the processor used in multi-band frequency compression, the program operation was tested by progressively decreasing the clock frequency from 120 MHz. For FFT length of 1024 the program worked satisfactorily down to 20 MHz, and for FFT length of 512 the program worked satisfactorily at a clock frequency of 12.288 MHz indicating a significant amount of unused computational capacity which may be useful in implementing other processing as needed for a hearing aid.

Chapter 5

SUMMARY AND CONCLUSION

Multi-band frequency compression using pitch-synchronous segmentation helps in removing the distortions associated with fixed-frame segmentation and improving speech perception. However, it is not suitable for real-time operation due to the algorithmic and computational delays associated with it. To avoid the artifact associated with the fixed-frame segmentation with 50% overlap and the additional algorithmic and computational delay associated with pitch estimation for the pitch-synchronous processing, Griffin-Lim's LSEE method [26] for signal estimation from modified short-time complex spectrum was investigated. The method works satisfactorily for speech as well as music and other audio signals.

For real-time operation, the LSEE method was implemented on a 16-bit fixed point processor TMS320C5515 based DSP board. Codec and DMA were used at a sampling rate of 10 kHz for continuous acquisition of the input signal and outputting of the processed signal. The on-chip FFT hardware accelerator facilitated the real-time processing. The data transfer and buffering operations were devised for an efficient realization of analysis-synthesis with 75 % overlap. The real-time processing with analysis window length of 26 ms and 512-point FFT was implemented, using about one-tenth of the computing capacity of the processor, with a processing delay under 35 ms, making it suitable for hearing aid applications. Informal listening tests showed that the processed output from the DSP board was perceptually similar to the corresponding output from the offline implementation for speech as well as other audio signals, for compression factors of 0.6, 0.8 and 1. The PESQ-MOS between offline and real-time speech outputs for compression factors of 0.6−1.0 was found to be 2.5−3.4.

For using the processing in hearing aids for persons with moderate sensorineural loss, frequency-selective gain and multi-band dynamic range compression, with the

gain and compression ratios settable in accordance with the loss characteristics of the individual listener, also need to be implemented.

REFERENCES

[1] H. Levitt, J. M. Pickett, and R. A. Houde, Eds., *Sensory Aids for the Hearing Impaired*, New York: IEEE Press, 1980, pp. 3–10.

[2] J. M. Pickett, *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*, Boston, Massachusetts: Allyn Bacon, 1999, pp 289–323.

[3] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, London, UK: Academic, 1997, pp 66–107.

[4] S. A. Gelfand, *Hearing*: *An Introduction to Psychological and Physiological Acoustics*, 3rd ed., New York: Marcel Dekker, 1998, pp. 314–318

[5] D. O' Shaughnaessy, *Speech Communications*: *Human and Machine*, 2nd ed., Hydrabad, India: Univ. Press, 2001, pp. 127–128

[6] T. Lunner, S. Arlinger, and J. Hellgren, "8-channel digital filter bank for hearing aid use: preliminary results in monaural, diotic, and dichotic modes," *Scand. Audiol. Suppl.*, vol. 38, pp. 75–81, 1993.

[7] P. N. Kulkarni, P. C. Pandey, and D. S. Jangamashetti, "Binaural dichotic presentation to reduce the effects of spectral masking in moderate bilateral sensorineural hearing loss," *Int. J. Audiol*., vol. 51, no. 4, pp. 334–344, 2012.

[8] D. S. Chaudhari and P. C. Pandey, "Dichotic presentation of speech signal with critical band filtering for improving speech perception," in *Proc. IEEE ICASSP* 1998, Seattle, Washington, pp. 3601–3604.

[9] A. N. Cheeran and P. C. Pandey, "Speech processing for hearing aids for moderate bilateral sensorineural hearing loss," in *Proc. IEEE ICASSP* 2004, Montreal, Quebec, IV-17-20.

[10] H. T. Bunnel, "On enhancement of spectral contrast in speech for hearing-impaired listeners," *J.Acoust. Soc. Am*., vol. 88, pp. 2546–2556, 1990.

[11] T. Baer, B. C. J. Moore, and S. Gatehouse, "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality, and response times," *Int. J. Rehab. Res*., vol. 30, no. 1, pp. 49–72, 1993.

[12] J. Yang, F. Luo, and A. Nehorai, "Spectral contrast enhancement: Algorithms and comparisons," *Speech Commun*., vol. 39, pp. 33–46, Feb. 2003.

[13] I. Cohen, "Speech spectral modeling and spectral enhancement based on autoregressive conditional heteroscendasticity models," *Signal Processing*, vol. 86, pp. 698-709, 2006.

[14] K. Yasu, K. Kobayashi, K. Shinohara, M. Hishitani, T. Arai, and Y. Murahara, "Frequency compression of critical band for digital hearing aids," *Proc. China-Japan Joint Conf. on Acoustics*, Nanjing, China, 2002, pp. 159–162.

[15] T. Arai, K. Yasu, and N. Hodoshima, "Effective speech processing for various impaired listeners," *Proc. 18th Int. Congr. Acoust*., Kyoto, Japan, 2004, pp. 1389–1392.

[16] K. Yasu, M. Hishitani, T. Arai, and Y. Murahara, "Critical-band based frequency compression for digital hearing aids," Acoustical Science and Technology, vol. 25, no. 1, pp. 61-63, 2004

[17] P. N. Kulkarni, P. C. Pandey, and D. S. Jangamashetti, "Multi-band frequency compression for reducing the effects of spectral masking," *Int. J. Speech Tech*., vol. 10, pp. 219–227, 2009.

[18] P. N. Kulkarni, P. C. Pandey, and D. S. Jangamashetti, "Multi-band frequency compression for improving speech perception by listeners with moderate sensorineural hearing loss," *Speech Commun*., vol. 54, no. 3, pp. 341–350, 2012.

[19] P. N. Kulkarni, "Speech processing for reducing the effects of spectral masking in sensorineural hearing loss," Ph.D. thesis, Electrical Engineering, Indian Institute of Technology Bombay, 2010.

[20] H. Dillon, *Hearing Aids*, New York: Thieme Medical Publisher, 2001.

[21] Robert E. Sandlin, *Textbook of Hearing Aid Amplification*, San Diego, Cal.: Singular 2000, pp. 210–220.

[22] D. G. Childers and H. T. Hu, "Speech synthesis by glottal excited linear prediction," *J. Acoust. Soc. Am*., vol. 96, no. 4, pp. 2026–2036, 1994.

[23] E. Zwicker, "Subdivision of the audible frequency range into critical bands (Freqenzgruppen)," *J. Acoust. Soc. Am*., vol. 33, no. 2, pp. 248, 1961.

[24] L. R. Rabiner and R. W Schafer, *Digital Processing of Speech Signals*, Englewood Cliffs, New Jersey: Prentice-Hall, 1978, pp. 274–277.

[25] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Englewood Cliffs, New Jersey: Prentice-Hall, 1994, pp. 548–560.

[26] D. W. Griffin and J. S. Lim, “Signal estimation from modified short-time Fourier transform,” *IEEE Trans. Acoustics, Speech, Signal Proc*., vol. 32, no. 2, pp. 236–243, 1984.

[27] X. Zhu, G. T. Beauregard, and L. L. Wyse, “Real-time signal estimation from modified short-time Fourier transform magnitude spectra,” *IEEE Trans. Audio, Speech, Language Proce*., vol. 15, no. 5, pp. 1645–1653, 2007.

[28] X. Zhu, G. T. Beauregard, and L. L. Wyse, “Real-time iterative spectrum inversion with look-ahead, ”in *Proc. IEEE Int. Conf. Multimedia and Expo*, Toronto, Canada, 2006 pp.229–232.

[29] V. Gnann and M. Spiertz, “Improving RTISI phase estimation with energy order and phase unwrapping,” in *Proc. 13th Int. Conf. Digital Audio Effects*, Graz, Austria, 2010, pp. 367–371.

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

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

[32] Texas Instruments Inc., “FFT Implementation on the TMS320VC5505, TMS320C5505, and TMS320C5515 DSPs,” 2010, [online] Available: http://www.ti.com/lit/an/sprabb6a/sprabb6a.pdf.

Acknowledgements

I would like to express my gratitude towards my guide Prof. P. C. Pandey for his invaluable guidance and support. I am also thankful to him for sparing his invaluable time to help me understand and implement the project. I would like to thank him sincerely for giving me the opportunity to learn and explore new concepts.

I am thankful to Dr. Panduranga Kulkarni for guiding me in project implementation. I would like to thank my seniors Rajath, Nataraj and Jagbandhu for their helpful advices in the different issues of the project. I would like to thank Santosh, Adithya and Pranava for helping me in real time implementation and sharing interesting discussions with me.

Nitya Tiwari

June 2012

Author's Resume

Nitya Tiwari: The author received the BE degree from Shri. G. S Institute of Technology and Science, Indore, affiliated to Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, in electronics and telecommunication in 2010. She is currently pursuing her MTech degree in electrical engineering at the Indian Institute of Technology Bombay. Her research interests include speech processing and digital signal processing.

Thesis related publication

N. Tiwari, P. C. Pandey, and P. N. Kulkarni, "Real-time implementation of multi-band frequency compression for listeners with moderate sensorineural impairment," accepted for publication in *Proc. Interspeech 2012*, Portland, Oregon, Sept. 9, 2012.