DYNAMIC RANGE COMPRESSION AND NOISE SUPPRESSION FOR USE IN HEARING AIDS

Thesis
submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

by

Nitya Tiwari
(Roll No. 124070014)

under the supervision of

Prof. P. C. Pandey

Department of Electrical Engineering
Indian Institute of Technology Bombay

June 2019

Left blank

Dedicated to my family and teachers

Left blank

Indian Institute of Technology Bombay
Department of Electrical Engineering

Ph.D. Thesis Approval

Thesis entitled **"Dynamic Range Compression and Noise Suppression for Use in Hearing Aids"** by **Nitya Tiwari (Roll No. 124070014)** is approved, after the successful completion of viva voce examination, for the award of the degree of **Doctor of Philosophy.**

Supervisor:	PCPandey 26/06/2019	(Prof. P. C. Pandey)
Internal Examiner:	Preeti Rao	(Prof. P. Rao)
External Examiner:	Sumesh	(Prof. S. Umesh)
Chairman	26/06/2019	(Prof. A. Datta)

Date: 26th June, 2019

Place: Mumbai

Left blank

DECLARATION

I declare that this written submission 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 also 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
(Roll No: 124070014)

Date: 26th June, 2019
Place: Mumbai

Left blank

N. Tiwari / Supervisor: Prof. P. C. Pandey, “Dynamic range compression and noise suppression for use in hearing aids,” *Ph.D. Thesis*, Department of Electrical Engineering, Indian Institute of Technology Bombay, June 2019.

Abstract

Sensorineural hearing loss is associated with elevated hearing thresholds, reduced dynamic range, and increased temporal and spectral masking, resulting in degraded speech perception, particularly in noisy environments. The research objective is to develop signal-processing techniques for dynamic range compression and background noise suppression to enhance the performance of hearing aids for the listeners with sensorineural loss, with considerations for low computational requirements, low audio latency, and low perceptible distortions.

A dynamic range compression technique, named as ‘sliding-band compression’ (SLBC), is proposed to overcome the shortcomings of the single-band and multiband compressions. The gain for each spectral sample is based on the short-time power in an auditory critical band centered at its frequency. The technique avoids the attenuation of high-frequency components in the presence of strong low-frequency components, which may occur in single-band compression. Further, it avoids distortions in the shape of spectral resonances and discontinuities during the resonance transitions, which may occur in multiband compression.

Two techniques for quantile-based noise estimation are developed for single-input speech enhancement. A technique is proposed for dynamic tracking of quantiles of a data stream, without storage and sorting of the past samples and without prior knowledge of the distribution. The quantile is estimated recursively by applying an increment, calculated as a fraction of the dynamically estimated range, such that the estimate converges to the sample quantile. This technique is applied for the tracking of the quantiles of the spectral samples of the noisy speech spectrum for noise spectrum estimation without voice activity detection, resulting in the technique named as ‘dynamic quantile tracking based noise estimation’ (DQTNE). For improving the tracking of nonstationary noises, the technique named as ‘adaptive dynamic quantile tracking based noise estimation’ (ADQTNE) and using an adaptive quantile and an adaptive convergence factor is proposed. The two techniques are evaluated and compared with some of the existing techniques in terms of computational requirement and noise tracking and in a speech enhancement framework using spectral subtraction based on the geometric approach. Considering residual noise and speech attenuation together, ADQTNE provides the highest quality output. DQTNE, having the lowest computational requirement, provides a similar output, except that the output has a higher residual noise in case of low SNRs. Considering the increase in PESQ scores for the different noises, ADQTNE and DQTNE provide SNR advantages of 4−11 and 3−10 dB, respectively.

The proposed techniques are implemented using a fixed-point processor for real-time processing with audio latency acceptable for face-to-face communication. They are also implemented as a smartphone app with a graphical touch interface for setting the processing parameters in an interactive and real-time mode.

Left blank

CONTENTS

List of Figures

List of Tables

Left blank

List of Symbols

Symbols

ABG	air-bone gap
CR	compression ratio
BW(k)	auditory critical bandwidth at kth spectral sample
$D(n, k)$	noise spectrum at nth frame and kth spectral sample
$\hat{D}(n,k)$	noise estimate at nth frame and kth spectral sample
$\hat{D}_{\text{DQTNE}}$	noise estimate obtained using DQTNE
$\hat{D}_{\text{ADQTNE}}$	noise estimate obtained using ADQTNE
FE	sound field to ear canal transform
$f(k)$	frequency of the kth spectral sample in kHz
f_s	sampling frequency
G	frequency-dependent gain in dB
G_{GA}	SNR-dependent gain function
$G_{\max}$	maximum value of the target gain
$G_{\min}$	minimum value of the target gain
$G_{2\text{PLC}}$	SNR-dependent proposed two-point linear compression
$G_{2\text{PSC}}$	SNR-dependent proposed two-point smooth compression
$G_T(n, k)$	target gain at nth frame and kth frequency sample
$G_{TdB}(n, k)$	target gain in dB, at nth frame and kth frequency sample
H	hearing threshold in dB HL
H_{BC}	bone-conduction hearing threshold in dB HL
$H_{3\text{FA}}$	three-frequency average of the hearing thresholds
HTL_{im}	hearing threshold for the hearing-impaired listener
HTL_n	hearing threshold for normal hearing
k	spectral sample
K	FFT size
L	number of samples in an analysis frame
$[\text{Level error}]_{\text{dB}}$	tone-level error in dB
m	time index
M	number of evenly spaced probabilities
MCL_{im}	most comfortable level for the hearing-impaired listener
MCL_n	most comfortable level for the normal-hearing listener

n	frame index
N	total number of frames
p	probability corresponding to a quantile of a random variable
$p_{\max}$	peak of the probability density function
p_{si}	instantaneous speech presence probability
p_{ss}	smoothed speech presence probability
P	peak estimate of a data stream
$P_{\|X\|}$	peak estimate of magnitude spectrum of noisy input
P_{IdB}	input level (dB)
P_{IdB1}	input level (dB) for compression threshold
P_{IdB2}	input level (dB) for output-limiting threshold
P_{OdB}	output level (dB)
P_{OdB1}	output level (dB) for compression threshold
P_{OdB2}	output level (dB) for output-limiting threshold
$P_{OdB,2PSC}$	output level (dB) for proposed two-point smooth compression
$P_{OdB,2PLC}$	output level (dB) for proposed two-point linear compression
P_{OdBCL}	output level (dB) corresponding to comfortable sounds
P_{OdBLL}	output level (dB) corresponding to loud sounds for hearing aid user
P_{OdBSL}	output level (dB) corresponding to soft sounds for hearing aid user
P_{IdBLL}	input level (dB) corresponding to loud sounds for normal-hearing listener
P_{IdBSL}	input level (dB) corresponding to soft sounds for normal-hearing listener
PC	profound correction term
q	p-quantile of a random variable
$\hat{q}$	estimate of p-quantile q
q_{fin}	final value of the estimated quantile
q_{init}	initial value of the estimated quantile
$q_{\|X\|}$	p-quantile of a short-time magnitude spectrum of noisy speech
$\hat{q}_{\|X\|}$	estimate of p-quantile of magnitude spectrum of noisy input
$\hat{q}_{\|X\|,i}$	estimate of p_i-quantile of magnitude spectrum of noisy input
$\hat{q}_{\|X\|,\text{ad}}(n,k)$	adaptive quantile of magnitude spectrum of noisy input
R	range of values of a data stream
$\bar{R}$	time-varying average range

$R_{\|X\|}$	range of the spectral values at a particular frequency
$\mathrm{RMS}_{\mathrm{expected}}$	RMS value of the tone in the expected output
$\mathrm{RMS}_{\mathrm{output}}$	RMS value of the tone in the processed output
S	number of samples in a frame shift
s	number of steps needed for estimated quantile to change from its initial value to its final value
s_a	number of steps during the attack
s_r	number of steps during the release
Th	compression threshold
T_a	attack time
T_r	release time
UCL_{im}	uncomfortable level for the hearing-impaired listener
UCL_n	uncomfortable level for the normal-hearing listener
V	valley estimate of a data stream
$V_{\|X\|}$	valley estimate of magnitude spectrum of noisy input
x	noisy signal
$X(n, k)$	noisy spectrum at nth frame and kth spectral sample
$\|X(n, k)\|$	noisy magnitude spectrum at nth frame and kth spectral sample
$Y(n, k)$	enhanced spectrum at nth frame and kth spectral sample
α	updating constant for estimation of speech presence probability
β	smoothing constant for estimation of speech presence probability
γ_a	gain ratio to control number of steps during the attack
γ_r	gain ratio to control number of steps during the release
δ	peak-to-peak ripple in the estimated quantile
Δ_+	increment on the previous quantile estimate
Δ_-	decrement on the previous quantile estimate
λ	convergence factor
$\lambda_{\max}$	upper bound on λ
λ_s	speech presence probability dependent convergence factor
ξ	smoothed *a priori* SNR
σ_p	coefficient to control fall time of peak detector
σ_v	coefficient to control fall time of valley detector
τ_p	coefficient to control rise time of peak detector
τ_v	coefficient to control rise time of valley detector
ψ	smoothed *a posteriori* SNR

Left blank

List of Abbreviations

Abbreviations	
ABG	air-bone gap
AC	air conduction thresholds
ADC	analog-to-digital convertor
ADQTNE	adaptive dynamic quantile tracking based noise estimation
ASIC	application specific integrated circuit
ASL	active speech level
AVC	automatic volume control
BC	bone conduction thresholds
BTE	behind the ear
CITC	completely in the canal
CL	comfortable level
CPU	central processing unit
CR	compression ratio
DAC	digital-to-analog convertor
dB	decibel
DD	decision directed
DFT	discrete Fourier transform
DMA	direct memory access
DQTNE	dynamic quantile tracking based noise estimation
DQTRE	dynamic quantile tracking using range estimation
DQTRSE	dynamic quantile tracking using range segment estimation
DRT	diagnostic rhyme test
DSL	desired sensation level
DSL[i/o]	desired sensation level input-output
DSLm[i/o]	multistage DSL[i/o]
DSP	digital signal processing
EC	ear canal
EWSA	exponentially weighted stochastic approximation
FFT	fast Fourier transform
FIG6	Figure 6 based fitting procedure
FIR	finite impulse response
FPGA	field programmable gate array

GA	geometric approach to spectral subtraction
GUI	graphical user interface
HINT	hearing in noise test
HL	hearing level
HMM	hidden Markov model
HTL	hearing threshold level
IFFT	inverse fast Fourier transform
IHAFF	Independent Hearing Aid Fitting Forum
IMCRA	improved minima-controlled recursive averaging
ITC	in-the-canal
ITE	in-the-ear
KB	kilo byte
kHz	kilo Hertz
LGOB	loudness growth in half-octave bands
LL	loud level
LPC	linear predictive coding
LSEE	least squares error estimation
MB	megabyte
MBC	multiband compression
MCL	most comfortable level
MCRA	minima-controlled recursive averaging
MCRA2	computationally efficient minima-controlled recursive averaging
ML	maximum likelihood
MMSE	minimum mean-square error
MMSE-LSA	minimum mean square error log-spectral amplitude estimator
MMSE-STSA	minimum mean-square error based short-time spectral amplitude
MMSE-STSA-ML	MMSE-STSA with a priori SNR estimated using ML
MMSE-STSA-DD	MMSE-STSA with a priori SNR estimated using DD
MMSE-STSA-DD-SPU	MMSE-STSA-DD with signal presence uncertainty
MS	minimum statistics
NAL	National Acoustic Laboratory formula
NAL-NL1	NAL non-linear formula
NAL-NL2	revised NAL non-linear formula
NAL-R	revised NAL formula
NAL-RP	NAL revised profound formula

OM-LSA	optimally modified log-spectral amplitude
OS	operating system
P2	piecewise-parabolic formula
PC	personal computer
PESQ	perceptual evaluation of speech quality
POGO	prescription of gain and output
POGO II	revised prescription of gain and output
PSD	power spectral density
RAM	random access memory
RIC	receiver-in-the-canal
RMS	root mean square
RMSE	root mean square error
ROM	read only memory
SBC	single-band compression
SII	speech intelligibility index
SF	sound field
SL	soft level
SLBC	sliding-band compression
SNR	signal-to-noise ratio
SPL	sound pressure level
SQ	sample quantile
SREE	segmental relative estimation error
SSA	smooth stochastic approximation
TIMIT	Texas Instruments/MIT speech corpus
TK	threshold knee-point
TRRS	tip ring ring sleeve
UCL	uncomfortable level
UI	user interface
UMMSE	unbiased minimum mean-square error
VAD	voice activity detector
WDRC	wide dynamic range compression
Weiner-ML	Weiner filter with *a priori* SNR estimated using ML
Weiner-DD	Weiner filter with *a priori* SNR estimated using DD
2PLC	two-point linear compression
2PSC	two-point smooth compression

Left blank

Chapter 1

INTRODUCTION

1.1 Problem Overview

Hearing-impaired persons suffer from different types of hearing losses such as conductive, sensorineural, central, functional, and mixed [1]–[4]. Conductive loss is caused by disorders of the middle ear and is characterized by frequency-dependent elevation of hearing thresholds without a change in the dynamic range of hearing. Sensorineural loss is caused by degeneration of the sensory hair cells of the inner ear or the auditory nerve. It is characterized by frequency-dependent elevation of hearing thresholds, significantly reduced dynamic range of hearing and abnormal loudness growth leading to distorted loudness relationship among speech components, increased temporal masking leading to poor detection of acoustic landmarks, and increased spectral masking leading to reduced ability to sense spectral shapes. Mixed loss refers to conductive and sensorineural disorders in the same ear.

A hearing aid is an electro-acoustic device used for partially overcoming the deficits associated with hearing loss. It transforms the input sounds to improve sound audibility, comfort, and speech intelligibility for the user. In addition to providing frequency-selective amplification, these devices employ signal processing for dynamic range compression and noise reduction. Generally, listeners having a mild-to-moderate loss, with a varying combination of conductive and sensorineural losses, are most likely to benefit from the use of these devices [5]–[8].

One of the major problems faced by the listeners with sensorineural loss is a reduction in the dynamic range of hearing, with a significant frequency-dependent elevation of hearing thresholds without a corresponding increase in the upper discomfort level. The sensory mechanism in the cochlea consists of inner and outer hair cells. The loss of inner hair cells results in elevated thresholds while that of outer hair cells leads to abnormal growth of loudness known as loudness recruitment and a significantly reduced dynamic range [1], [2]. Signal processing with dynamic range compression is used for presenting all the sounds comfortably within the limited dynamic range of the listener. The commonly used compression techniques in the hearing aids may introduce perceptible distortions, partly offsetting the advantages of the dynamic range compression.

Hearing-impaired listeners experience degraded speech perception due to increased temporal and spectral masking, particularly in the presence of background noise [3], [5].

Suppression of background noise as part of the signal processing in hearing aids can serve as a practical solution for improving speech quality and intelligibility for persons with sensorineural or mixed loss.

1.2 Research Objective

The research objective is to develop signal-processing techniques for dynamic range compression and background noise suppression, in order to overcome the shortcomings of the currently employed techniques and thus to enhance the performance of the hearing aids used by listeners with sensorineural loss. The techniques are developed with considerations for low memory and computational requirements for implementation in hearing aids, low input-output delay for face-to-face communication, and low perceptible distortions.

Single-band and multiband compression techniques [9]–[12] are used to compensate for the reduced dynamic range and loudness recruitment. However, single-band compression leads to reduced high-frequency audibility and multiband compression reduces spectral contrast and may distort the spectral shape of a formant spanning the band boundaries. A compression technique to overcome the disadvantages of the commonly employed single-band and multiband compression techniques needs to be developed.

Signal processing in hearing aids for speech enhancement by suppression of background noise helps in improving speech quality and intelligibility for persons with sensorineural hearing loss. The speech enhancement technique should be effective for suppression of stationary as well as nonstationary noises. Further, it should have a low algorithmic delay and low computational complexity for implementing it using a low-power processor in a hearing aid. Single-input speech enhancement techniques can be used for background noise suppression and improving speech perception of the hearing-impaired listeners. They involve estimation of the noise spectrum from the noisy speech spectrum and using the estimated noise spectrum along with a noise suppression function for speech enhancement. Underestimation of the noise results in residual noise and overestimation results in distortion leading to degraded quality and reduced intelligibility. It has been reported [13]–[16] that noise spectrum may be estimated by selecting a certain quantile value from previous frames of the noisy speech spectrum. Although these quantile-based noise estimation techniques perform well in presence of stationary and non-stationary noises, they are not suited for use in hearing aids as they require large memory and have high computational complexity.

In order to overcome the drawbacks of the existing techniques for dynamic range compression and background noise suppression, investigations are carried out to develop two signal-processing techniques for use in hearing aids: (i) sliding-band dynamic range

compression to compensate for frequency-dependent loudness recruitment and (ii) speech enhancement using dynamic quantile tracking for estimation of background noise.

The processing for sliding-band dynamic range compression aims at reducing the temporal and spectral distortions generally associated with the single-band and multiband compression techniques. The proposed technique is implemented for offline processing and is compared with single-band compression and multiband compression techniques using objective measures and different test inputs.

A technique for dynamic tracking of quantiles for use in applications involving real-time estimation of quantiles of a data stream is developed. This technique is subsequently applied for tracking of the quantiles of the noisy speech spectrum for noise spectrum estimation without voice activity detection. It permits the use of a different quantile for each sub-band without processing overheads. An improved noise estimation technique that selects the quantile adaptively is also presented. It involves estimating a quantile function (inverse of cumulative distribution function) for each spectral sample by dynamically tracking multiple quantiles. The proposed techniques are evaluated and compared with some of the existing techniques in terms of computational requirement and noise tracking. The proposed techniques in combination with spectral subtraction based on the geometric approach are used for suppression of background noise and evaluated in a speech enhancement framework.

The implementations of the sliding-band dynamic range compression and the speech enhancement using dynamic quantile tracking based noise estimation are carried out using a fixed-point DSP chip for real-time processing. A smartphone app implementing the proposed techniques with an interactive touch-controlled graphical user interface is also developed.

1.3 Thesis Outline

The second chapter presents a brief review of hearing impairment, its adverse effect on speech perception, single-band and multiband compression techniques, and techniques for noise estimation and noise suppression for speech enhancement. The proposed sliding-band compression technique for dynamic range compression, its offline and real-time implementations, and test results are presented in the third chapter. The proposed dynamic quantile tracking technique, noise estimation using dynamic quantile tracking, an improved noise estimation using adaptive dynamic quantile tracking, speech enhancement using the proposed noise estimation techniques, offline and real-time implementations of the speech enhancement, and test results are presented in the fourth chapter. The last chapter provides a summary of the investigations, conclusions, and suggestions for further work. A brief review of hearing aid fitting procedures is given in Appendix A. A description of the dynamic

quantile tracking technique, as developed for quantile-based noise estimation, along with the test results is presented in Appendix B. A smartphone app implementation of the proposed techniques with interactive touch-controlled graphical user interface is presented in Appendix C.

Chapter 2

SIGNAL PROCESSING IN HEARING AIDS

2.1 Introduction

This chapter provides the background material for signal processing in hearing aids. A brief description of hearing impairment and its adverse effect on speech perception is presented in Section 2.2, followed by a description of the signal processing techniques to improve speech perception in Section 2.3. The basics of dynamic range compression are presented in Section 2.4, followed by a review of the dynamic range compression techniques in the subsequent section. The basics of speech enhancement by suppression of background noise are presented in Section 2.6, followed by reviews of the techniques for noise estimation and noise suppression in the two subsequent sections. The scope of the research is presented in the last section.

2.2 Hearing Impairment

Hearing-impaired persons suffer from different types of hearing losses such as conductive, sensorineural, mixed, central, and functional [1]–[4]. Conductive loss is caused by disorders of the middle ear and is characterized by frequency-dependent elevation of hearing thresholds without any change in the dynamic range of hearing. It can be compensated by frequency-selective amplification. Sensorineural loss is caused by abnormalities in the sensory hair cells or the auditory nerve. It may be inherited genetically or may be caused by aging, infection, excessive exposure to noise, or use of ototoxic drugs. Persons with such loss experience difficulty in speech perception, particularly in noisy environments. Mixed loss refers to conductive and sensorineural disorders in the same ear. The central loss may occur due to skull trauma, damage to the auditory cortex, cerebral meningitis, or congenital defects. It is associated with a reduced ability to interpret and integrate speech. Functional loss occurs due to psychological or emotional factors and it is characterized by poor speech understanding, without an associated pathology of the auditory system.

The sensory mechanism in the cochlea consists of inner and outer hair cells. The loss of inner hair cells results in elevated hearing thresholds. The elevation in hearing thresholds is generally higher for frequencies above 1 kHz [5]. In the speech signal, the power is contributed mostly by the low-frequency components and the high-frequency components are generally weak. Therefore, a hearing-impaired listener may perceive the speech as sufficiently

loud due to the presence of strong low-frequency components, but may miss the high-frequency information leading to poor speech intelligibility.

The difference between the hearing threshold and uncomfortable levels is the dynamic range. A significant frequency-dependent elevation of hearing thresholds due to loss of inner hair cells without a corresponding increase in the uncomfortable level leads to reduced dynamic range in listeners with sensorineural loss [3]. Further, the loss of outer hair cells leads to abnormal growth in the sensation of loudness with increase in the acoustic signal level, known as loudness recruitment [1], [3]. A reduced dynamic range of hearing limits the perception of speech, with typical conversational levels of 30−60 dB. Further, loudness recruitment leads to a distorted loudness relationship among the components of speech sounds leading to degraded speech perception.

Masking of a weak sound by a preceding or following intense sound is known as temporal masking. Persons with sensorineural loss show forward masking (the masker preceding the signal) and backward masking (the masker following the signal) over 100–200 ms as compared to about 10 ms in case of normal-hearing persons [17], [18]. A normal-hearing listener is able to extract useful information during the gaps in the masker such as fluctuating background noise. This ability to hear the weaker sounds during the gaps in the intense masker decreases with an increase in sensorineural loss [5]. Even in the absence of background noise, the consonantal segments are susceptible to masking by their adjacent vowel segments resulting in degraded speech perception [6]. The loss of outer hair cells in sensorineural loss results in widening of the auditory filters, which results in increased spectral masking leading to reduced ability to sense spectral features [1]–[4]. In the presence of background noise, more noise gets through the relatively wider auditory filters and discrimination of spectral features gets degraded [3].

In summary, sensorineural loss is associated with frequency-dependent elevation of hearing thresholds leading to inaudibility of low-level sounds; significantly reduced dynamic range of hearing and abnormal loudness growth leading to distorted loudness relationship among speech components; increased temporal masking leading to poor detection of acoustic landmarks; and increased spectral masking leading to reduced ability to sense spectral features, particularly in the presence of background noise. The signal processing in hearing aids aims to alleviate these effects of sensorineural loss.

2.3 Hearing Aids

Hearing aids are used to compensate for varying combination of mild-to-severe conductive and sensorineural losses. In a hearing aid, the acoustic input is converted to electrical signal

using a microphone, the signal is processed to compensate for hearing loss, and it is converted back to acoustic output using a receiver and coupled into the ear canal of the listener. Digital hearing aids are based on digital signal processing (DSP) chips or application-specific integrated circuit (ASIC) chips. The DSP chips offer more programming flexibility and faster development cycles whereas the ASIC chips are specifically designed for high performance, low power, and low area. The currently available hearing aids are generally categorized depending on the place where they are worn as body-worn, behind-the-ear (BTE), receiver-in-the-canal (RIC), in-the-ear (ITE), in-the-canal (ITC), and completely-in-the-canal (CITC) hearing aids [5], [7].

Several signal-processing techniques have been reported for alleviating the effects of sensorineural hearing loss [17]–[28]. While frequency-selective linear amplification can be used to compensate for frequency-dependent elevation of hearing thresholds, its benefits are generally limited because the amplification has to be selected as a trade-off between sufficient amplification for low level sounds and little or no amplification for high level sounds. Linear amplification makes the weaker sounds (such as consonants) audible, but it may make the high-level sounds (such as vowels) uncomfortably loud [5], [6]. Dynamic range compression in hearing aids is provided with the objective of presenting all the sounds comfortably within the limited dynamic range of the listener without introducing significant perceptible distortions [5], [6]. It reduces the level differences between the high and low-level components of the signal in order to amplify the low-level sounds without making the high-level sounds uncomfortably loud. Frequency-selective amplification and dynamic range compression form the core of signal processing in hearing aids. Further, speech enhancement by suppression of background noise is needed, because this noise may cause excessive temporal and spectral masking and severely degrade speech perception.

Consonant-to-vowel ratio enhancement [19] has been reported for reducing the effects of increased temporal masking. The low-level consonants, generally preceded or succeeded by high-level vowels, are more susceptible to temporal masking. The consonant-to-vowel intensity ratio enhancement improves the perception of low intensity consonants. The duration modification of acoustic segments [21], [22] has also been reported for reducing the effects of increased temporal masking. Several signal-processing techniques, such as binaural dichotic presentation [23], spectral contrast enhancement [25], [26], and multiband frequency compression [27], [28], have been reported for reducing the effects of increased spectral masking. In binaural dichotic presentation, the input signal is split into two parts in a complementary manner such that the components likely to mask each other are presented to the different ears [23]. The processing for spectral contrast enhancement involves

enhancement of the spectral prominences that are perceptually significant. Multiband frequency compression concentrates the spectral energy towards the center of auditory critical bands without introducing any spectral tilt or compression of the broadband spectrum [27].

Digital hearing aids, designed to present the sounds at a comfortable level with reduced noise and distortions to the hearing aid user, are available from several manufacturers such as Starkey, GN ReSound, Phonak, Beltone, Siemens, Widex, Costco etc. They have features such as dynamic range compression, directional microphones, background noise suppression, wind noise suppression, feedback cancellation, environment learning, comfort noise feature for patients with tinnitus, ease of customization using mobile applications, and wireless connectivity to stereo sets, phone, TV, etc.

Starkey's hearing aid 'S Series iQ' [29] provides dynamic range compression and uses a fast-acting noise suppression 'voiceIQ' [30], which uses overall signal level, estimated signal-to-noise ratio, and signal modulation for classifying the input as speech or noise and reduces the gain when noise is detected in a particular frequency band. The hearing aid 'Alera' from GN ReSound [31] provides multiband dynamic range compression using 17 auditory critical bands, each with 6−9 knee points. For noise suppression, it uses 'NoiseTracker II' [32] based on spectral subtraction and adaptive estimation of the noise spectrum using an environmental classifier with four settings for mild, moderate, considerate, and strong background noise situations. The 'Core' hearing aid from Phonak [33] uses adaptive dual-path compression with slow-acting and fast-acting compressions, for selecting the most suitable attack and release time constants. Phonak's 'Ambra' [34] provides a zooming feature, with direction selective sensitivity in presence of background noise. The 'Legend' hearing aid from Beltone [35] uses multiband compression with 12−17 bands and curvilinear compression function. Several other models, such as 'Ally', 'Boost', and 'Bold' from Beltone [36] use 'Sound Cleaner Pro' [37] based on spectral subtraction for reducing the background noise, which uses 'Smart Gain Pro' [37] for noise spectrum estimation. This technique employs a combination of a signal detector, noise detector, and signal power estimator to analyze the acoustic environment and to calculate speech probability in the input signal. Noise suppression is carried out using frequency-dependent gain reduction based on the speech probability.

The following sections provide basics of dynamic range compression, a review of the dynamic range compression techniques, basics of speech enhancement by suppression of background noise, a review of noise estimation techniques, a review of noise suppression techniques, and the scope of the research.

2.4. Dynamic Range Compression

Dynamic range compression in hearing aids is provided with the objective of presenting all the sounds comfortably within the limited dynamic range of the listener without introducing significant perceptible distortions. A compression amplifier provides a high gain for low-level inputs and reduces the gain progressively for inputs above a threshold. Compression function relating the output level to the input level on a dB scale characterizes the static behavior of compression amplification. The compression ratio is the inverse of the slope of the compression function. For low-level input, the compression function generally has a linear amplification segment with the compression ratio as one. The compression threshold or knee-point is the input level at which compression ratio changes to a value greater than one. The compression ratio in the compression segment may be constant or it may vary with the input level. As the output level reaches uncomfortable level, compression-limiting is applied and the output level remains constant for further increase in the input level. In compression-limiting, the compression ratio is infinity and the gain decreases by a dB for each dB increase in the input level. The compression function may employ a piecewise linear or curvilinear relationship on a dB scale. A three-segment piecewise-linear compression function, comprising a linear segment, a compression segment, and a compression-limiting segment, is used most commonly. It is specified by the gain in the linear segment, the compression threshold, the compression ratio in the compression segment, and the output level in the compression-limiting segment.

A level detector is used in compression amplification to estimate the signal level and to determine the gain. Its temporal characteristics determine the dynamic behavior of the compression amplification, i.e. the speed with which the gain control reacts to changes in the input level. Compression attack is a gain decrease in response to an increase in the input level and compression release is a gain increase in response to a decrease in the input level. Generally, a fast attack is used to prevent the output level from exceeding the listener's uncomfortable loudness level, and a slow release is used to reduce audible perturbations in the gain. The attack time is defined as the time from an abrupt change from 55 to 90 dB SPL in the level of the input tone (at 0.25, 0.50, 1.0, 2.0, and 4.0 kHz) to the output reaching within 3 dB of its steady-state value. The release time (also known as recovery time) is defined as the time from an abrupt change from 90 to 55 dB in the level of the input tone to the output reaching within 4 dB of its steady-state value [38].

Dynamic range compression schemes may be classified on the basis of (i) compressor location, (ii) knee-point control, (iii) compression curve, (iv) dynamic behavior, (v) system realization, and (vi) number of bands [5], [6].

(i) Compressor location: The compression is generally used in combination with overall volume control, which enables the hearing aid user to adjust the output level depending on the listening environment. Based on the volume control location, the compression schemes are classified as input-controlled and output-controlled [5], [6]. In the input-controlled compression, the compressor precedes the volume control. In this scheme, the volume control setting does not affect the compression threshold and compression ratio. It shifts the compression function vertically, resulting in a change in the compression-limiting output. In the output-controlled compression, the compressor acts on the signal after the volume control. In this scheme, the volume control setting shifts the compression function horizontally and it does not affect the compression-limiting output. The output-controlled compression is suitable for severe-to-profound hearing loss with a small residual dynamic range. The input-controlled compression is suitable for listeners with a mild-to-moderate sensorineural loss with a larger residual dynamic range, as it permits adjustment of compression-limiting output.

(ii) Knee-point control: The knee-point (compression threshold) is generally adjusted by the audiologist in accordance with the input range for which the compression is to be applied. The knee-point can be adjusted using conventional control or using threshold knee-point (TK) control [6]. In the conventional control, the linear-segment gain remains constant and the compression-limiting output changes with the knee-point. In TK control, the knee-point is adjusted without any change in the compression-limiting output, but the linear-segment gain of the compressor changes with the knee-point. The compression ratio remains unchanged in both conventional control and TK control. The conventional control is mostly used with output compression and sometimes with input compression, whereas the TK control is mostly used with input compression.

(iii) Compression curve: Based on the compression curve, the compression schemes are classified as output limiting and wide dynamic range compression (WDRC). Output limiting uses a high knee-point (50−60 dB SPL) and a large compression ratio (greater than 5) to prevent the output from becoming uncomfortably loud [6], [39]. It gives a high linear gain to almost all the input levels and uses a very high compression ratio once the input exceeds the threshold. It is suitable for listeners with a severe-to-profound sensorineural loss. WDRC uses a low knee-point (below 55 dB SPL) and a small compression ratio (lesser than 5) with an aim to improve the audibility of low-level inputs and to make loudness perception of a hearing-impaired similar to that of a normal-hearing listener. It remains in compression for

almost all the input levels and is suitable for listeners with a mild-to-moderate sensorineural loss.

(iv) Dynamic behavior: The dynamic behavior is determined by the level detector used in the compressor. Based on the dynamic behavior, the compression schemes generally used in hearing aids may be classified as syllabic (or fast) compression and automatic volume control (AVC). Syllabic compression uses an attack time of 2−10 ms and a release time of 20−150 ms. To avoid distortions, it usually has a low compression ratio and a low compression threshold. AVC uses a long attack time (≈ 300 ms) and even longer release time (≈ 500 ms − 2 s) [9]. It can be used with a wide range of compression ratios and compression thresholds. A fast-acting syllabic compressor is suitable for listeners with a small dynamic range [5], [10], [40], [41]. It is provided to improve the audibility of weaker consonant sounds without making the intense vowel sounds uncomfortably loud by increasing the gain for weaker segments and reducing the gain for stronger segments. AVC reduces the long-term dynamic range without altering the short-term level relation between the syllables. It is suitable for listeners with a large residual dynamic range. The dual front-end compressor uses a slow-acting compressor cascaded with a fast-acting compressor to realize a compressor that provides protection against brief intense sounds without causing rapid gain fluctuations during high-level speech.

(v) System realization (feed-forward and feedback): In a feedback compression scheme, the gain is calculated based on the output level and is applied to the input as a corrective measure. It may result in overshoots and undershoots in the output level as the gain depends on the output level itself. In a feed-forward compression scheme, the gain depends on the input signal level and it is applied to an appropriately delayed input to avoid overshoots and undershoots in the output level. It is not suited for AVC type of compression having level detector with large time constants.

(iv) Number of bands: Based on the number of bands used in the processing, the compression schemes may be classified as single-band or multiband. In single-band compression, the gain is a function of the signal level over its entire bandwidth. As the power is mostly contributed by the low-frequency components, the amplification of the high-frequency components is affected by the energy of the low-frequency components. In multiband compression, the input signal is divided into several bands and the gain for each band is a function of the signal level in that band. Figure 2.1 shows a schematic representation of multiband compression using a feed-forward gain control in each band. A delay equivalent to the processing time required for gain estimation is introduced in the signal path. The delayed signal is multiplied by the gain estimated using the selected compression function and

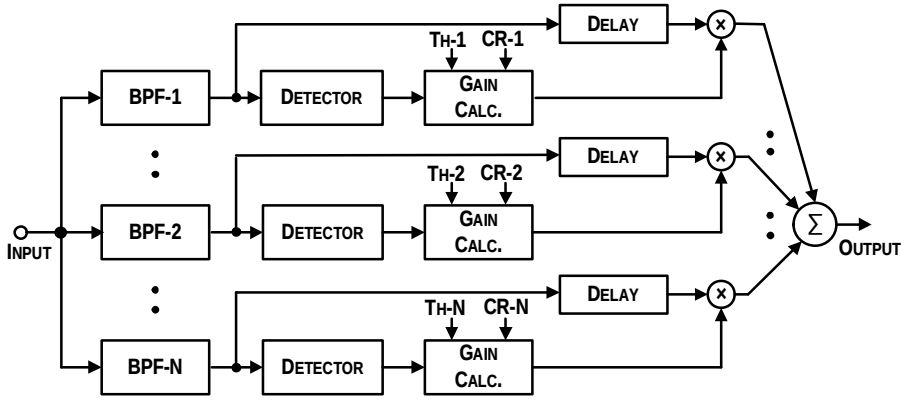


Figure 2.1 Multiband compression using feed-forward gain control with compression threshold Th and compression ratio CR (adapted from [5]).

the set compression threshold and compression ratio. The resultant signals from each band are added to produce the output.

2.5 Review of Dynamic Range Compression Techniques

Several studies using single-band dynamic range compression have been reported [11], [12], [42]–[45]. Dreschler et al. [11] compared the performance of hearing aids with linear amplification and single-band compression (compression threshold of 60 dB SPL, compression ratio of 3, attack time of 8 ms, release time of 50 ms). Listening tests, on 12 subjects with sensorineural hearing loss and using 13 sentences in quiet and in noise, resulted in a similar speech reception threshold for linear amplification and compression. Dreschler [12] compared the performance of a hearing aid having linear amplification and limiting with that having single-band compression (compression threshold of 50 dB SPL, compression ratio of 3, attack time of 6 ms, recovery time of 55 ms). Intelligibility tests, on 16 hearing-impaired listeners using 50 nonsense consonant-vowel-consonant syllables, showed that compression gave 15% higher recognition score than linear amplification and limiting. King and Martin [42] compared single-band compression (attack time of 1 ms, release time of 100 ms) with linear amplification. Listening tests were conducted on 13 hearing-impaired listeners using speech in quiet at 55, 65, 75, and 85 dB SPL and speech in presence of babble noise at 55 and 65 dB SPL with the listeners giving their preference based on clarity, comfort, pleasantness, and naturalness of the presented material. The listeners showed a preference for compression over linear amplification at higher speech levels (75 and 85 dB SPL) and no preference at lower levels (55 and 65 dB SPL). Boike and Souza [43] examined the effect of compression ratio of single-band compression on speech recognition and quality, using a compression

system (attack time of 3 ms, release time of 70 ms) with compression ratios of 1, 2, 5, and 10. Listening test, on normal-hearing listeners and listeners with mild-to-moderate sensorineural hearing loss using sentences from the Connected Speech Test [44] in quiet and in presence of babble noise, indicated that the compression ratio had no effect on speech-recognition scores in quiet, but the quality rating decreased with increase in compression ratio. In presence of noise, the quality rating as well as recognition score decreased with increase in compression ratio.

The studies using single-band compression indicate that it does not lead to significant improvement in the perceived speech quality and intelligibility. In single-band compression, the gain is calculated as a function of the signal level over its entire bandwidth. As the power in the speech signal is contributed mostly by the low-frequency components, the high-frequency components get affected by the level of the low-frequency components and may become inaudible in presence of strong low-frequency components [45], [46]. To overcome this shortcoming of single-band compression, most digital hearing aids use multiband compression. In this compression scheme, a filter bank is used to separate the input signal into several bands, level detectors are used to determine the signal level in each band, and a gain based on the input signal level and hearing loss is applied to the signal in each band. The resulting signals are added to get the compressed signal. Multiband compression may also be realized using DFT-based analysis-synthesis.

Several studies evaluating the performance of multiband compression have been reported [9], [10], [47]–[54]. Lippmann et al. [10] compared two 16-band compression schemes (C1 and C2) and four linear amplification schemes (L1, L2, L3, and L4). L1 had frequency response simulating the sound transmission path in a free field over a 1-m distance and the other three schemes provided different amounts of high-frequency emphasis. C1 used a lower compression ratio (1−3) and provided a lower high-frequency emphasis than required to restore the normal equal loudness contours. C2 was designed to restore the normal equal loudness contours for pure tones. The six schemes were implemented using sixteen one-third octave band filters with center frequencies of 160−8000 Hz, with attack time of 1.5−6 ms and release time of 20−32 ms. Intelligibility tests were conducted on five hearing-impaired subjects with moderate-to-severe sensorineural loss using nonsense consonant-vowel-consonant syllables and sentences as the test material presented in quiet and noisy environments at the most comfortable level. For speech material without large word-to-word level variations, L2, L3, and L4 gave a higher recognition score (68%) than L1 (39%). C1 performed slightly better than C2, but neither gave higher scores than the best linear scheme indicating that spectral flattening (reduction in peak to valley ratio of speech spectrum due to

highly varying gains in adjacent bands) degraded the acoustic clues important for speech perception. For speech material with large word-to-word level variations, C1 gave 7% higher score than the best linear scheme, indicating the usefulness of compression for low-level words.

Stone et al. [9] compared the effect of compression ratio, compression threshold, and attack and release times using single and four-band compressions implemented on a wearable digital hearing aid. They used a dual front-end compression with a slow system (attack time of 325 ms, release time of 1 s) that determined the normal operation and a fast system (attack time of 5 ms, release time of 75 ms) to provide protection against a sudden increase in level. They compared four compressors: dual front-end single-band compressor DUAL-HI with compression threshold of 63 dB SPL and compression ratio of 30 (fast system activated for the input level above 63 dB and the slow system output 8 dB above the mean value); dual front-end single-band compressor DUAL-LO with compression threshold of 55 dB SPL and compression ratio of 3 (fast system activated for the slow system output above a threshold); fast-acting four-band compressor FULL-4 (with the maximum compression ratio of 2.92 in a channel and compression thresholds as described in [55]); and compressor DUAL-4 as a combination of DUAL-LO and FULL-4. Listening tests were conducted on eight subjects with moderate-to-severe hearing loss using AB word lists at 50 and 80 dB SPL in quiet and using sentences in steady speech-shaped noise and amplitude modulated speech-shaped noise at 60 and 75 dB SPL. The word recognition score in quiet was higher than 90% for all four systems, indicating that the four compression systems were almost equally effective in compensating for loudness recruitment. The DUAL-LO and DUAL-HI compressor gave a better speech reception threshold for steady noise. For modulated noise, DUAL-4 performed better than the other compressors at low SNR.

It has been reported that compression schemes with several narrow bands produce more spectral distortion at the band boundaries and spectral flattening than the schemes with a small number of bands. Due to loudness recruitment accompanying sensorineural loss, a small change in the magnitude of a spectral component may lead to a large change in its percieved loudness. Therefore, the distortions due to compression may significantly degrade the speech quality for the hearing-impaired listeners with very small dynamic range. We have not come across psychophysical studies on the perception of synthetic stimuli after processing for compression or empirical listening data using speech material directly linking multiband compression artefacts and poor subjective quality. However, studies [56]–[58] using compression schemes aimed at reducing these artefacts have shown improvement in the listening test results. For reducing the distortions due to spectral discontinuities at the band

boundaries in a multiband compression, Tejero-Calado et al. [56] proposed compression based on the sinusoidal speech model. Their technique uses FFT based analysis-synthesis using a 30-ms Hamming window with 75% overlap. The speech is represented in each frame as the sum of up to six principal sinusoids, chosen as the largest peaks in the magnitude spectrum. For each sinusoid, the gain is calculated to make the ratio of peak sensation levels (number of decibels above hearing threshold) for normal-hearing and hearing-impaired listeners equal to the ratio of their respective dynamic ranges. The gain for the remaining spectrum is obtained using linear interpolation of the gains at the principal sinusoids. The technique was compared with five-band compression, each with 40-dB compression threshold. Listening tests were conducted on four listeners with a moderate hearing loss (flat and sloping) using nonsense consonant-vowel-consonant syllables, presented in quiet and noisy environments, and at the most comfortable level and soft level (10 dB below the most comfortable level). The phoneme recognition score for the proposed technique, under different listening conditions, were 4−24% higher than for the five-band compression. Rutledge et al. [57] compared the sinusoidal model based compression of [56] with two multiband compressions (one using the band RMS value for level calculation and the other using the peak spectral value). Listening tests were conducted on three listeners having a moderate flat hearing loss, using nonsense consonant-vowel-consonant syllables in steady and fluctuating noise generated with a gating frequency of 10 Hz and SNR of 5 dB. The multiband compressions showed fluctuations in the gain with the noise envelope, leading to an annoying pumping sound in the output. The scores were higher in gated noise than in steady noise for the sinusoidal model based compression, indicating a greater release from masking in gated noise.

Asano et al. [59] proposed a compression technique using an FIR filter to reduce the spectral distortions at band boundaries and to reduce flattening of the speech spectrum associated with multiband compression systems. It uses FFT-based spectral analysis of the windowed input signal for obtaining octave-band gains with compensation functions based on the relationship between the loudness for the hearing-impaired listener and that for normal-hearing listeners. An FIR filter response is obtained by interpolating the octave band gains using a spline function. The technique was compared with a linear system with a gain set as half of the hearing loss, using listening tests on 13 moderate-to-severe hearing-impaired subjects using Japanese monosyllables. The recognition scores for the proposed technique was 12% higher than that for the linear system.

To reduce the group delay associated with FIR filter based compression, Kates [51] and Kates and Arehart [60] proposed digital frequency warping to obtain a compression filter with

non-uniform frequency representation close to the auditory scale. A frequency-warped FIR filter, with the unit delays in the conventional FIR filter replaced by all-pass filters, was used to match its frequency resolution to that of the human auditory system. The compression uses an FFT-based side-branch for frequency analysis and gain computation, with compression gains applied to the signal through a frequency-warped FIR filter. Listening tests were conducted on 10 normal-hearing listeners and 11 hearing-impaired listeners, using clicks, sentences, and vowels as stimuli, to determine the optimal number of filter taps. The frequency resolution of 31-tap warped FIR filter using 32-point FFT was comparable to 128-point FFT based conventional compressor in terms of frequency resolution and resulted in inaudible delay under all listening conditions.

Although multiband compression can help in restoring near-normal loudness perception, use of a large number of bands reduces spectral contrasts and the modulation depth of speech, which may adversely affect certain speech cues. Further, different gains in adjacent bands of multiband compression may distort the spectral shape of a formant spanning the band boundaries, particularly during formant transitions [61], [62]. As it is difficult to analyze and quantify the effect of compression related distortions on the speech signal, these distortions have been demonstrated using synthesized waveforms with different tone combinations. The output of multiband compression for a time-varying narrow-band input, such as a swept sinusoid, shows unwanted peaks in the signal magnitude at the band boundaries. The power of a signal with frequency near a band boundary is split between adjacent bands and the applied gain is higher than that for signals with frequencies away from band boundaries. Lindemann [61] proposed to increase the number of bands and the overlap between bands, to minimize the distortion at the band boundaries. However, as the gain obtained using the signal power in a band is applied to all the spectral samples within the band, the resultant gain for a sinusoid at or near the boundary is higher than that at or near the center of the band. Vickers [62] proposed a multiband compression technique using FFT-based analysis-synthesis, with the frequency band boundary locations adjusted in every frame to prevent a spectral peak in the input signal from being located midway between two frequency bands. In this technique, spectral peaks corresponding to the formants are identified and the band boundaries are adjusted in each analysis frame to keep the spectral peaks away from the boundaries. However, locating these spectral peaks in each analysis frame is difficult.

As the distortions introduced by multiband compression may partly offset its advantages for the hearing-impaired listener, there is a need to develop a dynamic range compression scheme that overcomes the disadvantages of the commonly employed single-band and multiband compression schemes.

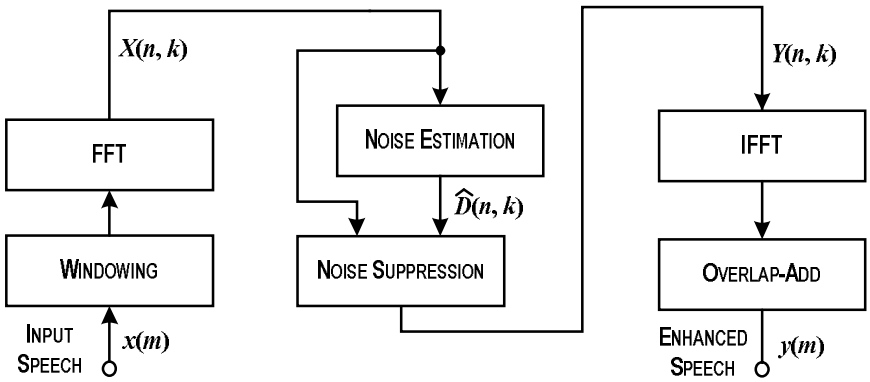


Figure 2.2 Single-channel speech enhancement using short-time spectral analysis-synthesis.

2.6 Speech Enhancement by Suppression of Background Noise

Signal processing in hearing aids for speech enhancement by suppression of background noise helps in improving speech quality and intelligibility for persons with sensorineural hearing loss. The speech enhancement technique should be effective for suppression of stationary as well as nonstationary noises. Further, it should have a low algorithmic delay and low computational complexity for implementing it using a low-power processor in a hearing aid. The techniques such as beamforming and spatial filtering can be used along with multiple microphones for improving the SNR in hearing aids [63], [64]. However, the improvement provided by them is restricted to situations where speech and interfering noise have different spatial locations. Further, power usage, computational complexity, and additional cost for matched microphones may limit the usability of multiple microphones in hearing aids. Single-input techniques do not have these restrictions. Such a technique can be used for improving speech perception as an independent enhancement stage or may be used as an additional enhancement stage after a multi-microphone technique.

Several single-input speech enhancement techniques such as spectral subtraction [65], [66], Wiener filtering [67], and minimum mean-square error short-time spectral amplitude estimation [68] have been reported. Figure 2.2 shows a general single-channel speech enhancement scheme using short-time spectral analysis-synthesis. The processing comprises windowing of the noisy input speech, FFT calculation, noise estimation, noise suppression, IFFT calculation, and overlap-add for re-synthesis of the enhanced output speech. Windowed segments of the noisy input $x(m)$ are used as the analysis frames and FFT is used to obtain the noisy spectrum, with $X(n, k)$ as the kth spectral sample of the nth frame. The noise magnitude spectrum $\widehat{D}(n,k)$ is estimated from $X(n, k)$ using a noise estimation technique. The noise

magnitude spectrum $\hat{D}(n,k)$ along with the noisy spectrum $X(n, k)$ are used for enhanced spectrum calculation using a noise suppression function to obtain the enhanced spectrum $Y(n, k)$. For effective speech enhancement, the technique should be able to track the noise spectrum accurately and the noise suppression function should not introduce perceptible distortions. The noise estimation techniques and noise suppression techniques for speech enhancement are reviewed in the following two sections.

2.7 Review of Noise Estimation Techniques

As described in the previous section, single-channel speech enhancement involves estimation of the noise spectrum from the input noisy speech signal and using the estimated noise spectrum along with a noise suppression function for speech enhancement. Due to nonstationary nature of the background noise, the noise spectrum needs to be estimated dynamically and accurately. Underestimation of the noise results in residual noise and over-estimation results in distortion leading to degraded quality and reduced intelligibility. Noise can be estimated during the speech pause identified using a voice activity detector (VAD). However, speech pause detection may not be satisfactory under low-SNR conditions and the noise may not be correctly tracked during long speech segments. Several techniques for noise estimation using the past segments of noisy speech and without using a VAD have been reported [69]–[72].

Hirsch and Ehrlicher [69] proposed two noise estimation techniques, without requiring an explicit speech pause detection and using about 0.4-s noisy speech segment. In the first technique, the noise estimate at each spectral sample is updated by first-order recursive averaging of the noisy spectral sample when the magnitude of the current spectral sample is lower than the previous noise estimate scaled by an overestimation factor. In the second technique, the peak of the histogram, dynamically calculated using 40 magnitude bins, for each sub-band is used as the noise estimate. The first and second techniques resulted in average relative estimation errors, for speech material degraded by car noise at –5 to 20 dB SNR, of 3.0−3.5% and 0.6−1.5%, respectively. Evaluation of the two techniques along with speech enhancement using spectral subtraction [73] and using noise from NOISEX database [74] at −6, 0, 6, 12, and 18 dB SNR for 10-digit recognition task by an HMM-based speech recognition system showed 10−50% improvement in recognition scores for both the techniques.

Several techniques based on minimum statistics (MS) for estimating the noise spectrum have been reported [70]–[72]. They are based on the assumption that the minimum value of

the smoothed spectrum of the noisy speech within a search window corresponds to the noise. As the minimum value is smaller than the mean value of noise, the MS-based estimate requires a bias compensation [71], [75]. Martin [71] proposed an MS-based noise estimation technique that uses a smoothing parameter to obtain the smoothed spectrum of the noisy speech and a bias compensation factor to obtain an unbiased noise estimate. The smoothing parameter is obtained using *a posteriori* SNR estimate, calculated as the ratio of previous noisy spectral sample to the previous noise estimate. For each spectral sample, the minimum is obtained from the smoothed spectra in a search window over the preceding frames. To reduce the tracking delay, the minimum is searched in small length sub-windows and updated after each sub-window. The objective measure of relative estimation error, defined as the error between the actual noise and the estimated noise with reference to the actual noise, during speech pauses established the usefulness of the time-varying smoothing parameter. Use of the technique along with a minimum mean square error log-spectral amplitude (MMSE-LSA) estimator [76], [77] for speech enhancement showed it to be effective in preserving weak sounds. The noise estimation in this technique depends on the length of the search window used for minima tracking. For a long search window, short-segment noise fluctuations in nonstationary noise may get treated as speech. A short search window may result in overestimation of the noise and attenuation of low energy speech regions.

Cohen [78] proposed the minima-controlled recursive averaging technique (MCRA), using a smoothing parameter for averaging, to prevent underestimation of noise and to avoid the use of a bias compensation factor. In this technique, the minimum for each spectral sample is obtained in a search window using storing and sorting operations. The ratio of the noisy spectral sample to the minimum is compared with a fixed threshold to obtain speech presence decision, which is smoothed to calculate the speech presence probability and used to calculate the smoothing parameter for averaging. The technique was used along with an optimally modified log-spectral amplitude (OM-LSA) estimator for speech enhancement and evaluated using sentences from TIMIT database [79] and white Gaussian, car interior, and cockpit noises from NOISEX [74] database at –5, 0, 5, and 10 dB SNR. The segmental SNR improvements for MCRA were approximately 0.1−1.4 dB higher than those for the weighted averaging technique proposed by Hirsch and Ehrlicher [69].

To make the minimum tracking robust during speech activity, Cohen [80] proposed an improved MCRA (IMCRA) technique that uses two iterations of smoothing and minimum tracking. The noise estimated in the first iteration is used to obtain an improved speech presence probability. It is used in the second iteration for smoothing the input spectrum in noise-only regions to avoid overestimation of noise during speech regions. The technique was

evaluated as in [78]. The segmental relative estimation errors under different test conditions for IMCRA (0.05−0.15) were lower than those for MS [69] (0.11−0.36). The segmental SNR improvements for IMCRA were approximately 1 dB higher than for MS [70] at low SNRs. It has been reported [81] that use of search window for minima tracking results in a delay in noise tracking, particularly in presence of nonstationary noise.

To avoid the use of search windows for minima tracking, Rangachari and Loizou [81] proposed MCRA2 as a variation of MCRA [78]. It uses a computationally efficient recursive minima tracking instead of storing and sorting operations. The ratio of the noisy spectral sample to the minimum is compared with a frequency-dependent threshold to obtain speech presence decision, which is smoothed to calculate the speech presence probability. The thresholds need to be tuned for different types of noise. The technique was evaluated along with wavelet thresholding based speech enhancement [82] and using sentences from HINT database [83] in presence of babble, factory, car, and triplet noise (concatenation of the three noises) at 5 dB SNR. The objective measures of segmental SNR and log-likelihood ratio indicated that MCRA2 performed better than the weighted averaging technique proposed by Hirsch and Ehrlicher [69] and IMCRA [80] and similar to MS [71] and quantile-based technique [13]. Subjective evaluation showed MCRA2 to be preferred over the other techniques in presence of triplet noise, establishing its suitability for use in presence of nonstationary noise.

Hendriks et al. [84] proposed a minimum mean-square error (MMSE) based noise estimation technique, with the noise estimate obtained by recursive averaging of the squared spectral sample using time and frequency dependent smoothing parameter. Ratio of the squared spectral sample to the previous noise power estimate is used as *a posteriori* SNR estimate. It is used for maximum likelihood based estimation of *a priori* SNR, which is used to calculate the smoothing parameter. This noise estimation was shown, by Gerkmann and Hendriks [85], to be similar to a VAD-based noise estimation, resulting in a biased noise estimate. They proposed a technique based on unbiased minimum mean-square error (UMMSE) using a soft speech presence probability, obtained using *a posteriori* SNR, for recursive averaging. It has lower computational complexity than MMSE as it does not require estimation of bias compensation factor. The technique was evaluated along with Weiner filter based speech enhancement and using sentences from TIMIT database [79] and synthetic and natural noises (stationary white Gaussian noise, modulated white Gaussian noise, traffic noise, nonstationary vacuum cleaner noise, and babble noise). UMMSE was compared with MMSE [84] and MS [71], using the log-error between the estimated and actual noise spectra as a measure of estimation error and the segmental SNR improvement as a measure of noise

reduction. MS resulted in highest estimation error and lowest segmental SNR improvement for nonstationary noises. The estimation error and segmental SNR improvement for UMMSE were similar to those for MMSE. UMMSE resulted in lowest estimation error for modulated white Gaussian noise, indicating its noise tracking capability.

Stahl et al. [13] reported a quantile-based technique for noise estimation from the noisy speech spectrum, without requiring prior knowledge of the probability distributions of the speech and noise signals. It is based on the observation that the signal energy in a frequency bin is low in most of the frames and high only in 10-20% frames corresponding to the voiced speech segments. The technique, with the quantiles calculated using storing and sorting operations over the spectra, was combined with Weiner filter based speech enhancement. It was evaluated using speech degraded by car noise, for digit recognition by an HMM-based speech recognition system. Use of 0.55-quantile for noise estimation resulted in maximum reduction (26%) in the word error rate. Noise estimation using quantile calculation over the preceding frames corresponding to a duration of 48 ms − 1.6 s showed that the word error rate decreased with increase in the buffer length for stationary noise, but it increased for nonstationary noise.

Evans and Mason [14] reported a time-frequency quantile-based noise estimation technique, with the noise estimated using a 0.5-s buffer. In this technique, 0.1-quantile for each frequency bin is compared with the current spectral sample for an approximate speech pause detection. The noise is estimated as the average of the 0.5-quantile and the current spectral sample in case of speech pause. Otherwise, it is estimated as the average of 0.5-quantiles of the frequency bin and two frequency bins corresponding to the neighboring valleys in the smoothed noisy spectrum. Evaluation of the technique along with spectral subtraction [65] for speech enhancement and using speech degraded with car noise showed 35% relative improvement in word recognition by a speech recognition system.

Bai and Wan [15] reported a two-pass quantile-based noise estimation technique. It uses *a posteriori* SNR estimated as a function of time and frequency using a fixed 0.21-quantile calculated over a 0.6-s buffer. The SNR estimate is used to determine a new quantile using an empirically obtained quantile-versus-SNR map for each sub-band. The technique was evaluated along with speech enhancement using the noise suppression technique reported in [86] and using speech sentences in presence of babble, car, pink, white, factory, machinegun noises from NOISEX database [74] with SNR of –6 to 12 dB. It was reported that the technique resulted in SNR improvement of 4−7 dB.

The histogram-based noise estimation [69], [87], as described earlier on page 18, is based on the assumption that the most frequently occurring value of each spectral sample of the

noisy spectrum is representative of the noise at that frequency. It involves a dynamic estimation of the histogram for each frequency sample, with a recursive or non-recursive updating of the count in each magnitude bin. There may be an overestimation of noise during long speech activity if the updating window is not long enough to include the speech as well as noise. A longer updating window may be used to avoid the noise overestimation but at the cost of poor tracking ability. Due to the computational complexity involved in dynamic estimation of the histograms and the requirement of prior selection of the magnitude bins, the histogram-based techniques for noise estimation pose severe implementation challenges for real-time processing. The quantile-based techniques ([13], [14], [15]) have been reported to perform well in presence of stationary and nonstationary noise. Estimation of the quantiles requires storing and sorting of the spectral samples, resulting in a large memory requirement and high computational complexity. Therefore, these techniques are not suitable for speech enhancement in hearing aids. Use of median, i.e. 0.5-quantile, considerably reduces the computational requirement and it has been reported to work in a robust manner [87]. Waddi et al. [88] used a cascaded-median, as an approximation to the median, in real-time speech enhancement. The technique was less effective for suppression of non-white and nonstationary noises, indicating a need for using SNR and frequency-dependent quantiles. Thus, there is a need for developing a noise estimation technique with low memory requirement and low computational complexity for single-channel speech enhancement in hearing aids for an effective suppression of nonstationary noise.

2.8 Review of Noise Suppression Techniques

Boll [65] proposed spectral magnitude subtraction as a single-input speech enhancement technique. In this technique, speech pause segments in the input signal are detected using a VAD and average of the magnitude spectra of these segments is used as the estimate of the noise spectrum. The estimated noise magnitude spectrum is subtracted from the noisy magnitude spectrum and any resulting negative values are set to zero to obtain the enhanced magnitude spectrum. The noise residual from the subtraction process sounds like a sum of tones generated with random fundamental frequencies and with pulsatile amplitude modulation. To suppress it, the enhanced magnitudes smaller than the maximum of the enhanced magnitudes during speech pause segments are replaced by the minimum of the enhanced magnitudes in the three adjacent analysis frames at the corresponding frequencies. The noise residual is further suppressed by 30-dB attenuation of the enhanced magnitudes having values 12 dB below the estimated noise. The enhanced magnitude spectrum is combined with noisy phase spectrum and the resulting complex spectrum is used to

resynthesize the speech signal. Evaluation of the technique using listening test on eight listeners, using 192 words from DRT database [89] in presence of helicopter noise, indicated improvements in quality and no change in intelligibility. Significant improvements in quality and intelligibility were reported for use of the technique as a preprocessor to an LPC-based vocoder.

The spectral subtraction results in isolated residual spectral peaks, which manifest themselves as varying tonal sounds. These sounds, known as 'musical noise', may significantly degrade the perceived speech quality. To reduce this musical noise, several variations of the spectral subtraction technique have been reported [66], [87], [90]. In the generalized spectral subtraction technique of Berouti et al. [66], the spectral subtraction is carried out using an exponent (0.5, 1, or 2) of the magnitude spectrum and the estimated noise is multiplied by a subtraction factor (3−6). To avoid musical noise, the results of subtraction are subjected to a floor, which is a fraction (0.005−0.06) of the estimated noise. The noise estimation and analysis-synthesis are similar to those in [65]. Evaluation by informal listening indicated that use of exponent as 2, corresponding to power spectrum subtraction, resulted in best speech quality. It was reported that processing resulted in improvement in speech quality and no change in speech intelligibility for inputs with −5 to 5 dB SNR.

McAulay and Malpass [91] proposed a maximum likelihood (ML) approach for estimating the enhanced magnitude. The ML-estimate is calculated as the average of the noisy magnitude and the magnitude as obtained from power spectrum subtraction. In order to decrease the residual noise and speech distortion in the output, an improved estimate of the enhanced magnitude is obtained by multiplying the ML-estimate with speech presence probability calculated using *a priori* SNR estimate (ratio of the ML-estimate to the estimated noise magnitude). The technique was evaluated along with a VAD-based noise estimation by using it as a preprocessor to an LPC vocoder and it was reported that the processing improved the speech quality.

Ephraim and Malah [68] proposed a speech enhancement technique using MMSE-based short-time spectral amplitude (MMSE-STSA) estimator along with *a priori* and *a posteriori* SNR estimates. The *a priori* SNR is estimated using ML or decision directed (DD) approaches. An improved estimate of the enhanced magnitude is obtained by multiplying the MMSE-STSA-estimate with signal presence uncertainty, calculated using *a posteriori* SNR estimate and the ratio of the MMSE-STSA-estimate to the estimated noise magnitude. The technique was evaluated along with noise estimation by averaging the noisy spectra obtained from initial 320 ms of the input and using sentences from a female and a male speaker and stationary uncorrelated additive wide-band noise at 5, 0, and −5 dB SNR. Informal listening

was used to compare MMSE-STSA-ML (MMSE-STSA with *a priori* SNR estimated using ML), MMSE-STSA-DD (MMSE-STSA with *a priori* SNR estimated using DD), MMSE-STSA-DD-SPU (MMSE-STSA-DD with signal presence uncertainty), Weiner-ML (Weiner filter with *a priori* SNR estimated using ML), Weiner-DD (Weiner filter with *a priori* SNR estimated using DD), spectral subtraction [67], and the McAulay-Malpass technique [91]. It was reported that MMSE-STSA-DD-SPU resulted in best speech enhancement with a significant noise reduction. Ephraim and Malah [76] reported an MMSE-based log-spectral amplitude estimator (MMSE-LSA) to reduce the residual noise, with the evaluation as in [68]. It was reported that MMSE-LSA resulted in lower residual noise than MMSE-STSA-ML and MMSE-STSA-DD and similar to MMSE-STSA-DD-SPU.

The main advantage of spectral subtraction techniques for real-time speech enhancement is their low computational complexity. These techniques are based on the assumption that the speech and noise are uncorrelated and the cross terms due to the phase difference between the speech and noise spectra are zero. Violations of this assumption in short-time spectra of noisy speech lead to excessive musical noise [87]. Lu and Loizou [90] proposed a technique, named as 'geometric approach (GA) to spectral subtraction', for reducing the musical noise associated with power spectrum subtraction. To suppress the effect of cross terms in the processed output, this technique uses a suppression function calculated using the *a priori* and *a posteriori* SNR estimates and the input spectrum is multiplied with the suppression function to obtain the enhanced spectrum. The processing resulted in no audible musical noise. The technique was evaluated for speech enhancement along with noise estimation using the MS technique [71] and using sentences from NOIZEUS database [92] and babble, street, and car noise from AURORA database [93] at 0, 5, and 10 dB SNR. The PESQ [94] scores for GA (1.76−2.53) were higher than those for spectral subtraction [66] (1.66−2.37) and comparable to those for MMSE-STSA [68] (1.76−2.74).

2.9 Scope of the Research

Frequency-selective amplification and dynamic range compression form the core of signal processing in hearing aids. A review of these techniques has been presented in the preceding sections.

Dynamic range compression in hearing aids is provided with the objective of presenting all the sounds comfortably within the limited dynamic range of the listener. Single-band and multiband compressions are commonly employed in hearing aids. The studies using single-band compression indicate that it leads to reduced high-frequency audibility and does not lead to significant improvement in the perceived speech quality and intelligibility. To overcome

the problem associated with single-band compression, most digital hearing aids use multiband compression. Although multiband compression can help in restoring near-normal loudness perception, use of a high number of bands reduces spectral contrasts and the modulation depth of speech, adversely affecting the perception of certain speech cues. Further, different gains in adjacent bands of multiband compression may distort the spectral shape of a formant spanning the band boundaries, particularly during formant transitions. The distortions in the speech signal introduced by multiband compression may partly offset its advantages for the hearing-impaired listener. Thus, there is a need to develop a dynamic range compression scheme that overcomes the disadvantages of the commonly employed single-band and multiband compression schemes.

Signal processing in hearing aids for speech enhancement by suppression of background noise can be used to improve speech quality and intelligibility for persons with sensorineural hearing loss. The processing technique should be effective for suppression of stationary as well as nonstationary noises. Further, it should have a low algorithmic delay and low computational complexity for implementing it on a low-power processor in a hearing aid. Single-input techniques can be used for improving speech perception as an independent enhancement stage or may be used as an additional enhancement stage after one of the multi-microphone techniques. They involve estimation of the noise spectrum from the input noisy speech signal and using the estimated noise spectrum along with a noise suppression function for speech enhancement. Underestimation of the noise results in residual noise and over-estimation results in distortion leading to degraded quality and reduced intelligibility. Quantile-based and histogram-based techniques for noise estimation from the noisy speech spectrum have been reported to perform well in presence of stationary and nonstationary noise. They do not require prior knowledge of the probability distributions of the speech and noise signals. However, they are not currently suitable for use in hearing aids due to large memory requirement and high computational complexity involved in storing and sorting the spectral samples. Thus, there is a need to develop a noise estimation technique with low memory requirement and low computational complexity that is effective for suppression of stationary as well as nonstationary noises.

The objective of the current research is to develop signal-processing techniques for dynamic range compression and background noise suppression, in order to overcome the shortcomings of the currently employed techniques and thus to enhance the performance of hearing aids used by listeners with sensorineural loss. The techniques are developed with considerations for low memory requirement and computational complexity for implementation using a low-power processor in a hearing aid, low input-output delay for

acceptability of the signal processing for face-to-face communication, and low perceptible distortions.

Towards the above research objective, two signal-processing techniques are developed for use in hearing aids: (i) sliding-band dynamic range compression to compensate for frequency-dependent loudness recruitment, and (ii) speech enhancement using dynamic quantile tracking for estimation of background noise. The techniques are implemented for offline processing and evaluations are carried out using different test materials and comparisons with some of the existing techniques. Subsequently, the techniques are implemented for real-time processing using a fixed-point DSP chip, to assess their suitability for use in hearing aids in terms of memory and computational requirements and input-output signal delay. To enable the use and evaluation of these techniques by a large number of users without incurring the expenses involved in the ASIC-based hearing aid development, a smartphone app implementing the techniques with an interactive touch-controlled graphical user interface is also developed.

Chapter 3

SLIDING-BAND DYNAMIC RANGE COMPRESSION

3.1 Introduction

Dynamic range compression in hearing aids is provided with the objective of presenting the sounds comfortably within the limited dynamic range of the listener without introducing significant perceptible distortions. A review of single-band and multiband dynamic range compression techniques has been presented in Section 2.5 of the second chapter. In single-band compression, the gain is calculated as a function of the signal power over its entire bandwidth. As the power in the speech signal is contributed mostly by the low-frequency components, the high-frequency components get affected by the level of the low-frequency components and may become inaudible in presence of strong low-frequency components. As a solution to this problem, several multiband compression techniques have been developed. In these techniques, the spectral components of the input signal are divided into multiple bands and the gain for each band is calculated on the basis of signal power in that band. These techniques avoid the problems associated with single-band compression, but result in decreased spectral contrasts and modulation depths in the speech signal. The bands in multiband compression get narrower with increase in the number of bands, and hence more than 8−16 bands are not used. Further, different gains in adjacent bands may distort the spectral shape of a formant spanning the band boundaries, particularly during formant transitions.

In order to reduce the temporal and spectral distortions associated with the currently used single-band and multiband compression techniques, a technique referred to as 'sliding-band compression' is proposed in Section 3.2. The implementations of the technique for offline processing and real-time processing along with corresponding test results are presented in Section 3.3 and Section 3.4, respectively, followed by discussion in the last section.

3.2 Sliding-Band Dynamic Range Compression

The proposed technique is aimed at compensating for frequency-dependent loudness recruitment associated with sensorineural hearing loss without introducing the distortions generally associated with the single-band and multiband compression techniques. The technique uses a frequency-dependent gain function, with the gain for each spectral sample

calculated based on the short-time power in a band centered at its frequency. The bandwidth of the 'sliding' band is selected to approximate the frequency resolution of the auditory system, varying from a small value at the low-frequency end to a large value at the high-frequency end. The bandwidth can be selected as one-third octave bandwidth, bandwidth corresponding to equal increments on the mel scale, or auditory critical bandwidth. As the gain for a spectral component is determined by the spectral components located within a band centered at its frequency, the proposed technique avoids the possibility of attenuation of high-frequency components due to the presence of strong low-frequency components, which may occur in single-band compression. Use of the band sliding with the frequency results in a smooth magnitude response. The technique provides a smooth magnitude response, but the operation is different from a smoothing operation on the frequency response of multiband compression as described by Asano et al. [59]. As there are no discrete bands for gain calculation, the relationship between the gain and signal level is independent of the position of the spectral component with respect to the band center frequency. The proposed technique avoids distortions in the shape of spectral resonances and discontinuities during the resonance transitions, which may occur in multiband compression.

The proposed technique is referred to as sliding-band compression, as the gain for each spectral sample is calculated based on the signal level in an auditory critical band centered at its frequency. Thus, the gain calculated for each spectral sample depends not only on the level corresponding to that spectral sample, but also on the level of the neighboring spectral samples. The technique is aimed at avoiding the distortions at the band boundaries and spectral contrast reduction, commonly associated with multiband compression with large number of bands. The main downside of the technique is its high computational complexity, but it may be acceptable in an FFT-based realization, particularly if the FFT computation is shared with other processing steps in the hearing aid.

A target gain is calculated for each spectral sample based on the short-time power in the band centered at it and in accordance with the selected compression function. The compression function may be calculated from the specified parameters. Use of a look-up table relating the target gain to the band power may be used to reduce the computation and to provide a frequency-dependent compression function most suited to compensate for the abnormal loudness growth curve of the hearing-impaired listener. The target gain for each spectral sample is used to update the gain in accordance with the specified attack and release times, resulting in time-varying frequency response with the magnitude response being smooth along time and frequency axes.

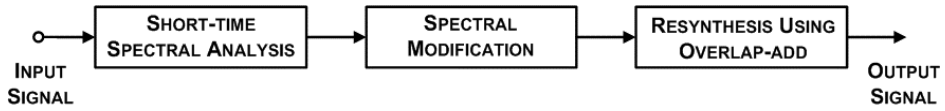


Figure 3.1 Sliding-band dynamic range compression using spectral modification.

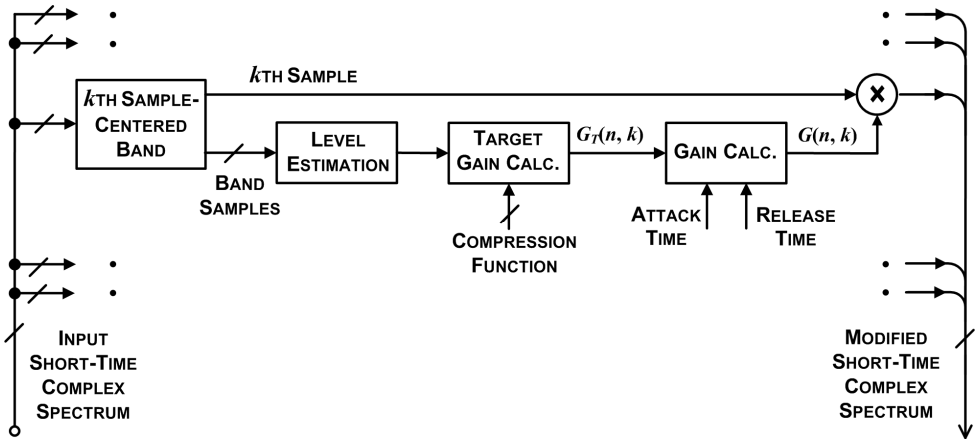


Figure 3.2 Spectral modification for compensation of increased hearing thresholds and decreased dynamic range using sliding-band dynamic range compression.

The proposed technique is implemented as a feed-forward compression system using short-time spectral analysis and synthesis. A block diagram of the signal processing, comprising the steps of short-time spectral analysis, frequency and level dependent spectral modification, and signal resynthesis, is shown in Figure 3.1. The analysis involves segmentation of the input signal into overlapping frames and calculating FFT to get short-time complex spectra. Spectral modification for dynamic range compression consists of frequency-dependent gain calculation and using it for calculating the modified complex spectrum. The output signal is resynthesized using IFFT, windowing, and overlap-add.

The processing for spectral modification is shown in Figure 3.2. For each discrete frequency sample *k* of the input complex spectrum, the processing path for calculating the frequency-dependent gain comprises the steps of level estimation, target gain calculation, and gain calculation. For compression with bands based on auditory critical bandwidths [95], the bandwidth in kHz at the frequency sample *k* can be approximated as

$$\mathrm{BW}(k) = 25 + 75(1 + 1.4(f(k))^2)^{0.69} \tag{3.1}$$

where $f(k)$ is the frequency of the *k*th sample in kHz. The level estimation involves calculation of the input power in the *n*th frame as the sum of squared magnitude of the spectral samples in the band centered at *k*. A frequency-dependent compression function, relating the output power to the input power, is used to calculate the target gain. The compression function is selected in

accordance with the desired hearing aid fitting procedure (as described in Appendix A) to compensate for the abnormal loudness growth. To reduce the computational requirement, the target gain may be obtained as a function of frequency and input power using a two-dimensional look-up table. In the gain calculation step, the present gain value is calculated as a smooth change from the previous value towards the target gain value using ratio steps to avoid distortions due to sudden gain changes. The gain applied to the *k*th spectral sample in the *n*th frame is obtained using set values of attack and release times by updating the gain from the previous value towards the target value, $G_T(n, k)$, and is given as

$$G(n,k) = \begin{cases} \max(G(n-1,k)/\gamma_a, G_T(n,k)), & G_T(n,k) < G(n-1,k) \\ \min(G(n-1,k)\gamma_r, G_T(n,k)), & G_T(n,k) \geq G(n-1,k) \end{cases} \quad (3.2)$$

The number of steps during the attack and release phases are controlled using gain ratios $\gamma_a = \left(G_{\max} / G_{\min}\right)^{1/s_a}$ and $\gamma_r = \left(G_{\max} / G_{\min}\right)^{1/s_r}$, respectively. Here $G_{\max}$ and $G_{\min}$ are the maximum and minimum values of the gains in the compression segment of the compression function and s_a and s_r are the number of steps during the attack and release, respectively.

The processing for sliding-band compression modifies the magnitude spectrum and the signal is resynthesized using the original phase spectrum. It may result in distortions due to phase discontinuities in the modified short-time complex spectrum. To mask these distortions, the analysis-synthesis method based on the least squares error estimation (LSEE) as proposed by Griffin and Lim [96] is used. It involves segmenting the input signal using *L*-sample frames with 75% overlap, i.e. frame shift $S = L/4$, and multiplying the segmented frames with modified Hamming window proposed in [96]. Complex spectrum is obtained by zero padding the *L*-sample frame to length *K* and calculating *K*-point FFT. After spectral modification as described earlier, *K*-point IFFT is calculated to get the modified output frame that is multiplied with the modified Hamming window. The successive output frames are added in accordance with the overlap of the input frames to provide the resynthesized output signal.

3.3 Implementation for Offline Processing and Test Results

3.3.1 Implementation for offline processing

For comparing the performance of the proposed sliding-band compression with single-band compression and multiband compression, the three compression techniques were implemented for offline processing. Spectral modification for single-band compression involved level estimation in a single-band comprising all the spectral samples of the input spectrum, gain calculation, and multiplication of the input spectrum with the gain to get the modified spectrum. Multiband compression was implemented with band frequencies

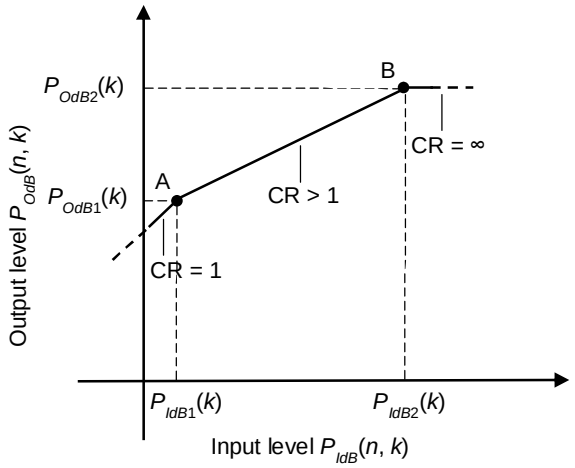


Figure 3.3 Example of compression function relating the output level (dB) to the input level (dB) and for *n*th frame and band centered at *k*th spectral sample.

corresponding to 18 critical bands [97]. Spectral modification for multiband compression involved level estimation using the spectral samples located within a band, gain calculation, and multiplication of spectral samples within the band with the calculated gain to obtain the modified spectral samples. In the implementation of sliding-band compression, the gain for each spectral sample was calculated on the basis of power in the auditory critical band centered at its frequency as described in Section 3.2.

The three compression techniques were implemented using a compression function with a piecewise linear three-segment relation between input level $P_{IdB}(n, k)$ and the output level $P_{OdB}(n, k)$ on a dB scale, as shown in Figure 3.3. The compression threshold and the output-limiting threshold are marked as the point A (P_{IdB1}, P_{OdB1}) and the point B (P_{IdB2}, P_{OdB2}), respectively. The gain for amplification in dB in terms of P_{IdB1}, P_{OdB1}, P_{IdB2}, and P_{OdB2} is given as

$$G_{TdB}(n,k) = \begin{cases} P_{OdB1}(k) - P_{IdB1}(k), & P_{IdB}(n,k) < P_{IdB1}(k) \\ P_{OdB1}(k) + \dfrac{P_{OdB2}(k) - P_{OdB1}(k)}{P_{IdB2}(k) - P_{IdB1}(k)} (P_{IdB}(n,k) - P_{IdB1}(k)) - P_{IdB}(n,k), & P_{IdB1}(k) < P_{IdB}(n,k) < P_{IdB2}(k) \\ P_{OdB2}(k) - P_{IdB}(n,k), & P_{IdB2}(k) < P_{IdB}(n,k) \end{cases} \tag{3.3}$$

The compression ratio (CR) is defined as the ratio of the change in the input level to the change in the output level. Its value for the compression-amplification segment is given as

$$\mathrm{CR}(k) = \frac{P_{IdB2}(k) - P_{IdB1}(k)}{P_{OdB2}(k) - P_{OdB1}(k)} \tag{3.4}$$

The gain for amplification can also be expressed in terms of P_{IdB1},P_{IdB2}, P_{OdB2}, and CR(k) as

$$G_{TdB}(n,k) = \begin{cases} P_{OdB2}(k) - \dfrac{(P_{IdB2}(k) - P_{IdB1}(k))}{\mathrm{CR}(k)} - P_{IdB1}(k), & \\ & P_{IdB}(n,k) < P_{IdB1}(k) \\ P_{OdB2}(k) - \dfrac{(P_{IdB2}(k) - P_{IdB}(n,k))}{\mathrm{CR}(k)} - P_{IdB}(n,k), & \\ & P_{IdB1}(k) < P_{IdB}(n,k) < P_{IdB2}(k) \\ P_{OdB2}(k) - P_{IdB}(n,k), & P_{IdB2}(k) < P_{IdB}(n,k) \end{cases} \tag{3.5}$$

For each of the three compression techniques, the gain was obtained using the target gain and the set attack and release times. For the attack time T_a, release time T_r, sampling frequency f_s, and window shift S samples, the number of steps s_a during the attack and the number of steps s_r during the release were set as the following:

$$s_a = T_a f_s / S \tag{3.6}$$

$$s_r = T_r f_s / S \tag{3.7}$$

The processing was carried out using a sampling frequency f_s of 10 kHz and window length L of 256, resulting in 25.6 ms segments. A 75% overlap-add was used, corresponding to a window shift S of 64. Analysis-synthesis was carried out using K-point FFT with K = 512. A sinusoid with the root-mean-square (RMS) value of 1 (i.e. peak value of $\sqrt{2}$) was used as the reference for calculation of level on dB scale. The static characteristics of the compression techniques were selected by setting {$P_{IdB1}(k)$, $P_{IdB2}(k)$, $P_{OdB2}(k)$, CR(k)} as {−12 dB, 0 dB, 0 dB, 1−10}. Two types of dynamic characteristics were used for testing, first with fast attack and fast release and second with fast attack and slow release. The fast attack and fast release was implemented with s_a = s_r = 1, corresponding to T_a = T_r = 6.4 ms. The fast attack and slow release were implemented with s_a = 1 and s_r = 30, corresponding to T_a = 6.4 ms and T_r =192 ms.

3.3.2 Test material and evaluation method for offline processing

The compression techniques were tested to examine the difference in their processed outputs, using inputs consisting of single-tone, two-tone, and speech signals. The single-tone input, comprising a sine wave with constant amplitude and time-varying frequency, was used to examine the effect of frequency variation on the output level. The frequency of the sine wave with amplitude as 0.6 (signal level of −8 dB) was linearly swept from 100 Hz to 4900 Hz. The two-tone input comprised two sine waves, a wave of low frequency f_1 with a time-varying amplitude and a wave of high frequency f_2 with a constant amplitude. It was used to examine the effect of the level of the low-frequency component on amplification of the high-frequency

component. The f_1-tone amplitude was varied in the range 0.1−2 with the f_2-tone amplitude as 0.2. Evaluation was carried out using two such two-tone inputs, the first input with f_1 as 570 Hz and f_2 as 2510 Hz and the second input with f_1 as 500 Hz and f_2 as 2510 Hz. The tone frequencies were selected to examine the effect of the location of the f_1-tone frequency with respect to the bands as used in multiband compression, with the f_1-tone in the first input at the band center and that in the second input at the band boundary. The f_2-tone was at the band center for both the two-tone inputs. The processing techniques were also tested using inputs comprising the speech signal modulated with different amplitude envelopes.

The spectrograms of the inputs and corresponding processed outputs were visually examined for undesirable level changes in the outputs of the three compression techniques. The error in the tone-level was quantified as the level error in dB of the processed output and given as

$$\left[\text{Level error}\right]_{\text{dB}} = 20\ \log_{10}\left(\frac{\text{RMS}_{\text{output}}}{\text{RMS}_{\text{expected}}}\right) \tag{3.8}$$

where $\text{RMS}_{\text{output}}$ and $\text{RMS}_{\text{expected}}$ are the RMS values of the tone in the processed and expected outputs, respectively. $\text{RMS}_{\text{output}}$ was measured after stabilization of the output and was not affected by the set attack and release times. $\text{RMS}_{\text{expected}}$ was obtained by multiplying the input RMS with the target gain as given in (3.5). The error measure is 0 for $\text{RMS}_{\text{output}} = \text{RMS}_{\text{expected}}$, positive for $\text{RMS}_{\text{output}} > \text{RMS}_{\text{expected}}$, and negative for $\text{RMS}_{\text{output}} < \text{RMS}_{\text{expected}}$.

3.3.3 Test results

The processed outputs for a single-tone with swept frequency as the input and compressions implemented using fast attack and fast release ($T_a = T_r = 6.4$ ms) and CR as 10 are shown in Figure 3.4, with the input and output waveforms and corresponding spectrograms for a visual comparison. Figure 3.4(a) shows the single-tone input with its frequency linearly swept from 125 Hz to 250 Hz over 200 ms and a constant amplitude. The output of single-band compression, shown in Figure 3.4(b), does not exhibit amplitude variation. Figure 3.4(c) shows the output of multiband compression, exhibiting amplitude variation during the tone frequency transition over a band boundary. The output of sliding-band compression, shown in Figure 3.4(d), does not exhibit amplitude variation. Similar results were obtained for different swept tones and narrowband noises with swept center frequencies. These results confirm that the sliding-band compression is successful in avoiding the level variations that occur in multiband compression.

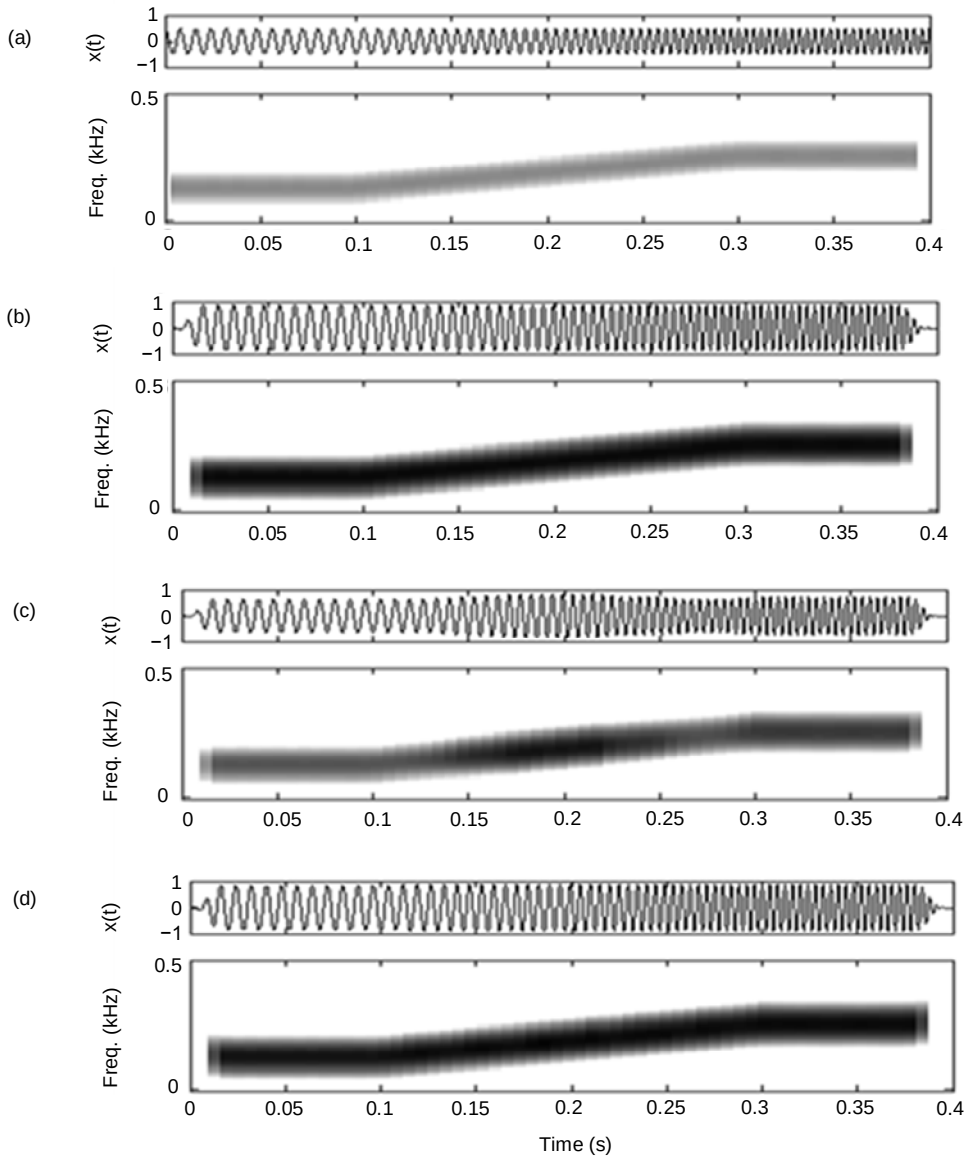


Figure 3.4 Example of offline processing of a single-tone input with constant amplitude and linearly-swept frequency (125−250 Hz over 200 ms): Waveforms and spectrograms of (a) input, (b) output of single-band compression, (c) output of multiband compression, and (d) output of sliding-band compression.

The error in the tone level for a single-tone input was quantified using the level error as given in (3.8) for the tone frequency varied from 100 Hz to 4900 Hz, and CR as 10. The plots of the level error in dB as a function of tone frequency are shown in Figure 3.5. The error for single-band compression is zero at all frequencies. For multiband compression, the error increases at frequencies close to band boundaries. This error increase occurs because a higher gain is applied to the tone when the frequency is close to a band boundary due to splitting of

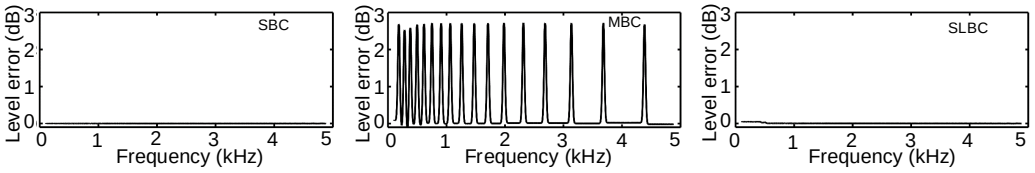


Figure 3.5 Level error (dB) in compression output for single-tone input with the frequency swept from 100 Hz to 4900 Hz, and CR of 10.

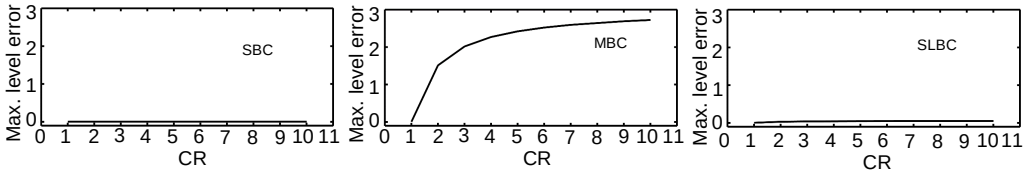


Figure 3.6 Maximum of the level error (dB) in compression output for single-tone input with the frequency swept from 100 Hz to 4900 Hz, and CR of 1−10.

the signal power between adjacent bands. The error for the sliding-band compression is zero at all frequencies, as in the case of single-band compression. The maximum error for multiband compression is a function of CR. Figure 3.6 shows the maximum error for CR of 1−10. For multiband compression, the error increases from 0 dB at CR of 1 to 1.5 dB at CR of 2 and 2.5 dB at CR of 10. The errors for single-band and sliding-band compressions are zero at all CR values. Thus, the results show that the multiband compression results in errors that vary with frequency and increase with CR. The single-band and sliding-band compressions do not result in these errors.

As an example of processing of a two-tone input, Figure 3.7 shows the spectrograms of the input and the processed outputs, with the input applied with f_1 as 570 Hz and f_2 as 2510 Hz. The compressions were implemented using T_a and T_r as 6.4 ms and CR as 10. Figure 3.7(a) shows the two-tone input with the f_1-tone amplitude varied as 0.1−2 over 200 ms and the f_2-tone amplitude constant at 0.2. In case of single-band compression, as seen in Figure 3.7(b), the f_2-tone output level decreases with increase in the f_1-tone input level. In case of multiband compression and sliding-band compression, as seen in Figure 3.7(c) and Figure 3.7(d), respectively, the f_2-tone output level does not show variation with increase in the f_1-tone input level. These results show that the high-frequency tone in the output gets affected by the variation in the low-frequency tone level in the input in the case of single-band compression but not in the cases of multiband and sliding-band compressions.

The error in the tone level for two-tone inputs was quantified using the level error as given in (3.8). The plots of the errors in the output level of f_1-tone and f_2-tone as a function of the

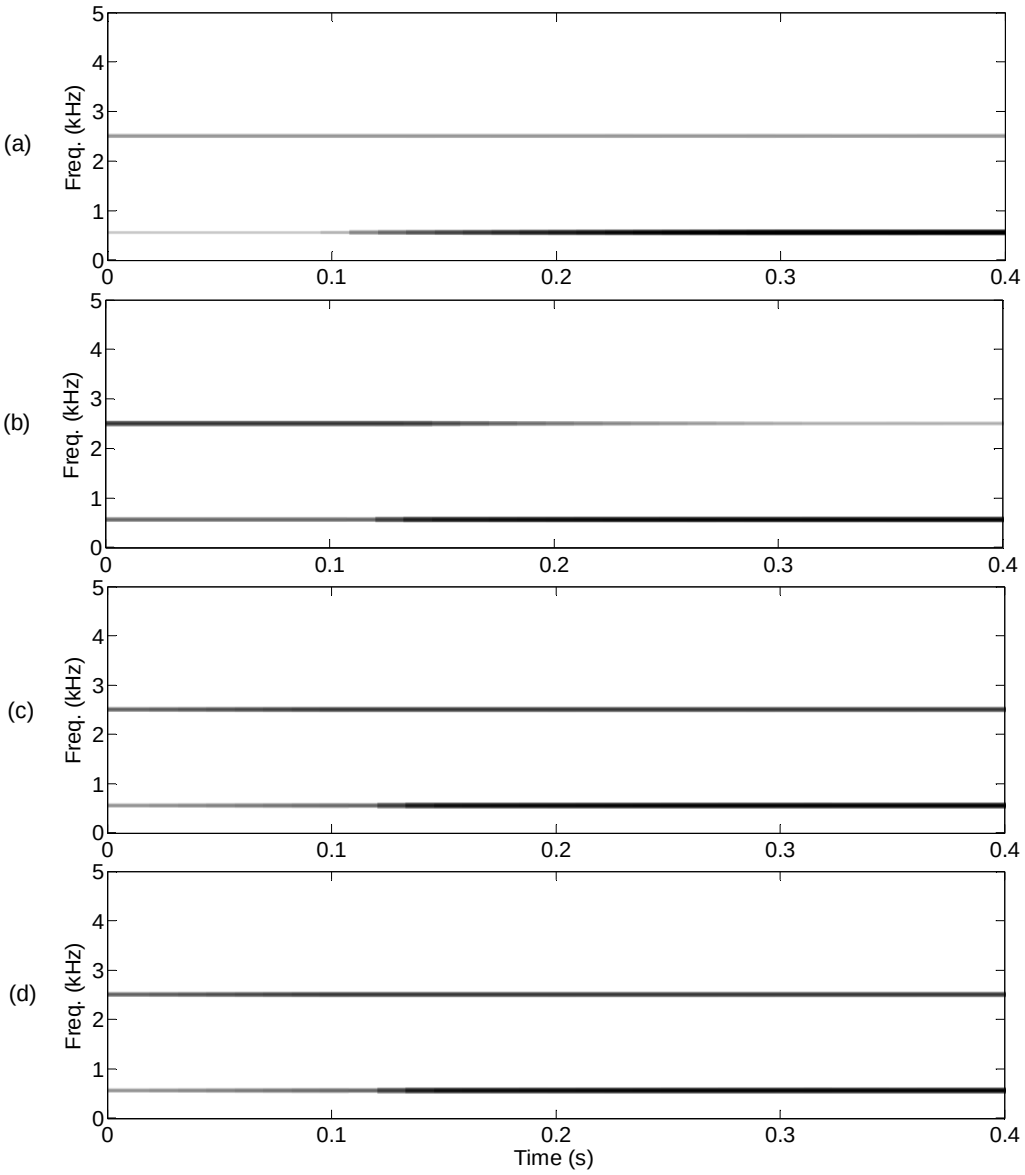
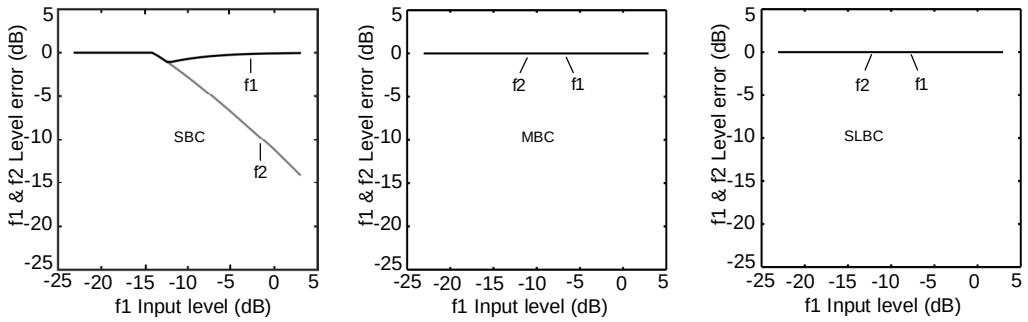
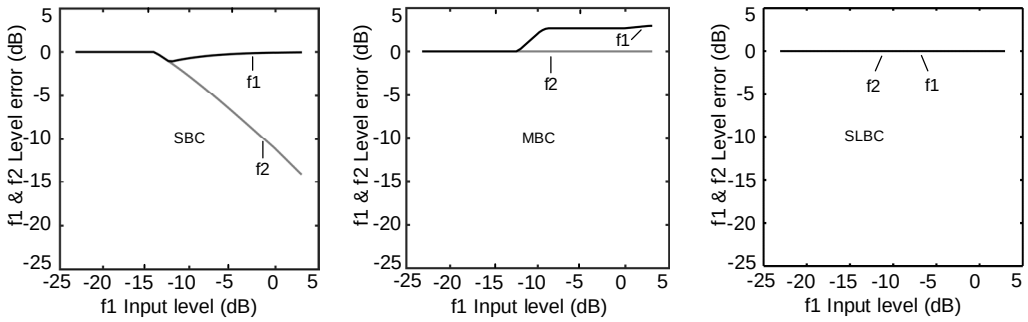


Figure 3.7 Example of offline processing of a two-tone input with f_1 as 570 Hz and f_2 as 2510 Hz with the f_1-tone varied as 0.1−2 over 200 ms and the f_2-tone amplitude constant as 0.2: Spectrograms of (a) input, (b) output of single-band compression, (c) output of multiband compression, and (d) output of sliding-band compression.

input level of f_1-tone for the three compression techniques are shown in Figure 3.8 for two f_1-f_2 combinations and CR as 10. Figure 3.8(a) shows the plots of the errors with f_1 as 570 Hz and f_2 as 2510 Hz, which are at band centers. The errors in the output level of the f_1-tone and f_2-tone are zero for multiband and sliding-band compression for all input levels of the f_1-tone, showing that the high-frequency tone is not affected by the variation in the low-frequency tone level. For single-band compression, the error in the output level of the f_1-tone is zero at low-



(a) f_1 = 570 Hz and f_2 = 2510 Hz (both at band centers)



(b) f_1 = 500 Hz and f_2 = 2510 Hz (f_1 at a band boundary and f_2 at a band center)

Figure 3.8 Level error (dB) in output for two-tone input and CR as 10: (a) f_1 = 570 Hz and f_2 = 2510 Hz (both at band centers) and (b) f_1 = 500 Hz and f_2 = 2510 Hz (f_1 at a band boundary and f_2 at a band center).

level, corresponding to the linear region of the compression function. The error becomes negative with increase in the input level of the f_1-tone at the onset of the compression. This error occurs because the compression gain is affected by f_2-tone level. With further increase in f_1-tone input level, the effect of f_2-tone input level becomes less significant and the error in the f_1-tone output level becomes zero. The error in the f_2-tone output level becomes progressively more negative with increase in the f_1-tone input level. Figure 3.8(b) shows the plots of the errors with f_1 as 500 Hz and f_2 as 2510 Hz, with f_1 at a band boundary and f_2 at a band center. The error for single-band compression is similar to that in the earlier case. The multiband compression shows a positive error in the output level of the f_1-tone. This error increase occurs because a higher gain is applied to the f_1-tone when its frequency is close to a band boundary due to splitting of the signal power between adjacent bands. There are no errors for sliding-band compression. These results confirm that the sliding-band compression avoids level variation in the spectral components that may be caused by the single-band and multiband compressions.

The results using three compression techniques with CR as 2 (for all spectral samples) and fast attack time and slow release time (T_a = 6.4 ms, T_r = 192 ms) are shown in Figure 3.9. The

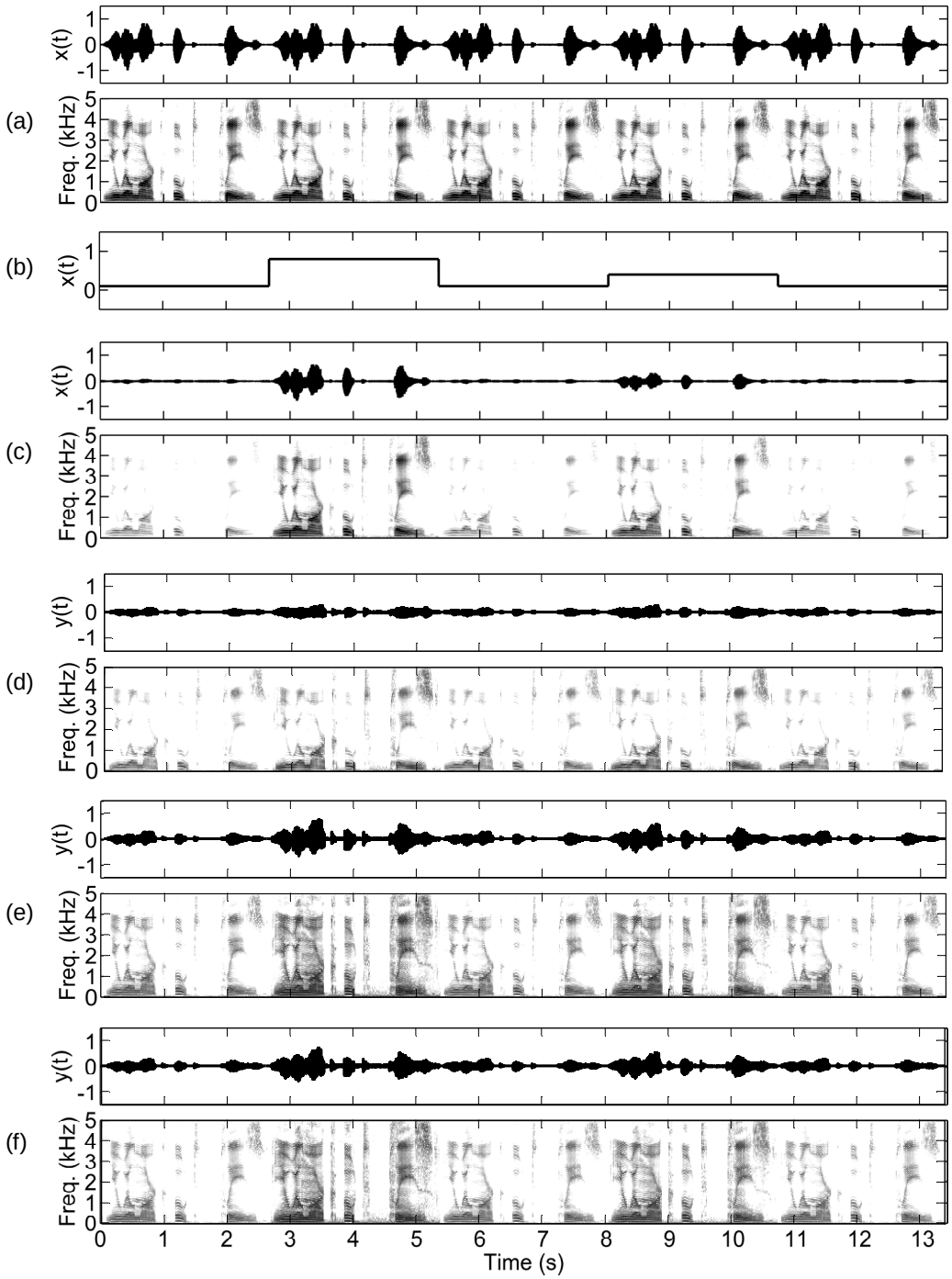


Figure 3.9 Example of offline processing, with fast attack and fast release (T_a = 6.4 ms, T_r = 192 ms) and CR as 2 (for all spectral samples), of sentence "you will mark ut please" concatenated with scaling factors of 0.1, 0.8, 0.1, 0.4, 0.1: (a) speech signal, (b) scaling factor, (c) input signal (speech signal multiplied by the scaling factor), (d) single-band compression output, (e) multiband compression output, and (f) sliding-band compression output.

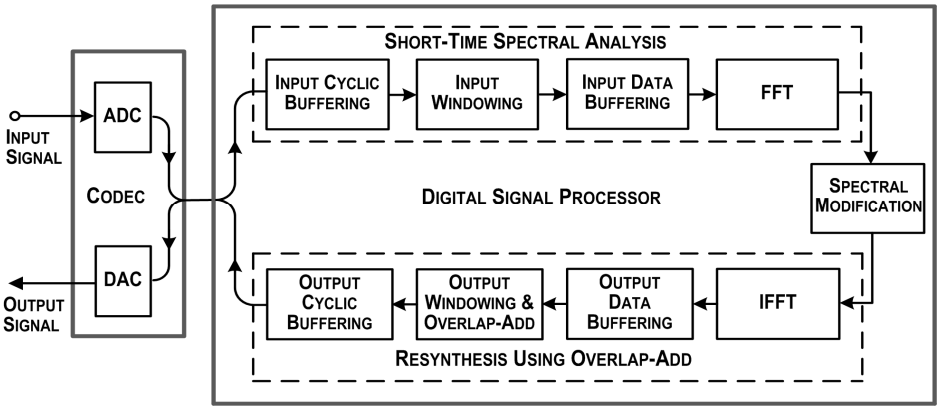


Figure 3.10 Implementation of sliding-band dynamic range compression for real-time processing using a DSP board.

input speech material consists of an English sentence "*you will mark ut please*" concatenated with different scaling factors to observe the effect of variation in the input level on the output waveform. It can be observed from the figure that as the input level increases, the gain decreases. In single-band compression, the gain is calculated as a function of the signal power over its entire bandwidth and hence the compression starts at a lower level. The outputs from multiband and sliding-band compression show the desired amplification and compression.

3.4 Implementation for Real-Time Processing and Test Results

3.4.1 Implementation for real-time processing

The technique has been implemented for real-time processing using a low-power DSP chip. For this purpose, a DSP board based on the 16-bit fixed-point processor TI/TMS320C5515 [98], with a maximum clock rate of 120 MHz and address space of 16 MB with 320 KB on-chip RAM (including 64 KB dual access RAM), and 128 KB on-chip ROM, is used. The chip has three 32-bit programmable timers, four DMA controllers each with four channels, and a tightly coupled FFT hardware accelerator supporting 8–1024 point FFT. The DSP board 'eZdsp' [99], used for the implementation, has 4 MB on-board NOR flash for user program and codec TLV320AIC3204 [100] with stereo ADC and DAC supporting 16/20/24/32-bit quantization and sampling frequency of 8–192 kHz. The program was written in C, using TI's 'CCStudio, ver. 4.0' as the development environment.

A block diagram of the implementation is shown in Figure 3.10. The input signal is acquired using ADC and the processed signal is output using DAC of the left channel of the codec, with 16-bit quantization and 10 kHz sampling. The digital samples from ADC are applied as input to the short-time spectral analysis comprising input cyclic buffering, windowing, zero-padding, and FFT. The output from the short-time spectral analysis is given

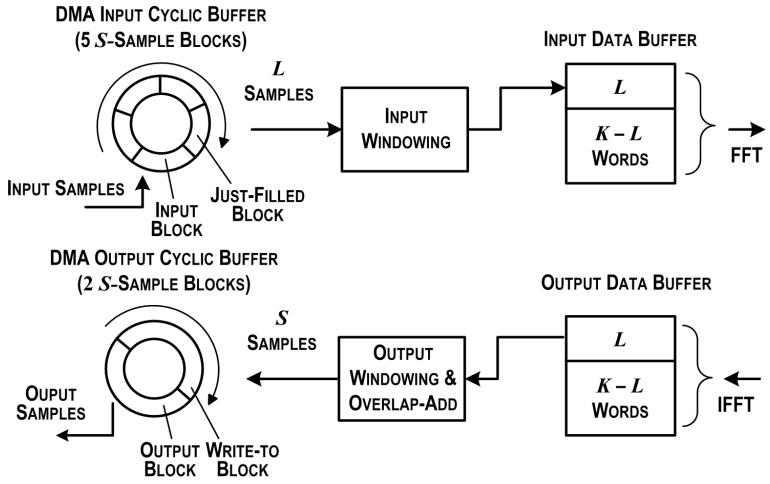


Figure 3.11 Data transfer and buffering on the DSP board ($S = L/4$) used in the real-time implementation of Figure 3.10.

as input to the spectral modification, which comprises frequency-dependent gain calculation and calculation of the modified spectrum. The modified spectrum is used for resynthesis using IFFT, output data buffering, output windowing, overlap-add, and output cyclic buffering. The digital signal obtained after overlap-add is stored in the output cyclic buffer and output through DAC.

Figure 3.11 shows the data transfer and buffering operations of the implementation. To reduce the conversion overheads, the input samples, the spectral values, and the processed samples are all stored as 4-byte words with 16-bit real and 16-bit imaginary parts. A 5-block DMA input cyclic buffer, with *S*-word blocks, is used for signal acquisition. An input data buffer of *K* words is initialized with zero values. Cyclic pointers are used to track the 'current input', 'just-filled input', 'current output', and 'write-to output' blocks. The pointers are initialized to 0, 4, 0, and 1, respectively. When an input block gets filled, a DMA interrupt is generated and all pointers are incremented. Input window with *L* samples is formed using the samples of the just-filled and the previous three blocks. These *L* samples multiplied by modified-Hamming window of length *L* are copied to the input data buffer. They are padded with $K-L$ zero-valued samples to serve as input to *K*-point FFT. This method of data handling results in an efficient realization of 75% overlap and zero padding. The complex spectrum obtained after the FFT calculation is used for spectral modification. The output of the *K*-point IFFT of the modified complex spectrum is copied to the output data buffer. The first *L* samples of the output data buffer are multiplied by the modified-Hamming window to get the time domain segment and is used as input for overlap-add operation to synthesize *S* samples of the output signal. The overlap-add operation uses a buffer of 3*S* 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. It may be noted that the processing has to get completed in S sampling intervals for real-time operation.

A two-dimensional look-up table is used for target gain calculation in accordance with the short-time spectrum of the signal. It reduces the computational requirement and permits use of a frequency-dependent compression function most suited to compensate for the abnormal loudness growth function of the hearing-impaired listener. The gain is calculated using (3.2), with attack and release times as set using s_a and s_r. Modified spectrum is obtained by multiplying the complex spectral samples with the corresponding gain. The first L samples of the K-point IFFT of the complex spectrum are multiplied by the modified Hamming window to get time domain signal. The output signal is synthesized using overlap-add operation. It may be noted that the processing has to get completed in S sampling intervals for real-time operation.

3.4.2 Test results for real-time processing

The real-time implementation was evaluated using a visual examination of the processed waveforms, informal listening, and objective evaluation using 'perceptual evaluation of speech quality (PESQ)' measure [94]. The technique was tested for different speech materials, music, and environmental sounds with large variation in the sound level. An example of the processing, with L as 256, K as 512, CR as 2, s_a as 1, and s_r as 30, is shown in Figure 3.12. The processed outputs from offline and real-time processing have the same amplitude variation, with a high gain at low input levels and decreased gain for increased input level.

Informal listening 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 and no perceptible distortions were noticed in the processed outputs. The processed output signal from the DSP board was acquired through a PC sound card. PESQ score for speech outputs from the real-time processing with reference to the output from offline processing was 3.50, indicating that the processing artifacts due to fixed-point processing were not significant.

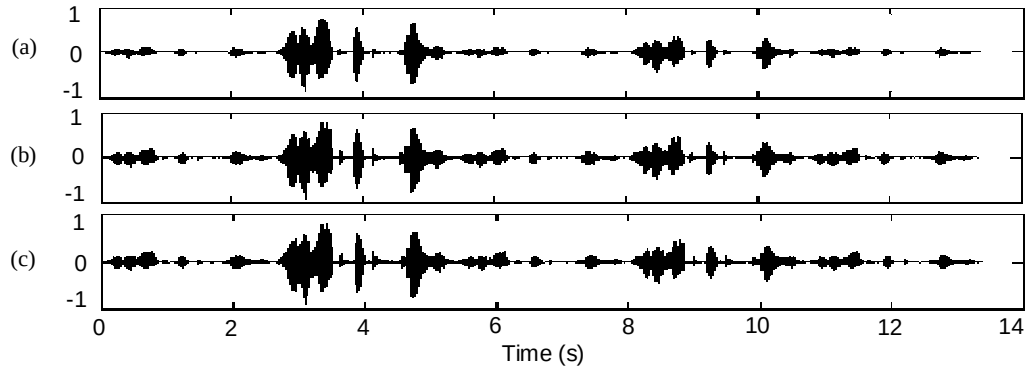


Figure 3.12 Example of real-time processing for the sliding band compression: (a) input signal (as in Figure 3.9(c)), (b) offline processed waveform, (c) real-time processed waveform.

To estimate the computational load on the processor, the system clock was progressively decreased from 120 MHz. It was found that the processing required a minimum clock frequency of 50 MHz indicating that the technique needed approximately 41% of the processing capacity. The audio latency (total signal delay) is the sum of the algorithmic delay and the input-output delay (delay in the input and output buffering operations and filters). It was measured by applying a 1 kHz tone burst of 200 ms from a function generator as the input and observing the delay from onset of the input tone burst to the corresponding onset in the output, using a digital storage oscilloscope. The total signal delay was found to be approximately 36 ms.

3.5 Discussion

A dynamic range compression technique, named as sliding-band compression, has been presented to compensate for frequency-dependent loudness recruitment associated with sensorineural hearing loss without introducing the distortions generally associated with the single-band and multiband compressions. In this technique, the gain for each spectral sample is calculated based on the power in a band sliding with the frequency. It results in time-varying frequency response with the magnitude response being smooth along time and frequency axes. As the gain for a spectral component is determined by the spectral components located within a band centered at its frequency, the technique avoids attenuation of high-frequency components due to the presence of strong low-frequency components, which may occur in single-band compression. The proposed technique avoids distortions in the shape of spectral resonances and discontinuities during the resonance transitions, which may occur in multiband compression. The bandwidth is selected to approximate the frequency resolution of the auditory system. The proposed technique permits use of settable attack and

release times and a frequency-dependent compression function selected in accordance with the desired hearing aid fitting procedure to compensate for the abnormal loudness growth.

Several studies evaluating the performance of single-band and multiband compression [9], [10], [47]−[54], using listening tests conducted on listeners with sensorineural hearing loss reported that multiband compression with smaller compression ratios and a smaller compression segment were preferred over those with larger compression ratios and with larger compression segment. It has also been reported that compression schemes with several narrow bands produce more spectral distortion at the band boundaries and spectral flattening than the schemes with a small number of bands [56]−[58]. Therefore, the proposed technique was compared with single-band compression and multiband compression techniques by visual examination of the spectrograms for undesirable level changes in the outputs and by quantifying the deviations from the expected output levels.

The three compression techniques were tested using single-tone, two-tone, and speech signals as the inputs to examine the difference in their processed outputs. The single-tone input comprised a sine wave with constant amplitude and time-varying frequency. It was used to examine the distortions at the band boundaries. The two-tone input comprised two sine waves, a wave of low frequency with a time-varying amplitude and a wave of high frequency with a constant low amplitude. It was used to examine the effect of the level of the low-frequency component on the level of the high-frequency component.

The visual examination of the spectrograms indicated that sliding-band compression significantly reduced the temporal and spectral distortions associated with the single and multiband compression techniques. The deviations from the expected output levels were quantified as the level error in dB. For single-tone inputs, the multiband compression showed frequency-dependent error, with the maxima at the frequencies close to the inter-band boundaries. The maximum errors increased with CR, from 1.5 dB at CR of 2 to 2.5 dB at CR of 10. The single-band and sliding-band compressions showed no errors for the single-tone inputs. For the two-tone inputs, the output of the single-band compression showed attenuation of the high-frequency tone with an increase in the level of the low-frequency tone. This attenuation was not observed in the outputs of the multiband and sliding-band compressions. Visual examination of the waveforms and spectrograms of the processed output of the sliding-band compression for the input speech signal with different modulation envelopes showed desired amplification and compression, without noticeable deviations.

The proposed technique avoids distortions associated with commonly employed compression techniques, but it has a high computational requirement because of the level estimation and gain calculation for each spectral sample. For reducing the computation, the

technique has been implemented using FFT-based analysis-synthesis. The algorithmic delay of this implementation is equal to the sum of the window length and window shift used for the analysis-synthesis. For real-time processing, the computation for each frame should be completed within a window shift. The audio latency (total signal delay) is the sum of the algorithmic delay and the input-output delay (delay in the input and output buffering operations and filters). It should be significantly lower than 120 ms to be acceptable for face-to-face conversation [180]. To examine the feasibility of the proposed technique for use in hearing aids with limited computational resources and power constraints, the technique was implemented on the 16-bit fixed-point processor TI/TMS320C5515, with a 120 MHz clock and 320 KB on-chip RAM. The processing used approximately 41% of the processor capacity and the audio latency was found to be approximately 36 ms. It may be noted that this DSP chip was introduced in 2010 and several chips with higher processing capacity and lower power consumption have become available since then. These advancements, in terms of increasing processing capacity and decreasing power consumption, may be expected to continue. Therefore, it should be feasible to use the proposed sliding-band compression technique for real-time processing in hearing aids without sacrificing their other features.

The proposed compression technique has also been implemented as a smartphone app [101] for use as a hearing aid. Another smartphone app [102] has been developed that combines the proposed compression technique with the noise suppression technique as presented in Chapter 4. This app, described in Appendix C, has been developed to make the proposed processing techniques conveniently available to the hearing-impaired user, with a graphical touch interface for setting the processing parameters in an interactive and real-time mode. The audio latency of the app tested using Nexus 5X handset was 45 ms. The app permits use and evaluation of the proposed processing techniques by a large number of users without incurring the expenses involved in the ASIC-based hearing aid development.

Chapter 4

SPEECH ENHANCEMENT USING NOISE ESTIMATION WITH DYNAMIC QUANTILE TRACKING

4.1 Introduction

Single-input speech enhancement techniques can be used for background noise suppression and improving speech perception of hearing-impaired listeners. They involve estimation of the noise spectrum from the noisy speech spectrum and using the estimated noise spectrum along with a noise suppression function for speech enhancement. Underestimation of the noise results in residual noise and overestimation results in distortion leading to degraded quality and reduced intelligibility. A review of noise estimation techniques has been presented in Section 2.7 of the second chapter. The quantile-based noise estimation, as reviewed in the second chapter, is based on the observation that the total short-time energy in each sub-band is close to the noise short-time energy for most of the time. Although not explicitly stated in the literature, it is based on the assumption that a real-world noise is more stationary than the speech signal in terms of the sub-band levels. Several quantile-based techniques for noise estimation [13]−[16] have been reported. These perform well in presence of stationary and nonstationary noises. Estimation of the quantiles by storing and sorting of the spectral samples requires a large memory and has high computational complexity. Therefore, these techniques are currently not suitable for speech enhancement in hearing aids. Thus, there is a need for developing a noise estimation technique with low memory and computational requirements for use in hearing aids.

Here we present a noise estimation technique based on tracking of quantiles in real-time without sorting of past spectral samples in order to reduce the memory and computational requirements. Towards this objective, a technique for dynamic tracking of quantiles for use in applications involving real-time estimation of quantiles of a data stream is developed. This technique approximately tracks a quantile without prior knowledge of the distribution of the data stream and without storage and sorting of the past samples. The proposed dynamic quantile tracking is subsequently applied for the tracking of the quantiles of the noisy speech spectrum for noise spectrum estimation without voice activity detection. It permits the use of a different quantile for each sub-band without processing overheads. An improved noise estimation technique that selects the quantiles adaptively is also presented. It involves estimating a quantile function (inverse of cumulative distribution function) for each sub-band by dynamically tracking multiple quantiles. The two proposed noise estimation techniques are

evaluated and compared with some of the existing techniques in terms of computational complexity and noise estimation accuracy. The proposed techniques in combination with spectral subtraction based on the geometric approach [90] are used for suppression of background noise.

The proposed dynamic quantile tracking technique for data streams is presented in Section 4.2. The noise estimation techniques based on dynamic quantile tracking are presented in Section 4.3. Evaluation of the noise estimation techniques in terms of computational requirement and noise tracking, along with a comparison with some of the existing techniques, is presented in Section 4.4. Evaluation of the noise estimation techniques in a speech enhancement framework is presented in Section 4.5. Implementation for real-time processing and test results are presented in Section 4.6, followed by the discussion in the last section.

4.2 Dynamic Quantile Tracking for Data Streams

The p-quantile (or 100p-percentile) q of a random variable X satisfies the condition $\text{Prob}(X \le q) = p$. The most common method for obtaining quantile estimate involves storing previous N samples and sorting them in ascending order as $\{x((1)), x((2)), ..., x((N))\}$ to obtain a point estimate of the p-quantile as $\hat{q} = x((\lceil pN \rceil))$. This estimate obtained from the order statistics of the samples is known as the sample quantile. A computationally efficient dynamic quantile tracking technique [103], based on stochastic approximation [104], is proposed for estimating the quantile. It recursively updates the quantile estimate eliminating the need for storing and sorting operations. It has lower memory and computational requirements than the earlier reported quantile estimation techniques based on stochastic approximation [105], [106].

The quantile is dynamically estimated as the input sample of the data stream arrives by applying an increment Δ_+ or a decrement Δ_- on the previous estimate. The values of Δ_+ and Δ_- are calculated as appropriate fractions of the range of the input samples such that the estimate after a sufficiently large number of input samples converges to the sample quantile. As the underlying distribution of the data is unknown, the range also needs to be dynamically estimated. The technique uses a step-size control factor for a trade-off between variance and adaptivity of the estimation. For input sample $x(n)$, the estimate of p-quantile is calculated recursively as

$$\hat{q}(n) = \begin{cases} \hat{q}(n-1) + \Delta_+, & x(n) \ge \hat{q}(n-1) \\ \hat{q}(n-1) - \Delta_-, & \text{otherwise} \end{cases} \tag{4.1}$$

The values of Δ_+ and Δ_- should be such that the quantile estimate converges to the sample quantile and sum of the increments approaches zero. For stationary data and a sufficiently large number of input samples N, the change is expected to be $-\Delta_-$ for pN samples and Δ_+ for $(1-p)N$ samples. Therefore, we should have

$$(1-p)N\Delta_+ - pN\Delta_- = 0 \tag{4.2}$$

which results in $\Delta_+ / \Delta_- = p/(1-p)$. Therefore, Δ_+ and Δ_- may be selected as

$$\Delta_+ = \lambda pR \tag{4.3}$$

$$\Delta_- = \lambda(1-p)R \tag{4.4}$$

where R is the range (difference between the maximum and minimum values) and λ is a convergence factor that controls the step size λR and determines the convergence and ripples of the estimate. It can be shown, as in [107], that $\lim_{n\to\infty} E(\hat{q}(n)) = q$, if $0 < \lambda < 1/(p_{\max}R)$, where $p_{\max}$ is the peak of the probability density function. Thus, the upper bound on λ is given as

$$\lambda_{\max} = 1/(p_{\max}R) \tag{4.5}$$

For tracking $p_{\max}$, the stochastic approximation based quantile estimation techniques [106], [108] estimate the probability density recursively and use a lower bound on it to prevent the estimation from becoming unstable. However, tracking $p_{\max}$ requires additional calculations for recursively estimating the probability density function. To avoid these calculations, the value of λ may be selected empirically such that convergence is ensured for a given application, as described in Section B.2.1 of Appendix B.

Near convergence, the peak-to-peak ripple δ in the estimated values is $\Delta_+ + \Delta_-$ and therefore it is given as

$$\delta = \lambda R \tag{4. 6}$$

During tracking, the maximum number of steps needed for the estimated value to change from its initial value q_{init} to its final value q_{fin} is given as

$$s = \max\{(q_{\text{fin}} - q_{\text{init}})/\Delta_+,\ (q_{\text{init}} - q_{\text{fin}})/\Delta_-\} \tag{4.7}$$

Since $(|q_{\text{fin}} - q_{\text{init}}|)_{\max} = R$, the maximum number of steps using (4.3) and (4.4) is given as

$$s = \max\left\{\frac{1}{\lambda p},\ \frac{1}{\lambda(1-p)}\right\} \tag{4.8}$$

It may be noted that s becomes very large for very low or high values of p. The value of λ is selected to ensure convergence and for an appropriate trade-off between δ and s which are related to variance and adaptivity, respectively.

The range is estimated using dynamic peak and valley detectors. The peak estimate $P(n)$ and the valley estimate $V(n)$ are updated without any restriction on their polarity, using the following first-order recursive relations:

$$P(n) = \begin{cases} \tau_p P(n-1) + (1-\tau_p)x(n), & x(n) \geq P(n-1) \\ \sigma_p P(n-1) + (1-\sigma_p)V(n-1), & \text{otherwise} \end{cases} \tag{4.9}$$

$$V(n) = \begin{cases} \tau_v V(n-1) + (1-\tau_v)x(n), & x(n) \leq V(n-1) \\ \sigma_v V(n-1) + (1-\sigma_v)P(n-1), & \text{otherwise} \end{cases} \tag{4.10}$$

and the range is tracked as

$$R(n) = P(n) - V(n) \tag{4.11}$$

The coefficients τ_p, τ_v, σ_p, and σ_v are selected in the range [0, 1] to control the rise and fall rates of the range estimation. An overestimation of the range results in increased variance in the quantile estimation and an underestimation of the range results in lower adaptivity. Considering adaptivity to be more important for applications involving nonstationary data, we select small τ_p and τ_v values to provide fast response to an increase in the range and large σ_p and σ_v for a slow response to a decrease in the range.

With the range $R(n)$ tracked as in (4.11), the dynamic quantile tracking as given by (4.1), (4.3), and (4.4) can be rewritten as the following:

$$\hat{q}(n) = \begin{cases} \hat{q}(n-1) + \lambda p R(n), & x(n) \geq \hat{q}(n-1) \\ \hat{q}(n-1) - \lambda(1-p)R(n), & \text{otherwise} \end{cases} \tag{4.12}$$

The technique, comprising the computation steps as given by (4.9), (4.10), (4.11), and (4.12) and using sample delay operations, is shown as a block diagram in Figure 4.1. The details of the proposed dynamic quantile tracking using range estimation (DQTRE) including the trade-off between convergence and ripple, evaluation of the technique using synthetic and real data with different distributions, and comparison with some of the techniques having similar features are presented in Appendix B. The quantile values estimated using DQTRE were compared with the sample quantiles as the reference values and with those obtained using the piecewise-parabolic formula (P2) by Jain and Chlamtac [109], the smooth stochastic approximation (SSA) technique by Amiri and Thiam [110], and the exponentially weighted stochastic approximation (EWSA) technique by Chen et al. [108]. The techniques were tested using synthetic stationary data with several symmetric and asymmetric density functions, synthetic nonstationary data with time-varying mean and standard deviation, and real data streams with different distributions. The comparisons and test results presented in Appendix B show that DQTRE has low memory and computational requirements as compared with low-variance techniques such as P2 and SSA. As it does not keep a track of the sample

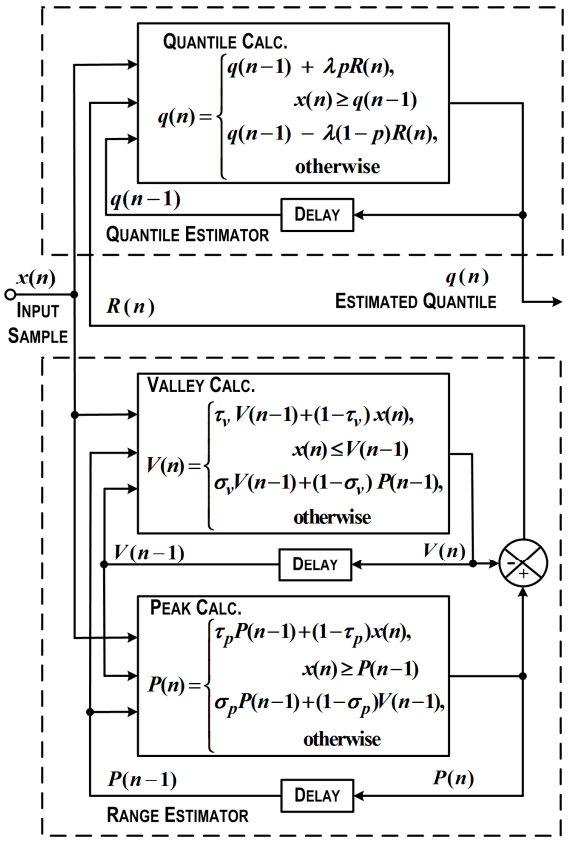


Figure 4.1 Dynamic quantile tracking using range estimation.

number, it is suitable for sample-by-sample or window-based tracking of quantiles of nonstationary data and can be used without any restriction on the sequence length. As compared to technique with fast adaptivity such as EWSA, it gives much lower variance during stationary segments and an acceptable adaptivity during transitions.

Due to its low memory and computational requirements, DQTRE can be used for real-time quantile tracking of multiple variables using a single processor. Noise spectrum estimation using DQTRE is described in the next section.

4.3 Noise Spectrum Estimation Using Dynamic Quantile Tracking

The DQTRE technique presented in the previous section for tracking quantile of a stream of data is applied for noise spectrum estimation by estimating a fixed quantile for each spectral sample of the short-time spectrum of the noisy input signal. This technique is referred to as dynamic quantile tracking based noise estimation (DQTNE). An improved technique using an adaptive quantile for each spectral sample is also presented. This technique is referred to as

adaptive dynamic quantile tracking based noise estimation (ADQTNE). The two techniques are presented in the following subsections.

4.3.1 Noise estimation using dynamic quantile tracking with fixed quantile (DQTNE)

The processing for single-channel speech enhancement comprises steps of windowing the noisy input signal, short-time spectrum calculation, noise spectrum estimation, enhanced magnitude spectrum calculation, estimation of enhanced short-time complex spectrum, and resynthesis using overlap-add. The noisy signal $x(m)$ is divided into overlapping frames by the application of a window function and the short-time spectrum $X(n, k)$ is calculated at nth frame and kth spectral sample. Assuming additive noise, the noisy spectrum $X(n, k)$ is sum of clean spectrum $Y(n, k)$ and noise spectrum $D(n, k)$. In the proposed DQTNE technique, the noise spectrum is estimated by tracking the p-quantile estimate $\hat{q}_{|X|}(n,k)$ of $|X(n,k)|$ at kth spectral sample of the nth frame by applying an increment on its previous estimate $\hat{q}_{|X|}(n-1,k)$ as

$$\hat{q}_{|X|}(n,k) = \begin{cases} \hat{q}_{|X|}(n-1,k) + \Delta_+(k), & |X(n,k)| \geq \hat{q}_{|X|}(n-1,k) \\ \hat{q}_{|X|}(n-1,k) - \Delta_-(k), & \text{otherwise} \end{cases} \tag{4.13}$$

The values of the increment $\Delta_+(k)$ and the decrement $\Delta_-(k)$ should be such that the quantile estimate converges to the sample quantile and sum of the changes in the estimate approaches zero. Therefore $\Delta_+(k)$ and $\Delta_-(k)$ may be selected as

$$\Delta_+(k) = \lambda R_{|X|}(n,k)\, p \tag{4.14}$$

$$\Delta_-(k) = \lambda R_{|X|}(n,k)(1-p) \tag{4.15}$$

where $R_{|X|}$ is the range (difference between maximum and minimum of the spectral values at a particular frequency) and λ is a convergence factor that controls the step size $\lambda R_{|X|}(n, k)$ and determines the convergence and ripples of the estimate. Estimation of the quantile $\hat{q}_{|X|}(n,k)$ as given by (4.13), (4.14), and (4.15) can be written as

$$\hat{q}_{|X|}(n,k) = \begin{cases} \hat{q}_{|X|}(n-1,k) + \lambda p R_{|X|}(n,k), & |X(n,k)| \geq \hat{q}_{|X|}(n-1,k) \\ \hat{q}_{|X|}(n-1,k) - \lambda(1-p) R_{|X|}(n,k), & \text{otherwise} \end{cases} \tag{4.16}$$

The range is estimated using dynamic peak and valley detectors. The peak $P_{|X|}(n, k)$ and the valley $V_{|X|}(n, k)$ are updated, using the following first-order recursive relations:

$$P_{|X|}(n,k) = \begin{cases} \tau_p P_{|X|}(n-1,k) + (1-\tau_p)|X(n,k)|, & |X(n,k)| \geq P_{|X|}(n-1,k) \\ \sigma_p P_{|X|}(n-1,k) + (1-\sigma_p) V_{|X|}(n-1,k), & \text{otherwise} \end{cases} \tag{4.17}$$

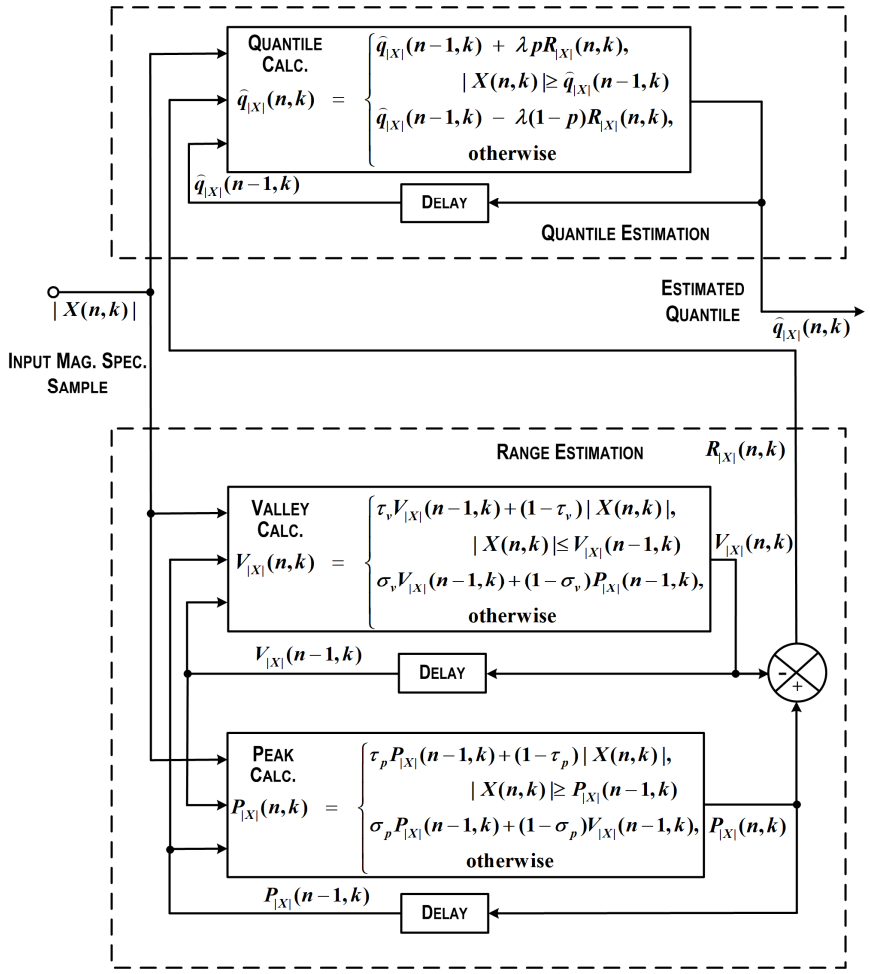


Figure 4.2 Estimation of the noise spectral samples using dynamic quantile tracking technique based on range estimation.

$$V_{|X|}(n,k) = \begin{cases} \tau_v V_{|X|}(n-1,k) + (1-\tau_v)\,|X(n,k)|, & |X(n,k)| \leq V_{|X|}(n-1,k) \\ \sigma_v V_{|X|}(n-1,k) + (1-\sigma_v)P_{|X|}(n-1,k), & \text{otherwise} \end{cases} \quad (4.18)$$

The constants τ_p, τ_v, σ_p, and σ_v are selected in the range [0, 1] to control the rise and fall rates. As the peak and valley samples may occur after long intervals, τ_p and τ_v should be small to provide fast detector responses to an increase in the range and σ_p and σ_v should be relatively large to avoid ripples. The range is dynamically tracked as

$$R_{|X|}(n, k) = P_{|X|}(n, k) - V_{|X|}(n, k) \quad (4.19)$$

A block diagram of the technique with its computation steps is shown in Figure 4.2. The noise estimate at *n*th frame and frequency index *k* is updated as

$$\hat{D}_{\mathrm{DQTNE}}(n,k) = \hat{q}_{|X|}(n,k) \quad (4.20)$$

To find the most suitable quantile for the noise estimation, the offline processing was carried out using sample quantile (SQ). The noisy signal was obtained using white noise and

babble from the NOISEX database [74] and street, station, restaurant, subway, exhibition, lobby, and airport noises from the AURORA database [93] added at −6, −3, 0, 3, … 15 dB SNR to the speech sentences from the GRID database [111] on the active speech level (ASL) [112] basis. The FFT-based analysis-synthesis was carried out using 25.6-ms frames with 75% frame overlap using sampling frequency of 10 kHz, window length of 256, shift length of 64, and FFT size of 512. For each frame, the sample quantile values were estimated for p as 0.1, 0.25, 0.50, and 0.75, and using previous 256 frames (corresponding to 1.6 s at the sampling frequency of 10 kHz). Figure 4.3 shows the spectrograms for SQ-estimated noise using different p values for speech with white noise at 3 dB SNR. The spectrograms of clean signal, noisy signal, and actual noise are also shown for reference. SQ-estimate with p as 0.1 exhibits underestimation whereas SQ-estimate with p as 0.75 exhibits overestimation in presence of strong low-frequency components of speech. SQ-estimate with p as 0.50 matches the noise at high frequencies but exhibits overestimation in presence of strong low-frequency components of speech. SQ-estimate with p as 0.25 exhibits a satisfactory tradeoff between underestimation at high frequencies and overestimation at low frequencies. Similar results were observed for other noises. An example for babble at 3 dB SNR is shown in Figure 4.4. These results indicate that 0.25-quantile is appropriate for noise estimation with low computational complexity. A frequency-dependent quantile may be used for improved noise estimation.

An experiment involving estimation of histogram using storing and sorting of the spectral samples was carried out for an empirical estimation of $\lambda_{\max}$ (upper bound on λ). At kth spectral sample and nth frame, a 50-bin histogram was updated using previous 256 frames (corresponding to 1.6 s at sampling frequency of 10 kHz). The bins of the histogram were evenly distributed between maximum and minimum of the spectral values obtained from the previous 256 frames. The range $R_{|X|}$ was calculated as the difference between the maximum and minimum values. For each n and k, the peak of the probability density function $p_{\max}$ was calculated as the maximum bin count divided by the number of frames and the bin width used for the histogram estimation. The empirical value of $\lambda_{\max}$ was calculated, in accordance with (4.5), as

$$\lambda_{\max} = \min\{(p_{\max}(n,k)R_{|X|}(n,k))^{-1} \quad \forall n, \forall k\} \tag{4.21}$$

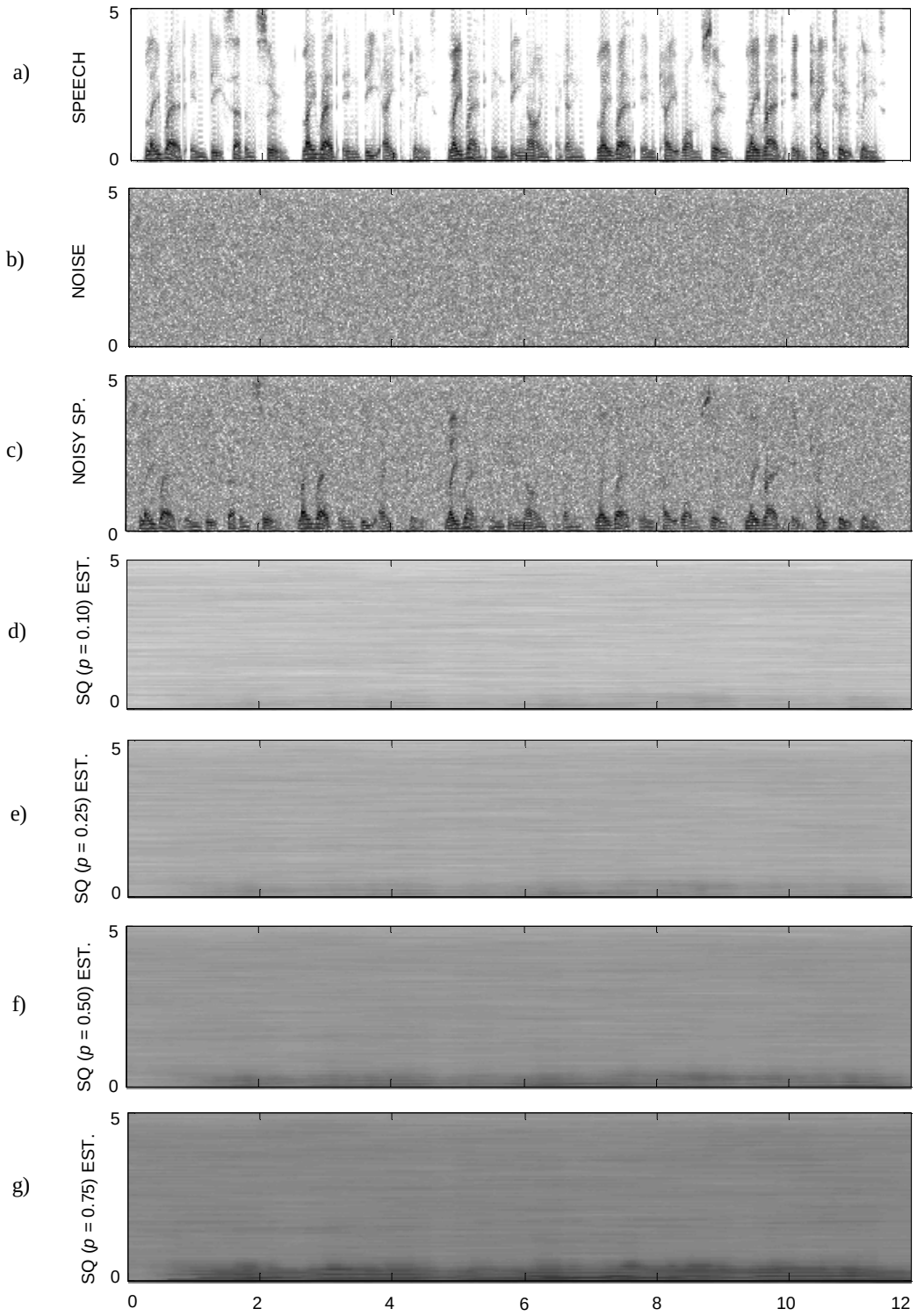


Figure 4.3 Noise tracking for speech signal degraded by white noise at 3 dB SNR: Spectrograms of (a) speech, (b) noise, (c) noisy speech, (d) SQ-estimated noise with p as 0.10, (e) SQ-estimated noise with p as 0.25, (f) SQ-estimated noise with p as 0.50, and (g) SQ-estimated noise with p as 0.75, with time (s) on the x-axis and frequency (kHz) on the y-axis.

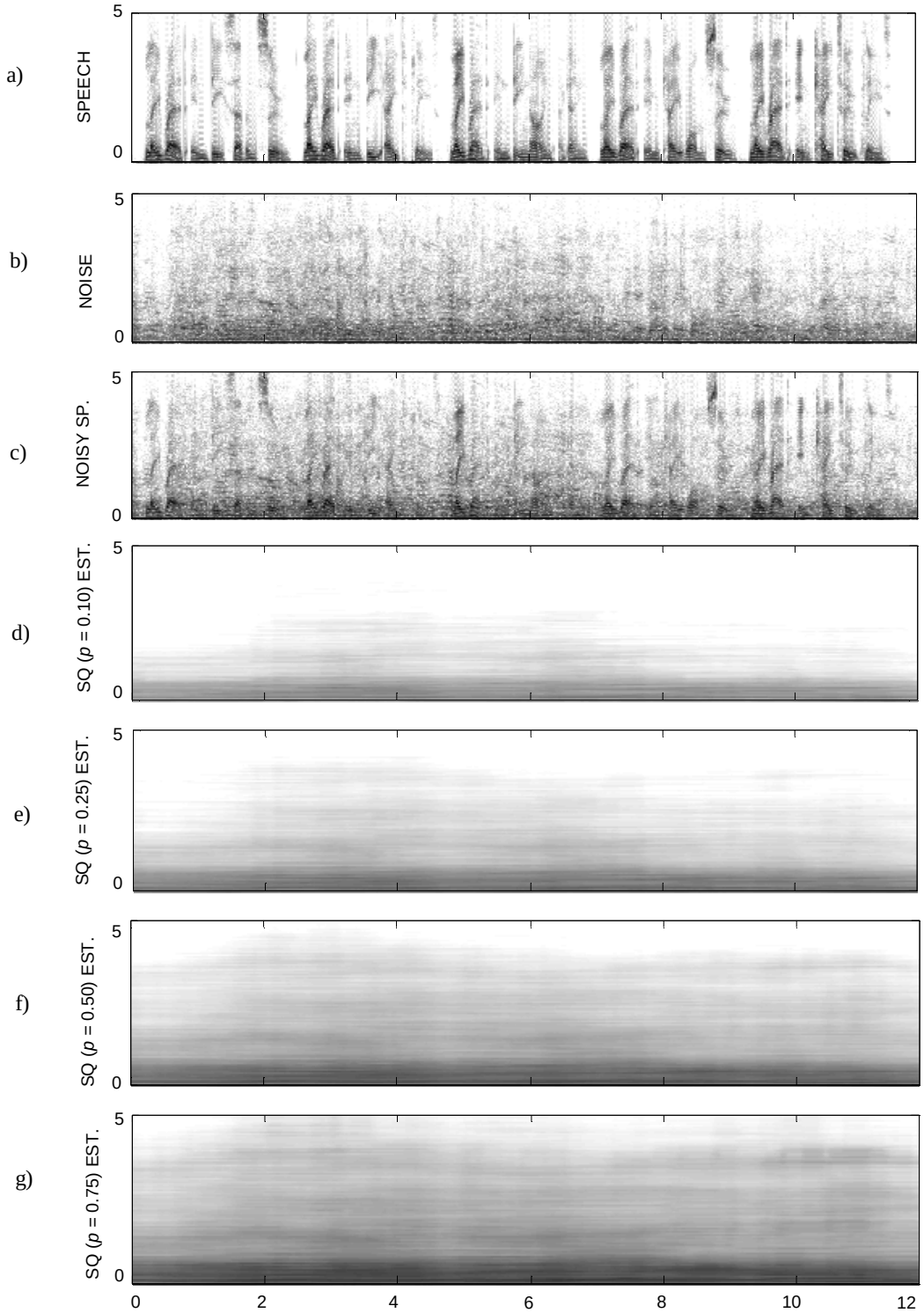


Figure 4.4 Noise tracking for speech signal degraded by babble at 3 dB SNR: Spectrograms of (a) speech, (b) noise, (c) noisy speech, (d) SQ-estimated noise with p as 0.10, (e) SQ-estimated noise with p as 0.25, (f) SQ-estimated noise with p as 0.50, and (g) SQ-estimated noise with p as 0.75, with time (s) on the x-axis and frequency (kHz) on the y-axis.

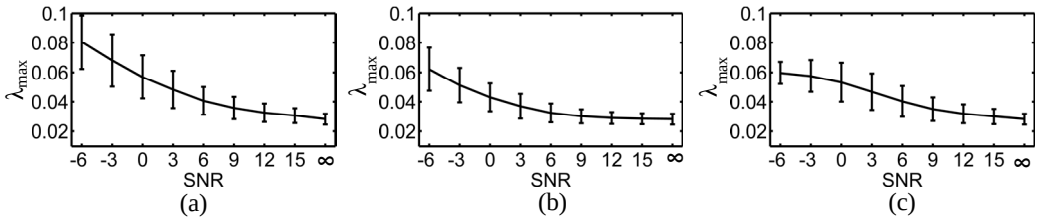


Figure 4.5 λ_{max} (mean and deviation) as a function of SNR for sentences from GRID database and three noises: (a) babble, (b) street noise, and (c) exhibition noise.

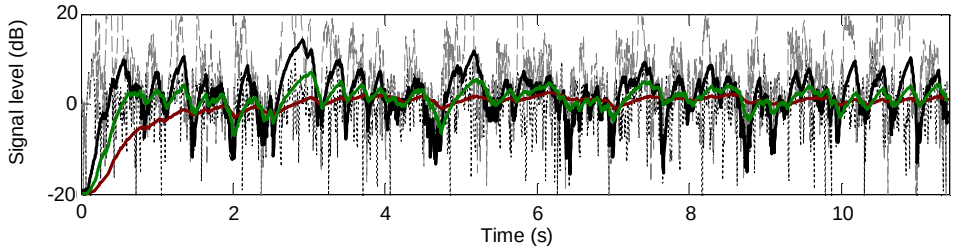


Figure 4.6 Noise estimation example for speech degraded by white noise at 3 dB SNR using 0.25-quantile for a frequency sample *k* as 15 (293 Hz); with noisy speech as thin dashed gray trace, noise as thin dotted black trace, estimated noise using λ as 1/64 as thick black trace, estimated noise using λ as 1/256 as thick green trace, and estimated noise using λ as 1/1024 as thick brown trace.

Figure 4.5 shows plots of λ_{max} as a function of SNR for babble, street noise, and exhibition noise. It can be seen that λ_{max} decreases with an increase in SNR and the lowest value is approximately 0.02. Similar λ_{max} versus SNR plots were observed for other noises. To ensure that $\widehat{q}_{|X|}$ converges in L_1 to true quantile $q_{|X|}$ for various types of noises and SNRs, the upper bound on λ is selected as 0.02.

The value of λ should be selected for an appropriate trade-off between the peak-to-peak ripple δ and the maximum number of steps needed for convergence *s*, which are related to variance and adaptivity, respectively. An experiment involving the estimation of noise spectrum using dynamic quantile tracking was carried out for empirical estimation of λ. The speech and noise materials used were the same as that used in the previous two investigations. Several values of λ (1/64, 1/128, 1/256, 1/512, 1/1024), with $\lambda < \lambda_{max}$, were used to estimate noise as 0.25-quantile. It was observed that using λ as 1/64 and 1/128 resulted in fast convergence but large ripples, whereas using λ as 1/512 and 1/1024 resulted in slow convergence and smaller ripples. It was found that the estimation using λ as 1/256 resulted in an appropriate tradeoff between ripple and convergence and was suitable for noise estimation. Figure 4.6 shows an example of tracking of 0.25-quantile using dynamic quantile tracking of noisy speech input in a speech-dominated frequency with three values of the convergence factor λ (1/64, 1/256, 1/1024) for input speech degraded by white noise at 3 dB SNR. Noisy speech and noise are also shown for comparison. The estimation using λ as 1/64 shows fast

tracking (as seen in the initial 1 s) but higher ripples in the presence of speech, whereas the estimation using λ as 1/1024 shows smaller ripples but slow tracking. The estimation using λ as 1/256 shows a tradeoff between ripple and tracking and thus a fixed λ of 1/256 may be used for reduced computational complexity. These results indicate a need for adaptive λ for a controlled tradeoff between ripple and tracking.

4.3.2 Noise estimation using dynamic quantile tracking with adaptive quantile (ADQTNE)

The DQTNE technique presented in the previous subsection uses empirically selected values of p as 0.25 and λ as 1/256. An improved noise estimation technique that uses adaptive p and λ at each frame and spectral sample is presented. The technique, referred to as adaptive dynamic quantile tracking technique for noise estimation (ADQTNE), involves estimation of a quantile function (inverse of cumulative distribution function). For each spectral sample, the quantile function is coarsely estimated by dynamically tracking multiple quantiles for a set of probabilities. Each quantile is updated recursively, without storage and sorting of past spectral samples, using the DQTRE technique with an increment determined by the dynamically estimated range. The adaptive quantile representing the noise is obtained by finding the quantile where the quantile function has the lowest slope, which approximately corresponds to the peak of the probability density function of the noisy signal. We use a set of evenly spaced probabilities for calculating the quantile function. The quantile for lowest slope is located as the quantile at which the difference between adjacent quantiles is minimum.

In the context of DQTNE, the p-quantile estimate of the magnitude spectrum at kth spectral sample of the nth frame is represented as $\hat{q}_{|X|}(n,k)$. For ADQTNE, we estimate quantiles for multiple probabilities and the p_i-quantile estimate of the magnitude spectrum at kth spectral sample of the nth frame is represented as $\hat{q}_{|X|,i}(n,k)$. The set of quantiles $\{\hat{q}_{|X|,1}(n,k), \hat{q}_{|X|,2}(n,k), \ldots \hat{q}_{|X|,M}(n,k)\}$ corresponding to the set of evenly spaced probabilities $\{p_1, p_2, \ldots p_M\}$ are obtained using (4.16) with a common range $R_{|X|}$ tracked using (4.19) as shown in Figure 4.7. The lowest quantile for which the difference between the adjacent quantiles is minimum is used as the adaptive quantile $\hat{q}_{|X|,\mathrm{ad}}(n,k)$. It is expressed as

$$\hat{q}_{|X|,\mathrm{ad}}(n,k) = \min\{\hat{q}_{|X|,i}(n,k) \mid (\hat{q}_{|X|,i}(n,k) - \hat{q}_{|X|,i-1}(n,k)) \le (\hat{q}_{|X|,j}(n,k) - \hat{q}_{|X|,j-1}(n,k)) \quad \forall \ 2 \le j \le M\} \tag{4.22}$$

The adaptive quantile at the nth frame and kth spectral sample is used as the noise estimate

$$\hat{D}_{\mathrm{ADQTNE}}(n,k) = \hat{q}_{|X|,\mathrm{ad}}(n,k) \tag{4.23}$$

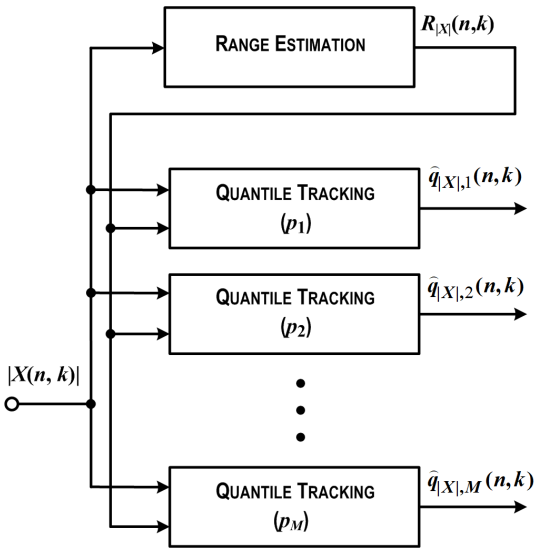


Figure 4.7 Dynamic tracking of multiple quantiles with a common range estimator.

The quantile function is estimated by tracking eight quantiles corresponding to p as 0.25, 0.30, 0.35,...., and 0.60. These p values are used for locating the adaptive quantile, because of the observation that a quantile corresponding to a lower p resulted in significant underestimation and that to a higher p resulted in significant overestimation. An example of noise estimation using adaptive quantile and that using 0.25-quantile for two frequencies is shown in Figure 4.8. The noisy speech and noise at two frequencies are shown as thin dashed gray trace and thin dotted black trace, respectively. For frequency corresponding to k as 15 (293 Hz) where speech is present, the noisy speech trace shows deviations from the noise trace. For frequency corresponding to k as 90 (1.75 kHz) where speech is absent, the noisy speech and noise traces are similar. For k as 90, the noise estimated using 0.25-quantile shows underestimation whereas noise estimated using adaptive quantile is closer to the noise trace. For k as 15, noise estimation using 0.25-quantile and adaptive quantile are similar except that the adaptive quantile based estimate deviates from the noise trace when peaks occur in the noisy speech trace, indicating the need for selecting λ dependent on speech presence probability. A larger λ should be used for faster tracking of noise in absence of speech and a smaller λ should be used in presence of speech to avoid overestimation of noise. Therefore, an adaptive λ dependent on the speech-presence probability is used for improved noise estimation.

The speech presence probability, for estimation of the noise spectrum from the noisy speech spectrum, has been conventionally calculated using the ratio of the spectral sample to the local minimum [78], [80], [81]. This calculation results in an undesirable increase in the

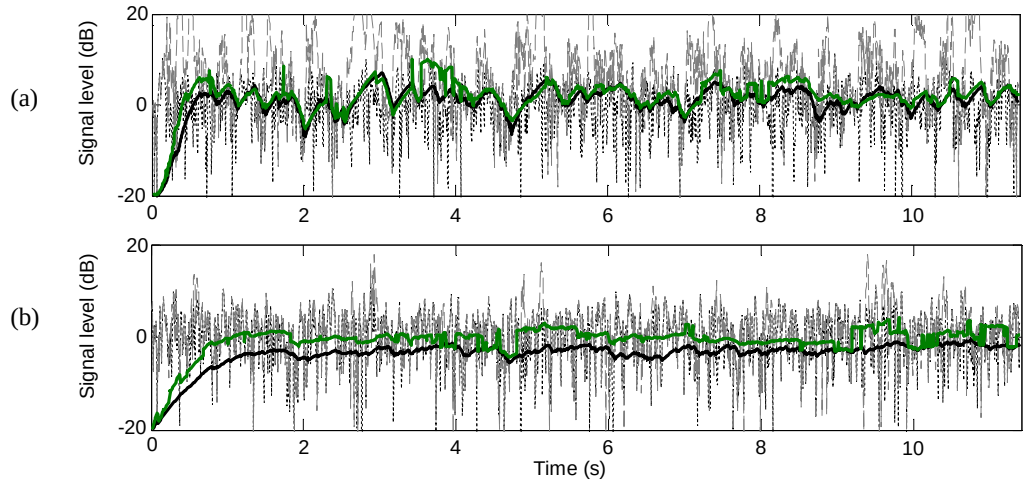


Figure 4.8 Noise estimation example for speech degraded by white noise at 3 dB SNR for two frequency samples: (a) *k* as 15 (293 Hz) and (b) *k* as 90 (1.75 kHz); with noisy speech as thin dashed gray trace, noise as thin dotted black trace, estimated noise using 0.25-quantile (λ = 1/256) as thick solid black trace; estimated noise using adaptive quantile (λ = 1/256) as thick solid green trace.

value of speech presence probability in case of a sudden increase in the noise floor that results in an increased spectral sample without a corresponding increase in the minimum. To avoid this problem, we calculate an instantaneous speech presence probability $p_{si}(n, k)$ based on the ratio of the spectral sample and a robust estimate of the noise floor, followed by a recursive averaging of $p_{si}(n, k)$ to obtain the smoothed speech presence probability $p_{ss}(n, k)$. The across-frequency mean of the spectral-sample range $R_{|X|}(n,k)$ is recursively averaged to calculate the time-varying average range $\bar{R}(n)$ as

$$\bar{R}(n) = \beta \bar{R}(n-1) + (1-\beta) \sum_{k=0}^{(K/2)-1} \frac{R_{|X|}(n,k)}{(K/2)-1} \tag{4.24}$$

where β is a smoothing constant selected as 0.95 and K is the FFT size. In the absence of speech, $R_{|X|}(n,k)$ is small for most of the frequency samples resulting in a small mean range. In the presence of speech, $R_{|X|}(n,k)$ is large for frequency samples corresponding to speech, but the mean range is close to that during the absence of speech. The instantaneous speech presence probability $p_{si}(n, k)$ is calculated using $\bar{R}(n)$, as a robust indicator of the noise floor, and the spectral sample $|X(n, k)|$ as

$$p_{si}(n,k) = 1 - \min\left\{ \frac{\bar{R}(n)}{|X(n,k)|}, 1 \right\} \tag{4.25}$$

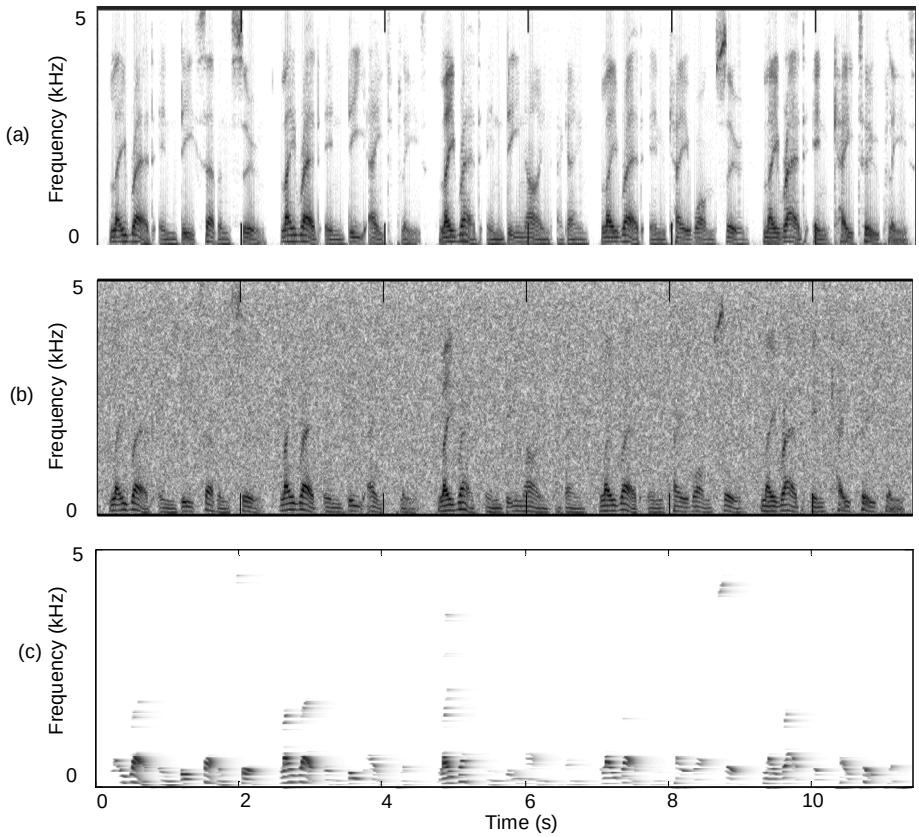


Figure 4.9 Speech presence probability calculation using (4.26), for input speech degraded by white noise at 3 dB SNR: (a) Spectrogram of speech; (b) Spectrogram of noisy speech; and (c) Plot of speech presence probability (gray scale: 0 indicated by white, 1 indicated by black) as a function of time and frequency.

In absence of speech, the ratio $\bar{R}(n)/|X(n,k)|$ is close to one and $p_{si}(n, k)$ is close to zero. In presence of speech, the ratio is close to zero and $p_{si}(n, k)$ is close to one. The instantaneous speech presence probability $p_{si}(n, k)$ is recursively averaged to calculate the smoothed speech presence probability $p_{ss}(n, k)$ as

$$p_{ss}(n,k) = \alpha\, p_{ss}(n-1,k) + (1-\alpha)\, p_{si}(n,k) \tag{4.26}$$

where α is an updating constant selected as 0.2. An example of the speech presence probability calculated using (4.26) is shown in Figure 4.9, along with spectrograms of clean and noisy speech for reference. It shows values close to one (indicated by black regions) in presence of speech and values close to zero (indicated by white regions) in absence of speech.

The speech presence probability is used to calculate the convergence factor

$$\lambda_s(n,k) = \begin{cases} \dfrac{1}{256}\left[1 - \dfrac{255\,(p_{ss}(n,k) - 0.5)}{256}\right], & p_{ss}(n,k) \geq 0.5 \\ \dfrac{1}{256} & \text{otherwise} \end{cases} \tag{4.27}$$

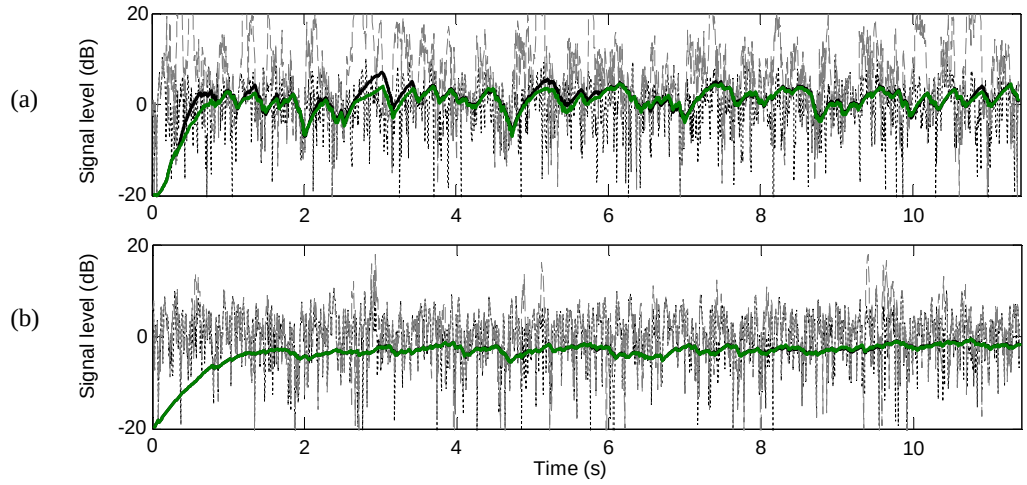


Figure 4.10 Noise estimation example for speech degraded by white noise at 3 dB SNR for two frequency samples: (a) k as 15 (293 Hz) and (b) k as 90 (1.75 kHz); with noisy speech as thin dashed gray trace; noise as thin dotted black trace; estimated noise using p as 0.25 and λ as 1/256 as thick solid black trace; estimated noise using p as 0.25 and λ_s as calculated using (4.27) as thick solid green trace.

which is 1/256 for $p_{ss}(n, k) < 0.5$ and decreases linearly to 1/512 for $p_{ss}(n, k) = 1$. Estimation of p_i-quantile using a convergence factor dependent on the speech presence probability can be written as the following:

$$\hat{q}_{u|X|,i}(n,k) = \begin{cases} \hat{q}_{u|X|,i}(n-1,k) \; + \; \lambda_s(n,k)\, p_i R_{|X|}(n,k), & |X(n,k)| \geq \hat{q}_{u|X|,i}(n-1,k) \\ \hat{q}_{u|X|,i}(n-1,k) \; - \; \lambda_s(n,k)(1-p_i) R_{|X|}(n,k), & \text{otherwise} \end{cases} \tag{4.28}$$

The quantile estimated using $\lambda_s(n, k)$, as in (4.28), is further smoothed recursively using speech presence probability as

$$\hat{q}_{|X|,i}(n,k) = \hat{q}_{u|X|,i}(n-1,k)\, p_{ss}(n,k) \; + \hat{q}_{u|X|,i}(n,k)(1-p_{ss}(n,k)) \tag{4.29}$$

An example of tracking of 0.25-quantile estimated from the noisy speech input using (4.28) and (4.29) for two frequencies, with $\lambda_s(n, k)$ as calculated in (4.27), is shown as thick solid green trace in Figure 4.10. The noisy speech and the noise at two frequencies are shown as thin dashed gray trace and thin dotted black trace, respectively. The estimation using λ as 1/256 is also shown for comparison as thick solid black trace. For frequency corresponding to k as 15 where speech is present, the noisy speech trace shows deviations from the noise trace. For frequency corresponding to k as 90 where speech is absent, the two traces are similar. For k as 90, the noise estimation using λ as 1/256 and that using $\lambda_s(n, k)$ are the same. For k as 15, noise estimation using $\lambda_s(n, k)$ shows smaller deviations from noise trace as compared to that using λ as 1/256. Thus use of speech presence probability to calculate the convergence factor tracks the noise spectrum for both frequencies, showing the suitability of this approach for noise tracking with variable SNR.

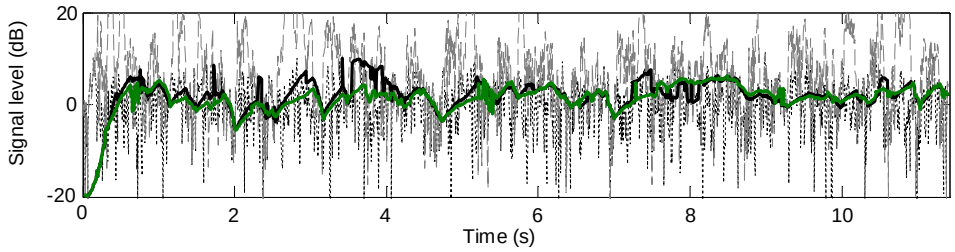


Figure 4.11 Noise estimation example for speech degraded by white noise at 3 dB SNR for k as 15 with noisy speech as thin dashed gray trace, actual noise as thin dotted black trace, noise estimated using adaptive p and λ as 1/256 as thick solid black trace, and noise estimated using adaptive p and λ_s as calculated using (4.27) as thick solid green trace.

Table 4.1 Parameters used in proposed DQTNE and ADQTNE techniques and their optimal values.

Noise estimation	Parameter	Optimum value
DQTNE	Range estimation: τ_p, σ_p, τ_v, σ_v	$\tau_p = \tau_v = 0.1$, $\sigma_p = (1 - \tau_p)^{1/64} = 0.9984$, $\sigma_v = (1 - \tau_v)^{1/1024} = 0.9999$
	Convergence factor: λ, p	$\lambda = 1/256$, $p = 0.25$
ADQTNE	Range estimation: τ_p, σ_p, τ_v, σ_v	$\tau_p = \tau_v = 0.1$, $\sigma_p = (1 - \tau_p)^{1/64} = 0.9984$, $\sigma_v = (1 - \tau_v)^{1/1024} = 0.9999$
	Convergence factor calculation: α, β	$\alpha = 0.2$, $\beta = 0.95$

The quantiles estimated using (4.29) are used for estimating adaptive quantile using (4.22). An example of noise estimated using adaptive quantile as calculated in (4.22) with adaptive λ_s at the frequency corresponding to k as 15 is shown in Figure 4.11, along with noise estimated using adaptive quantile with λ as 1/256, and noisy speech and noise for reference. The noise estimate using λ_s is close to the noise trace whereas the noise estimate using λ as 1/256 shows deviations from the noise trace in presence of speech.

4.4 Evaluation of Noise Estimation

Comparison of the proposed noise estimation techniques with some of the existing ones in terms of computational complexity, evaluation using a visual examination of spectrograms of the estimated noise, and quantification of the estimation errors using an objective measure are presented in the following subsections.

4.4.1 Computational requirements

The computations involved in DQTNE and ADQTNE are much lower than the computations involved in the existing quantile-based noise estimation techniques such as those in [13]–[16]. The noise estimation techniques such as MS [71], MCRA2 [81], and UMMSE [85] are computationally efficient and therefore are used for comparison with DQTNE and ADQTNE. The parameters used in DQTNE and ADQTNE techniques along with their optimal values are shown in Table 4.1. The total number of parameters in DQTNE, ADQTNE, MS, MCRA2, and

Table 4.2 Computation steps and operations per frame per spectral sample using ADQTNE, DQTNE, MS, MCRA2, and UMMSE. (K is FFT size)

Technique	Computation steps with number of operations
ADQTNE (5 parameters)	1) Peak and valley calculation: ≤ 2 comparisons, 2 additions, 4 multiplications. 2) Range calculation: 1 addition. 3) Convergence factor calculation: 2 comparisons, (4+1/K) additions, (3+2/K) multiplications. 4) Quantile function calculation: 8 comparisons, 24 additions, 32 multiplications. 5) Adaptive peak calculation: 7 comparisons, 7 additions.
DQTNE (5 parameters)	1) Peak and valley calculation: ≤ 2 comparisons, 2 additions, 4 multiplications. 2) Range calculation: 1 addition. 3) Quantile calculation: 1 comparison, 1 addition, 1 multiplication.
MS [71] (14 parameters)	1) Squaring input noisy spectrum: 1 multiplication. 2) Smoothing parameter calculation: (1+1/K) comparisons, (5+3/K) additions, (4+6/K) multiplications, 1 exponential. 3) Smoothed power calculation: 2 additions, 2 multiplications. 4) Inverse normalized variance calculation: 3 comparisons, 6 additions, 11 multiplications. 5) Bias correction calculation: 4 additions, 4 multiplications. 6) Average normalized variance calculation: (1+1/K) additions, (1+2/K) multiplications, 1/K square root. 7) Running minimum estimation ($\approx$ 1 s): 5+ (14K+17)/15K comparisons, (1/K) additions, (6+1/15) multiplications.
MCRA2 [81] (7 parameters)	1) Smoothing noisy spectrum: 1 addition, 3 multiplications. 2) Minimum tracking: 2 additions, 3 multiplications, 1 comparison. 3) Speech presence probability calculation: 1 addition, 3 multiplications, 1 comparison. 4) Frequency dependent subtraction factor calculation: 1 addition, 1 multiplication. 5) Recursive averaging for noise estimation: 1 addition, 3 multiplications, 1 square root.
UMMSE [85] (7 parameters)	1) Squaring input noisy spectrum: 1 multiplication. 2) *a posteriori* SNR calculation: 1 multiplication. 3) Likelihood ratio calculation: 1 addition, 2 multiplications, 1 comparison, 1 exponent. 4) *a posterior*i speech presence probability calculation: 1 addition, 1 multiplication. 5) Smoothed a posteriori speech presence probability calculation: 1 addition, 2 multiplications. 6) Avoiding stagnation: 2 comparisons. 7) Noise periodogram estimation using speech presence probability: 1 addition, 2 multiplications. 8) Noise PSD estimation: 1 addition, 2 multiplications, 1 square root.

UMMSE are 5, 5, 14, 7, and 7, respectively. A comparison of the computational complexity of ADQTNE and DQTNE with the MS, MCRA2, and UMMSE is shown in Table 4.2. The total operations per frame per spectral sample using ADQTNE, DQTNE, MS, MCRA2, and UMMSE are 95, 12, 58, 22, and 28, respectively. The computational complexity of DQTNE with 0.25-quantile is lowest and it involves 4 additions, 5 multiplications, and 3 comparisons. MCRA2 involves 6 additions, 13 multiplications, 2 comparisons, and a square root operation and has higher computational complexity than DQTNE. ADQTNE involves 37 addition, 39 multiplication, and 19 comparison operations. The computations for ADQTNE are contributed mainly by tracking of quantile function, which requires estimation of multiple quantiles. It has higher computational complexity than MCRA2, but lower than MS that involves 19 additions, 29 multiplications, 9 comparisons, and calculation of exponential operation and UMMSE that

involves 5 additions, 11 multiplications, 3 comparisons, a square root, and calculation of exponential operation.

4.4.2 Noise tracking

The ADQTNE and DQTNE techniques were implemented in MATLAB. Their performances for tracking the noise spectrum from the noisy speech spectrum were evaluated and compared with those of MS, MCRA2, and UMMSE, using the implementations available in [87] and [113]. The evaluation was carried out by visual examination of spectrograms of the estimated noise and using an objective measure for quantification of estimation errors.

The speech material comprised sentences from GRID database [111], consisting of 1000 sentences spoken by 34 speakers (18 male, 16 female). Five test sentences were concatenated to generate a test segment of approximately 12 s duration. Twenty such test segments from six speakers, resulting in 120 test segments, were used for the evaluation. White noise and babble from the NOISEX database [74] and street, station, restaurant, subway, lobby, exhibition, and airport noises from the AURORA database [93] were used for generating noisy speech. A concatenation of restaurant, lobby, and exhibition noises was used to generate a 'triplet' noise as an example of the background noise with fast-changing characteristics. The speech and noise were added for SNR of 12, 9, 6, 3, 0, –3, and –6 dB. To make the SNR calculation independent of the speech activity duration in the utterance, the SNR was calculated using the noise RMS and the active speech level of the clean speech signal, in accordance with method B of [112], using the code provided in [87].

The spectrograms of the clean speech, noisy speech, actual noise, and estimated noise were visually examined for overestimation and underestimation of noise and for instances of speech getting tracked by estimated noise. The error in the noise estimation was quantified using segmental relative estimation error (SREE), [71], [80]. It is the squared sum of errors in the estimated noise spectrum $\widehat{D}(n,k)$ with reference to the actual noise spectrum $D(n,k)$, normalized by the noise power, and averaged over the frames. It is calculated as

$$\text{SREE} = \frac{1}{N}\sum_{n=0}^{N-1}\left(\sum_{k=0}^{K-1}(D(n,k)-\widehat{D}(n,k))^2 \Big/ \sum_{k=0}^{K-1}D^2(n,k)\right) \tag{4.30}$$

where K is the FFT size and N is the total number of frames. The mean error and standard deviation were calculated for the 120 test segments at each SNR for all types of noises. As the SREE does not differentiate between overestimation and underestimation errors, visual examination of spectrograms is used to analyze the nature of the errors.

As an example of noise estimation using different noise estimation techniques is shown in Figure 4.12. It shows the spectrograms of clean speech, noisy speech, actual noise, and noise estimated using the six noise estimation techniques for speech degraded with babble at 3 dB SNR. The spectrogram of the noise estimated using DQTNE and ADQTNE are similar and have a close match with the actual babble. The spectrogram of MS-estimated noise has residual noise indicating underestimation in several time-frequency regions, which may be due to inadequate bias compensation. The spectrogram of MCRA2-estimated noise shows noise overestimation in the high-frequency region, which may be due to the assumption involved in the calculation of speech presence probability that speech is not present above 2 kHz. The spectrogram of UMMSE-estimated noise shows overestimation in the presence of speech, which may occur due to the use of a soft speech presence probability that is calculated assuming that the DFT coefficients have a complex-Gaussian distribution, which may not be valid in presence of speech.

Figure 4.13 shows an example of noise estimation for speech degraded with the triplet noise at 3 dB SNR. The noise estimated using DQTNE and ADQTNE are similar and have a close match with the actual triplet noise. The spectrogram of MS-estimated noise shows under-estimation. The noise estimated using MCRA2 and UMMSE show overestimation in the presence of speech. DQTNE and ADQTNE are able to track the changes in noise characteristics when the noise changes from restaurant noise to lobby noise. However, MS and MCRA2 are slow in tracking the noise. UMMSE tracks the noise, but with an overestimation. The results indicate that DQTNE and ADQTNE result in lower overestimation and under-estimation than MS, MCRA2, and UMMSE.

The error in the noise estimation was quantified using SREE as the error measure as given in (4.30). The means and the standard deviations of the SREE values calculated over the 120 test segments, with different noises at 12, 9, 6, 3, 0, and –3 dB SNR, for the five noise estimation techniques are given in Table 4.3. The standard deviation of the error measure is much smaller than the mean across the techniques, noises, and SNRs, indicating that the error does not vary significantly with speech material. Across noises and SNRs, the errors for ADQTNE, DQTNE, MS, MCRA2, and UMMSE are 0.29−0.84, 0.31−0.65, 0.31−0.61, 0.27−1.44, and 0.24−2.07, respectively. The errors for ADQTNE are lower than those for DQTNE for SNRs below 9 dB, as DQTNE underestimates the noise at low SNRs. At SNRs higher than 6 dB, the errors for ADQTNE are higher than DQTNE. ADQTNE has lower errors than MS for all noises except for street and station noises. As street and station noises are concentrated in lower frequencies overlapping with the speech-dominant frequencies, ADQTNE overestimates the noise in lower frequencies leading to higher errors than MS. The

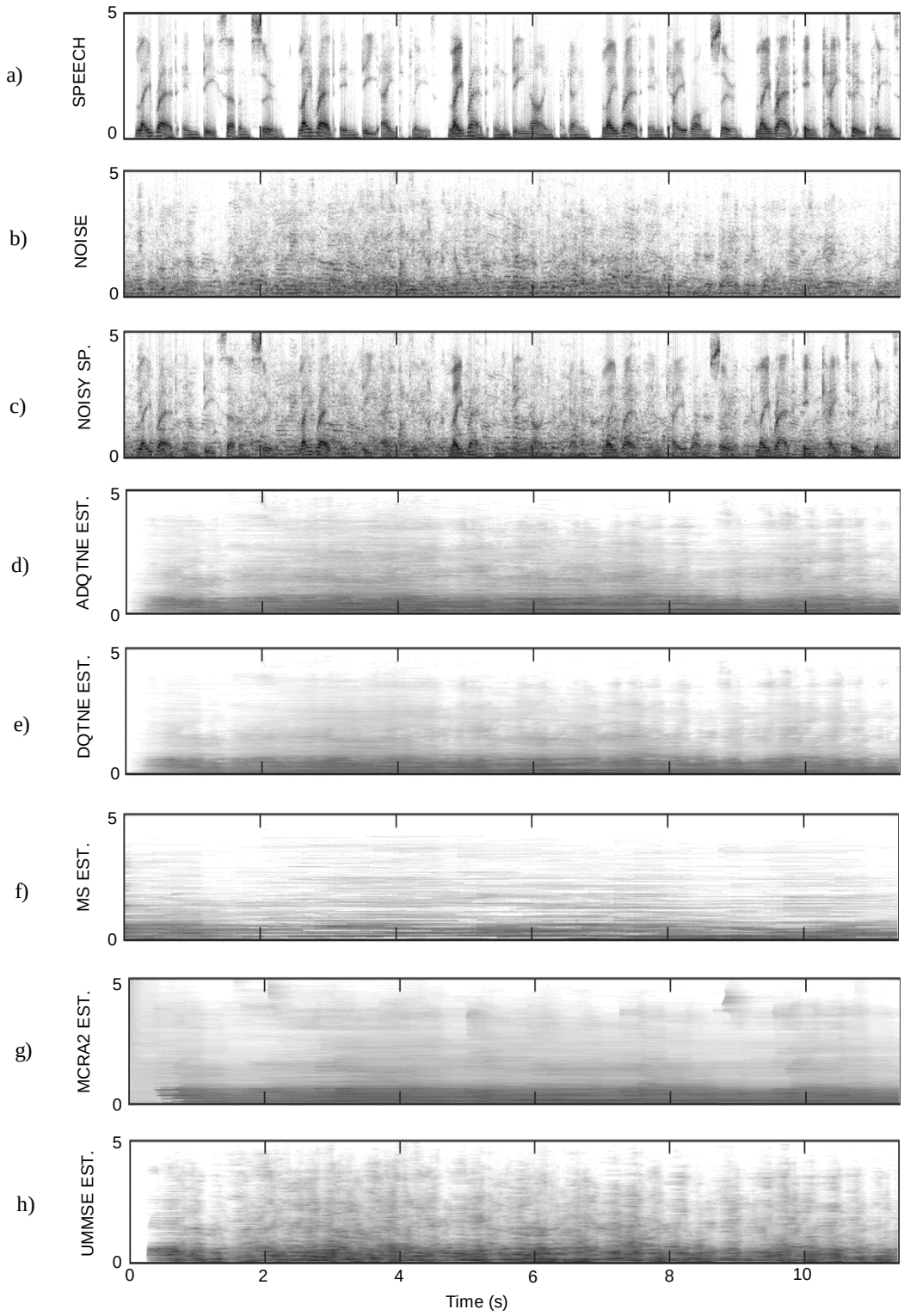


Figure 4.12 Noise tracking for speech degraded by babble at 3 dB SNR: Spectrograms of (a) speech, (b) noise, (c) noisy speech, (d) DQTNE-estimated noise, (e) ADQTNE-estimated noise, (f) MS-estimated noise, (g) MCRA2-estimated noise, and (h) UMMSE-estimated noise, with time (s) on x-axis and frequency (kHz) on y-axis.

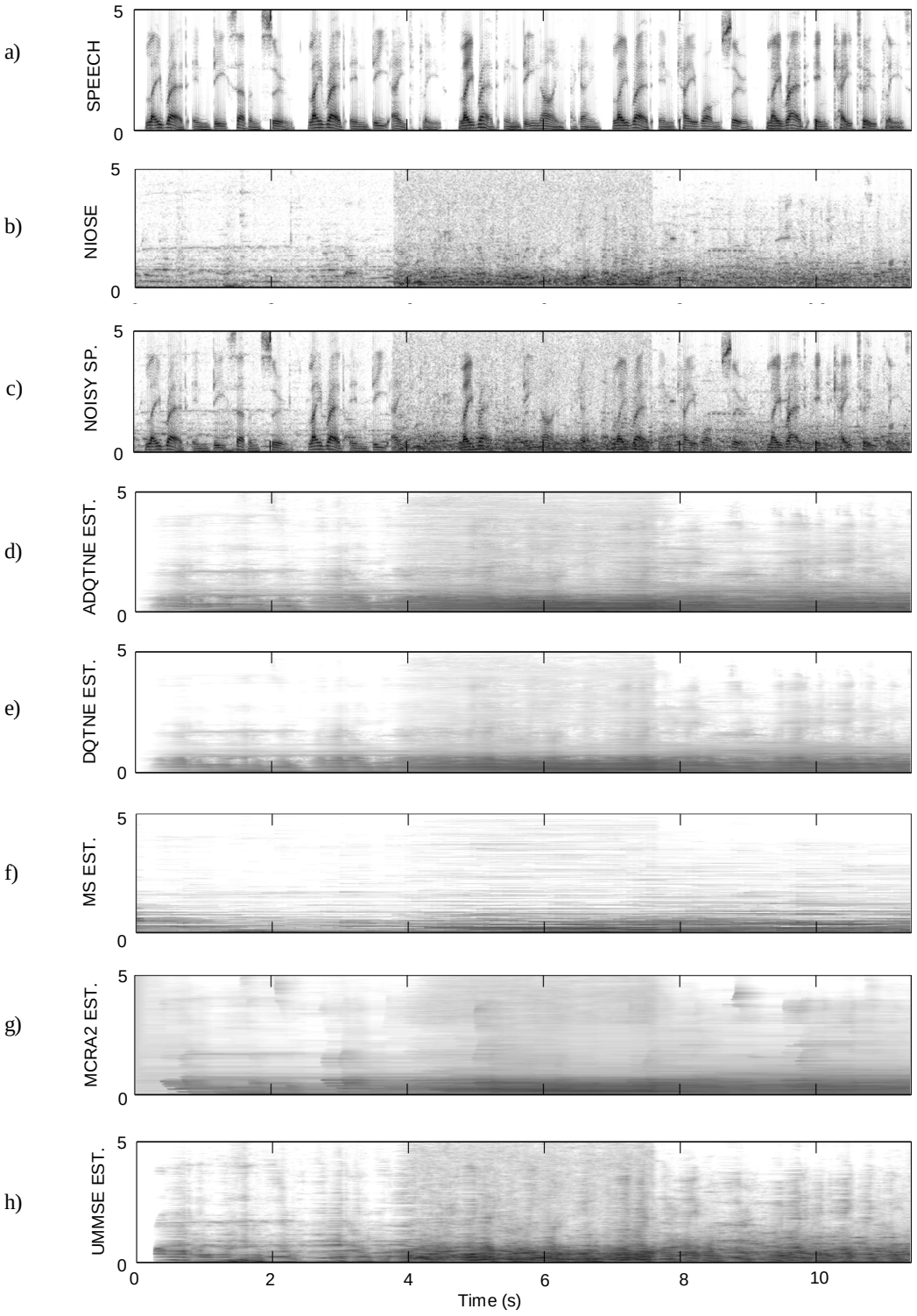


Figure 4.13 Noise tracking for speech degraded by the triplet noise at 3 dB SNR: Spectrograms of (a) speech, (b) noise, (c) noisy signal, (d) DQTNE-estimated noise, (e) ADQTNE-estimated noise, (f) MS-estimated noise, (g) MCRA2-estimated noise, and (h) UMMSE-estimated noise, with time (s) on x-axis and frequency (kHz) on y-axis.

Table 4.3 Segmental relative estimation error (SREE) for noise estimation using ADQTNE, DQTNE, MS [71], MCRA2 [81], and UMMSE [85] (Mean: mean error, S.D.: standard deviation of errors, No. of test segments = 120).

Noise	SNR	SREE									
		ADQTNE-Est.		DQTNE-Est.		MS-Est.		MCRA2-Est.		UMMSE-Est.	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
White	-3	0.30	0.00	0.37	0.01	0.35	0.00	0.27	0.00	0.24	0.01
	0	0.29	0.00	0.37	0.01	0.36	0.01	0.27	0.01	0.27	0.02
	3	0.29	0.01	0.36	0.01	0.38	0.01	0.27	0.01	0.32	0.03
	6	0.30	0.01	0.35	0.01	0.39	0.01	0.30	0.02	0.39	0.06
	9	0.34	0.03	0.37	0.02	0.42	0.01	0.37	0.04	0.52	0.11
	12	0.43	0.06	0.43	0.05	0.45	0.02	0.58	0.08	0.75	0.20
Street	-3	0.41	0.02	0.36	0.01	0.35	0.01	0.33	0.01	0.27	0.02
	0	0.38	0.01	0.36	0.01	0.35	0.01	0.34	0.01	0.32	0.03
	3	0.37	0.01	0.36	0.01	0.36	0.01	0.36	0.01	0.40	0.06
	6	0.37	0.02	0.36	0.02	0.37	0.02	0.42	0.02	0.53	0.11
	9	0.43	0.04	0.41	0.03	0.39	0.03	0.57	0.05	0.76	0.20
	12	0.59	0.09	0.53	0.06	0.43	0.05	0.92	0.10	1.13	0.37
Station	-3	0.36	0.02	0.33	0.01	0.31	0.01	0.28	0.01	0.24	0.01
	0	0.33	0.02	0.32	0.01	0.31	0.01	0.28	0.01	0.28	0.02
	3	0.30	0.01	0.31	0.01	0.32	0.01	0.29	0.01	0.35	0.04
	6	0.30	0.01	0.31	0.02	0.33	0.02	0.32	0.02	0.46	0.08
	9	0.33	0.02	0.33	0.02	0.35	0.02	0.41	0.03	0.66	0.16
	12	0.43	0.05	0.40	0.03	0.38	0.03	0.65	0.07	0.97	0.29
Babble	-3	0.38	0.01	0.42	0.01	0.46	0.01	0.38	0.01	0.31	0.01
	0	0.37	0.00	0.40	0.01	0.46	0.01	0.37	0.01	0.37	0.03
	3	0.36	0.00	0.38	0.01	0.47	0.01	0.36	0.01	0.47	0.05
	6	0.36	0.01	0.36	0.01	0.49	0.02	0.38	0.01	0.63	0.10
	9	0.40	0.02	0.37	0.01	0.51	0.02	0.46	0.03	0.90	0.19
	12	0.50	0.06	0.42	0.04	0.54	0.04	0.67	0.07	1.34	0.32
Restaurant	-3	0.38	0.01	0.41	0.01	0.43	0.01	0.38	0.01	0.33	0.02
	0	0.37	0.01	0.40	0.01	0.44	0.01	0.38	0.01	0.40	0.04
	3	0.37	0.01	0.38	0.01	0.46	0.01	0.39	0.01	0.51	0.08
	6	0.39	0.02	0.38	0.01	0.48	0.02	0.43	0.02	0.68	0.13
	9	0.44	0.04	0.41	0.03	0.50	0.03	0.55	0.05	0.95	0.23
	12	0.59	0.09	0.50	0.06	0.53	0.05	0.85	0.11	1.38	0.41
Lobby	-3	0.38	0.01	0.41	0.01	0.44	0.01	0.37	0.01	0.30	0.02
	0	0.37	0.01	0.39	0.01	0.45	0.02	0.36	0.01	0.37	0.04
	3	0.36	0.01	0.37	0.01	0.46	0.02	0.36	0.01	0.48	0.07
	6	0.36	0.01	0.36	0.01	0.48	0.02	0.37	0.01	0.66	0.11
	9	0.40	0.02	0.36	0.01	0.50	0.02	0.45	0.03	0.95	0.21
	12	0.50	0.06	0.42	0.04	0.53	0.03	0.67	0.06	1.40	0.37
Exhibition	-3	0.37	0.01	0.41	0.01	0.43	0.01	0.36	0.01	0.28	0.01
	0	0.35	0.01	0.39	0.01	0.43	0.01	0.35	0.01	0.34	0.03
	3	0.34	0.01	0.37	0.01	0.45	0.01	0.34	0.01	0.44	0.05
	6	0.34	0.01	0.35	0.01	0.47	0.02	0.36	0.01	0.61	0.10
	9	0.37	0.02	0.35	0.01	0.49	0.02	0.42	0.02	0.86	0.18
	12	0.47	0.04	0.40	0.03	0.52	0.03	0.60	0.06	1.26	0.32
Triplet	-3	0.40	0.01	0.42	0.01	0.49	0.01	0.39	0.01	0.36	0.05
	0	0.39	0.01	0.40	0.01	0.50	0.02	0.40	0.01	0.46	0.08
	3	0.39	0.01	0.39	0.01	0.51	0.02	0.43	0.02	0.62	0.14
	6	0.42	0.03	0.40	0.02	0.53	0.03	0.53	0.04	0.87	0.25
	9	0.54	0.06	0.46	0.04	0.56	0.04	0.80	0.10	1.29	0.43
	12	0.84	0.15	0.65	0.11	0.61	0.07	1.44	0.22	2.07	0.82

errors for ADQTNE are comparable to MCRA2 at –3, 0, and 3 dB and are much lower at higher SNRs, where MCRA2 overestimates the noise. UMMSE has highest errors for SNRs greater than 3 dB due to overestimation of noise. For SNRs of 9 and 12 dB, UMMSE has SREE values greater than one. The spectrograms show that some of the speech regions were tracked as part of noise in these cases.

The plots of mean SREE vs SNR for different noise estimation techniques are shown in Figure 4.14. It can be seen that considering all SNRs and different types of noises, the error measures are lowest for ADQTNE and the techniques can be ranked as ADQTNE, DQTNE, MCRA2, MS, and UMMSE.

4.5 Evaluation in a Speech Enhancement Framework

The two proposed noise estimation techniques, as described in Section 4.3, were used for speech enhancement along with the spectral subtraction based on geometric approach (GA) as described in [90]. Unlike spectral subtraction using power spectrum, GA does not assume speech and noise to be uncorrelated and thus the cross terms between the speech and noise spectrum are not assumed to be zero in the subtraction, and it and has been reported to result in smaller residual noise. The speech enhancement framework used for the evaluation, evaluation method, and the evaluation results are presented in the following subsections.

4.5.1 Noise suppression using geometric approach to spectral subtraction

Figure 4.15 shows a block diagram of speech enhancement using GA-based noise suppression. The processing comprises windowing, FFT calculation, magnitude spectrum calculation, noise spectrum estimation, SNR-dependent gain calculation, enhanced complex spectrum calculation, IFFT calculation, and resynthesis using overlap-add. Windowed segments of the input $x(m)$ are used as the analysis frames and FFT is used to obtain the complex spectrum. The magnitude spectrum, calculated from the complex spectrum, is used for noise estimation. The noise spectrum is estimated using the noise estimation techniques as described in Section 4.3. The gain dependent on the cross terms is calculated using the estimates of *a priori* and *a posteriori* SNRs. The SNR estimates are calculated using the estimate of noise magnitude, previous enhanced magnitude, and noisy input magnitude. The enhanced spectrum is obtained by multiplying the input complex spectrum with the SNR-dependent gain. The output is resynthesized using IFFT and overlap-add.

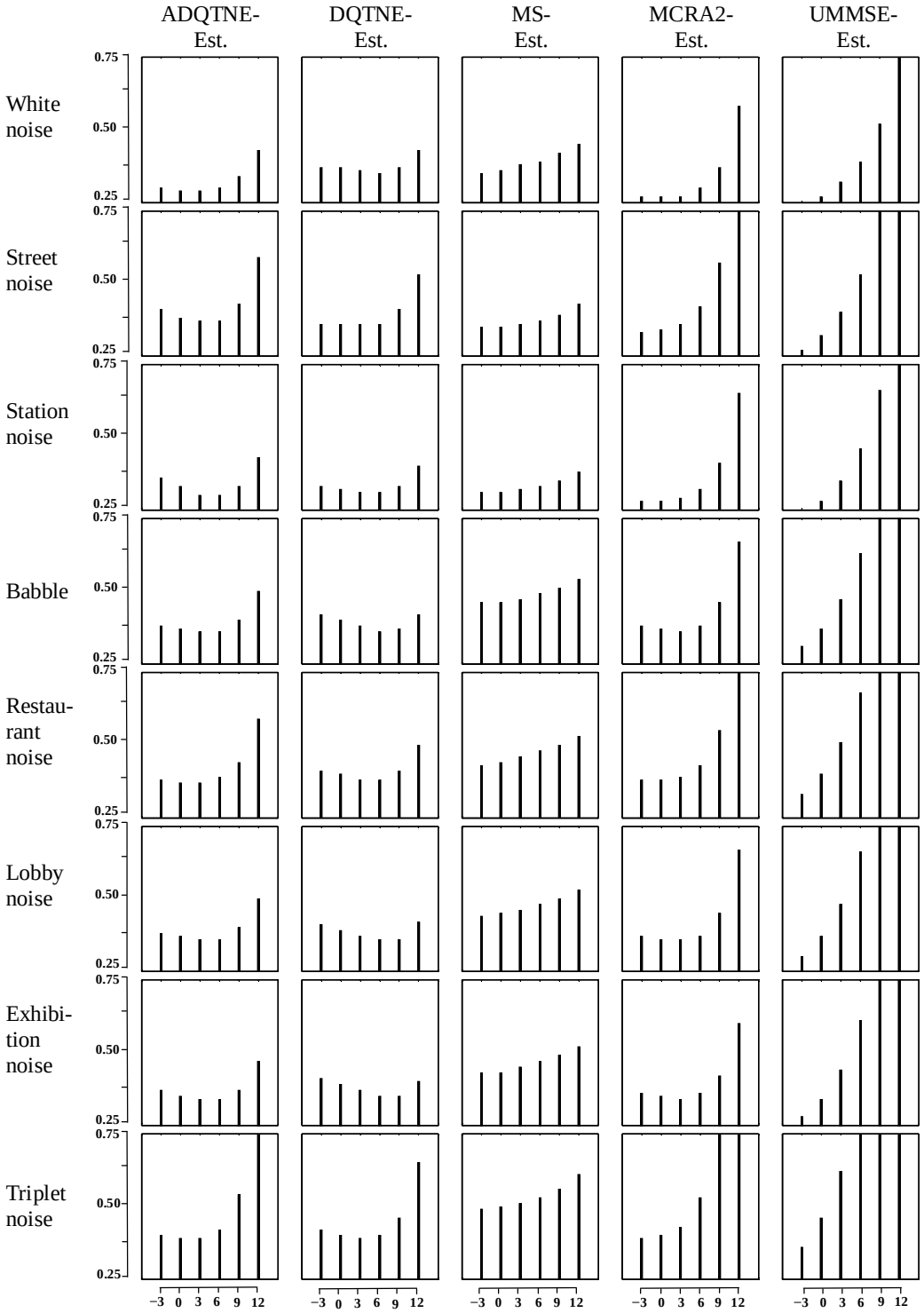


Figure 4.14 Noise tracking: SREE (0.25−0.75, on y-axis) vs SNR (−3, 0, 3, 6, 9, 12 dB, on x-axis) for noise estimation using ADQTNE, DQTNE, MS [71], MCRA2 [81], and UMMSE [85].

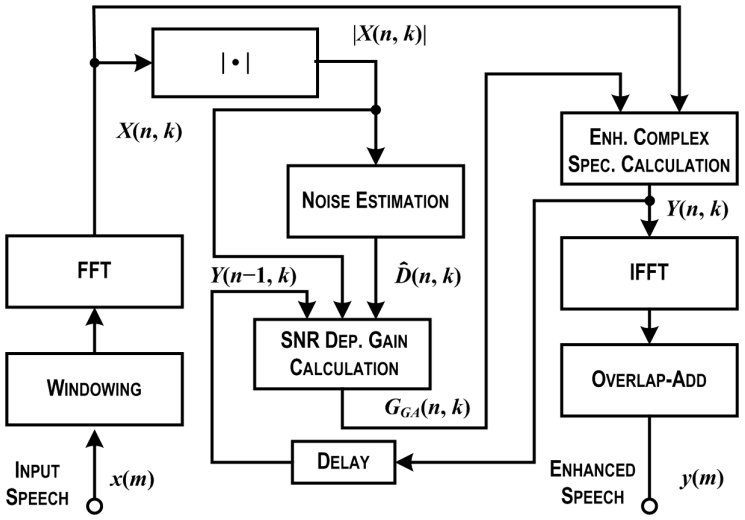


Figure 4.15 Speech enhancement by spectral subtraction based on geometric approach.

The estimated noise magnitude $\hat{D}(n,k)$ is used to calculate instantaneous *a posteriori* SNR as $(|X(n, k)|/ |\hat{D}(n,k)|)^2$, which is recursively averaged with an upper bound to calculate the smoothed *a posteriori* SNR $\psi(n, k)$ as

$$\psi(n,k) = \rho\psi(n-1,k) + (1-\rho)\min\left(\left(\frac{|X(n,k)|}{|\hat{D}(n,k)|}\right)^2, 20\right) \tag{4.31}$$

where ρ is the smoothing constant. The *a priori* SNR $\xi(n, k)$ is estimated from the previous-frame enhanced magnitude spectrum $Y(n-1,k)$, smoothed *a posteriori* SNR, and a decision-directed approach with a weighting factor κ as

$$\xi(n,k) = \kappa\left(\frac{|Y(n-1,k)|}{|\hat{D}(n-1,k)|}\right)^2 + (1-\kappa)\left(\sqrt{\psi(n,k)} - 1\right)^2 \tag{4.32}$$

Smoothing constant ρ and weighting factor κ are selected as 0.6 and 0.98, respectively, as reported in [90]. The SNR-dependent gain function at $G_{GA}(n, k)$ is calculated using smoothened *a posteriori* SNR estimate $\psi(n, k)$ and smoothened *a priori* SNR estimate $\xi(n, k)$ as

$$G_{GA}(n,k) = \sqrt{\left(1 - \frac{(\psi(n,k)+1-\xi(n,k))^2}{4\psi(n,k)}\right)\left(1 - \frac{(\psi(n,k)-1-\xi(n,k))^2}{4\xi(n,k)}\right)^{-1}} \tag{4.33}$$

The enhanced spectrum $Y(n, k)$ is obtained by multiplying the input spectrum $X(n, k)$ with the SNR-dependent gain function $G_{GA}(n, k)$, as

$$Y(n,k) = G_{GA}(n,k)X(n,k) \tag{4.34}$$

4.5.2 Evaluation method

For speech enhancement, the implementation of the GA-based spectral subtraction available in [114] was used along with noise estimation using ADQTNE and DQTNE. For comparison spectral subtraction was also implemented along with noise estimation using MS [71], MCRA2 [81], and UMMSE [85]. The implementation used FFT-based analysis-synthesis, with signal sampling frequency of 10 kHz, 256-point window (L = 256) with 64-point shift (S = 64) corresponding to 25.6-ms frames with 75% frame overlap, and 512-point FFT (K = 512). The speech and noise materials were the same as used for evaluation of noise estimation as described in Section 4.4.

The evaluation was carried out using informal listening, visual inspection of the spectrograms, and objective measure. In informal listening, the perceptual quality of the processed speech was examined in terms of residual noise, roughness, musical noise, and attenuation of weak speech segments. The spectrograms of the clean speech, noisy speech, actual noise, and enhanced speech were visually compared to examine the underestimation and overestimation of noise, distortions, and preservation of weak regions of speech. The perceptual evaluation of speech quality (PESQ) measure [94] was calculated for an objective evaluation. This measure, based on the difference between the loudness spectra of the level-equalized and time-aligned processed output and clean speech signals, provides a score on 0–4.5 scale.

4.5.3 Evaluation results

Informal listening indicated no audible roughness or musical noise in any of the processed outputs. In terms of low residual noise across the SNRs, the techniques could be ranked as UMMSE, MCRA2, ADQTNE, DQTNE, and MS. In terms of low speech attenuation, the techniques could be ranked as MS, ADQTNE, DQTNE, MCRA2, and UMMSE. MS had significant residual noise than other techniques, whereas UMMSE and MCRA2 had significant speech attenuation. Considering residual noise and speech attenuation together, ADQTNE appeared to provide the highest quality output. The output of DQTNE was similar to that of ADQTNE, except for having a higher residual noise at low SNRs.

Figure 4.16 shows an example of the processing of speech degraded with babble at −3 dB SNR, showing the spectrograms of noise-free speech, noisy speech, and outputs of processing using ADQTNE, DQTNE, MS, MCRA2, and UMMSE. ADQTNE has lower residual noise than DQTNE. The highest residual noise is shown by MS, indicating an underestimation of noise. UMMSE shows attenuation of speech regions, indicating an overestimation of noise in

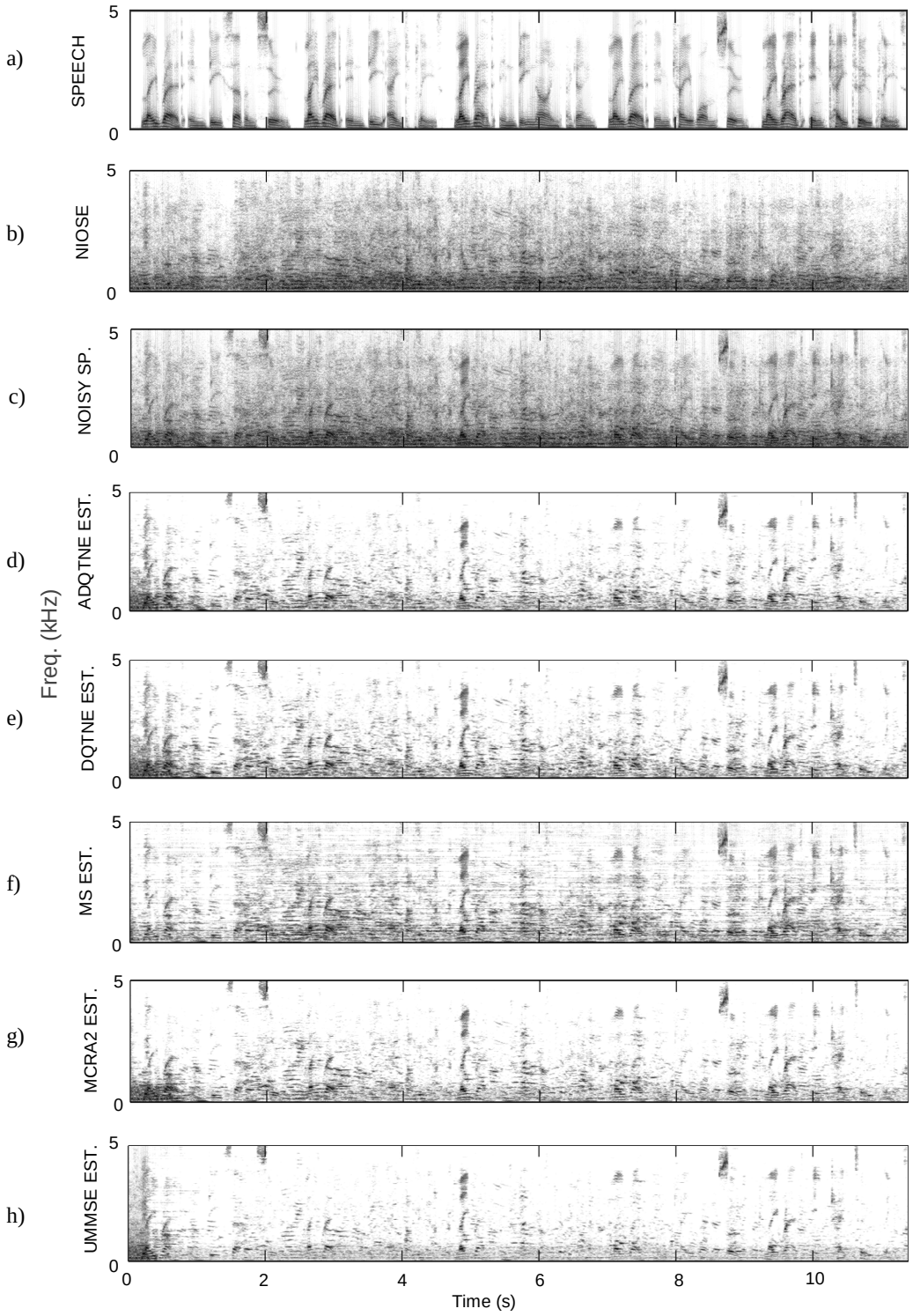


Figure 4.16 Processing of a sentence with babble at SNR of −3 dB: Spectrograms of (a) clean, (b) noise, (c) noisy signal, (d) DQTNE-enhanced speech, (e) ADQTNE-enhanced speech, (f) MS-enhanced speech, (g) MCRA2-enhanced speech, and (h) UMMSE-enhanced speech.

the presence of speech. MCRA2 shows lower residual noise than MS and DQTNE, but it shows attenuation of some speech regions above 2 kHz. An example of the processing for speech degraded with the triplet noise at −3 dB SNR is shown in Figure 4.17. It can be observed that ADQTNE, DQTNE, MCRA2, and UMMSE were able to track the changing noise characteristics. However, MS could not track the changes and resulted in excessive residual noise in the output. UMMSE shows attenuation of speech regions, indicating an overestimation of noise in the presence of speech. MCRA2 has lower residual noise than DQTNE, but it shows attenuation of some speech regions above 2 kHz. ADQTNE shows lower residual noise than DQTNE and no visible attenuation of speech regions. Visual examination of the spectrograms of the output signals for other noises showed similar results, confirming the observations from informal listening that ADQTNE and DQTNE were able to track nonstationary noises and provided speech enhancement without noticeable speech attenuation.

The performances of different noise estimation techniques for speech enhancement were compared using PESQ scores for objective evaluation. The means and standard deviations of PESQ scores, calculated over the 120 test segments, for unprocessed noisy input and the speech enhanced using different techniques, for different noises and SNRs, are given in Table 4.4. The mean scores for the unprocessed speech are lowest for white noise and highest for street noise. The standard deviations are highest for exhibition noise and lowest for white noise. Mean scores for enhanced speech obtained using all the noise estimation techniques are higher than the unprocessed scores. For most of the processing conditions, the ADQTNE scores are higher than the other scores, with the MS scores being the lowest. The DQTNE scores are lower than the ADQTNE scores for SNR of 6 dB and lower. These scores are similar for SNRs higher than 6 dB. The MCRA2 scores are lower than the ADQTNE and DQTNE scores, particularly for SNRs higher than 6 dB. The UMMSE scores are higher than the MS scores but lower than the MCRA2 scores.

Figure 4.18 shows the plots of mean PESQ score as a function of SNR for the unprocessed and processed signals for babble and triplet noise, for different techniques. For unprocessed speech, the score increased monotonically from 1.36 at SNR of −3 dB to 2.33 at SNR of 12 dB for babble and from 1.37 at SNR of −3 dB to 2.51 at SNR of 12 dB for triplet noise. The scores for babble were lower than the corresponding ones for triplet noise at all SNRs. The processing using all techniques resulted in increased scores. The increase in PESQ scores after processing as a function of SNR may also be interpreted as an equivalent SNR advantage, which was calculated as the difference between the corresponding SNRs for the

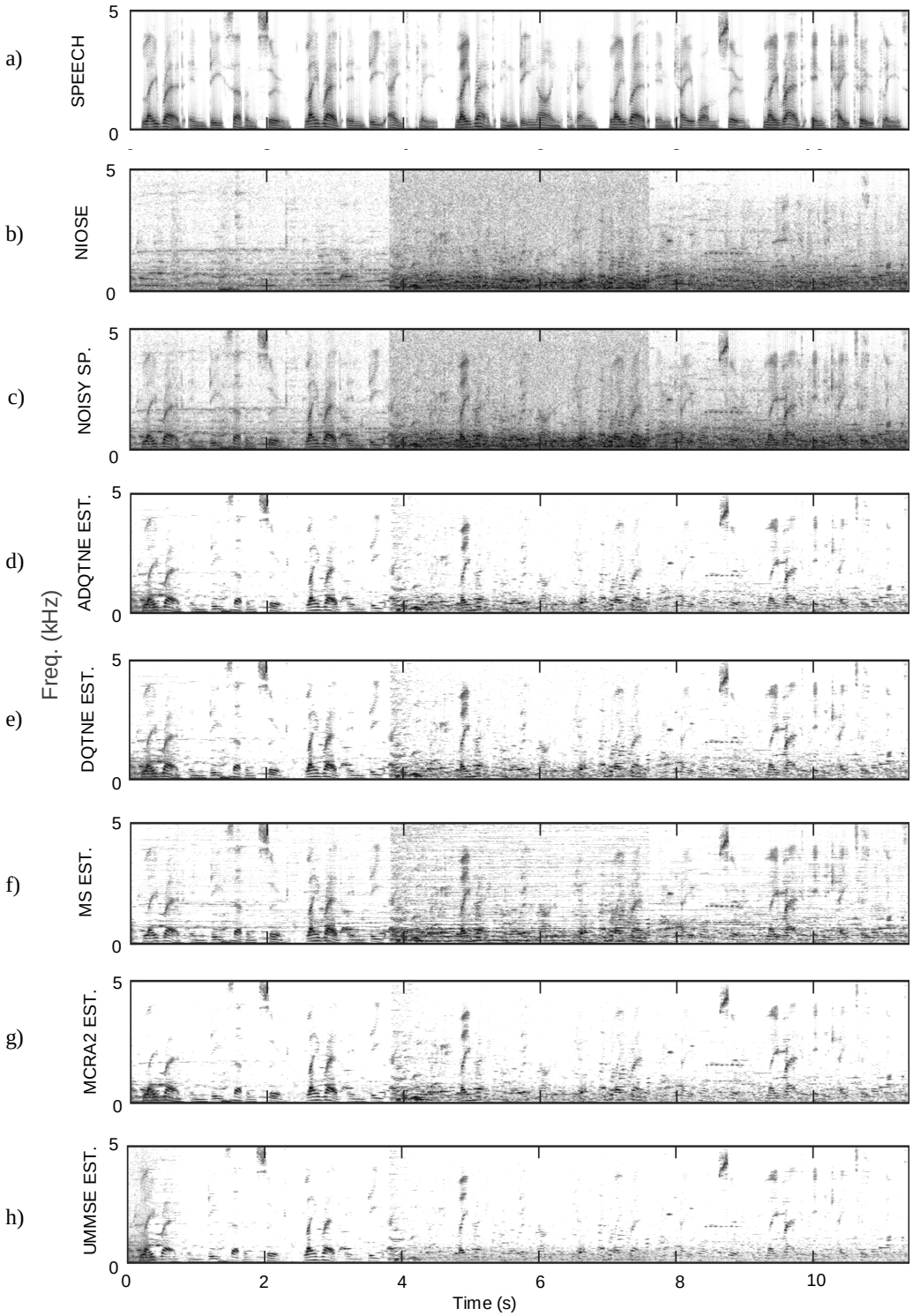


Figure 4.17 Processing of a sentence with triplet noise at SNR of −3 dB: Spectrograms of (a) clean, (b) noise, (c) noisy signal, (d) DQTNE-enhanced speech, (e) ADQTNE-enhanced speech, (f) MS-enhanced speech, (g) MCRA2-enhanced speech, and (h) UMMSE-enhanced speech.

Table 4.4 PESQ scores for unprocessed noisy speech and for speech enhanced using the ADQTNE, DQTNE, MS [71], MCRA2 [81], and UMMSE [85] (Mean: mean score, S.D.: standard deviation of scores, No. of test segments = 120).

Noise	SNR	PESQ											
		Unproc.		ADQTNE-Enh.		DQTNE-Enh.		MS-Enh.		MCRA2-Enh.		UMMSE-Enh.	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
White	-3	1.25	0.17	1.49	0.17	1.49	0.16	1.42	0.15	1.50	0.17	1.47	0.16
	0	1.34	0.17	1.70	0.17	1.68	0.17	1.60	0.15	1.70	0.17	1.68	0.16
	3	1.45	0.17	1.94	0.17	1.90	0.17	1.81	0.14	1.92	0.17	1.92	0.15
	6	1.59	0.17	2.18	0.16	2.13	0.17	2.01	0.14	2.17	0.17	2.17	0.14
	9	1.76	0.16	2.42	0.15	2.37	0.16	2.21	0.14	2.41	0.15	2.40	0.13
	12	1.96	0.16	2.64	0.13	2.60	0.15	2.42	0.14	2.61	0.13	2.61	0.11
Street	-3	1.73	0.19	2.20	0.16	2.19	0.16	2.03	0.16	2.20	0.15	2.17	0.16
	0	1.94	0.17	2.44	0.14	2.43	0.15	2.23	0.16	2.43	0.13	2.38	0.14
	3	2.16	0.16	2.66	0.13	2.65	0.13	2.44	0.15	2.63	0.12	2.58	0.13
	6	2.37	0.15	2.85	0.12	2.85	0.12	2.65	0.15	2.78	0.11	2.76	0.12
	9	2.58	0.14	3.01	0.12	3.02	0.12	2.85	0.14	2.90	0.10	2.95	0.12
	12	2.78	0.13	3.15	0.11	3.18	0.12	3.05	0.14	2.99	0.09	3.11	0.12
Station	-3	1.65	0.22	2.17	0.16	2.15	0.16	1.96	0.18	2.15	0.16	2.13	0.15
	0	1.87	0.2	2.42	0.13	2.40	0.14	2.18	0.17	2.39	0.13	2.33	0.13
	3	2.08	0.18	2.63	0.11	2.62	0.12	2.40	0.16	2.60	0.11	2.52	0.11
	6	2.29	0.16	2.81	0.10	2.81	0.11	2.61	0.14	2.76	0.09	2.69	0.10
	9	2.5	0.15	2.97	0.09	2.99	0.10	2.81	0.14	2.88	0.08	2.86	0.11
	12	2.71	0.14	3.11	0.10	3.14	0.10	3.00	0.14	2.96	0.08	3.04	0.11
Babble	-3	1.36	0.26	1.52	0.18	1.49	0.18	1.50	0.19	1.49	0.18	1.48	0.19
	0	1.51	0.23	1.77	0.17	1.74	0.18	1.69	0.19	1.75	0.18	1.74	0.19
	3	1.7	0.21	2.04	0.17	1.99	0.17	1.90	0.18	2.01	0.16	2.00	0.17
	6	1.91	0.2	2.29	0.15	2.25	0.15	2.11	0.17	2.26	0.15	2.24	0.14
	9	2.12	0.18	2.53	0.13	2.50	0.14	2.33	0.15	2.49	0.13	2.46	0.12
	12	2.33	0.16	2.75	0.11	2.73	0.12	2.55	0.15	2.69	0.11	2.67	0.11
Restau-rant	-3	1.41	0.21	1.51	0.17	1.51	0.16	1.51	0.16	1.47	0.16	1.42	0.18
	0	1.57	0.2	1.79	0.17	1.78	0.17	1.74	0.17	1.76	0.17	1.71	0.17
	3	1.75	0.19	2.07	0.15	2.05	0.15	1.97	0.16	2.04	0.15	2.00	0.15
	6	1.95	0.18	2.33	0.14	2.31	0.15	2.19	0.15	2.29	0.14	2.26	0.14
	9	2.16	0.16	2.56	0.12	2.55	0.13	2.41	0.15	2.51	0.12	2.49	0.12
	12	2.37	0.15	2.77	0.11	2.77	0.12	2.62	0.14	2.70	0.11	2.71	0.12
Lobby	-3	1.48	0.25	1.94	0.17	1.91	0.17	1.80	0.16	1.93	0.16	1.90	0.17
	0	1.67	0.23	2.18	0.16	2.14	0.17	2.00	0.16	2.16	0.16	2.12	0.16
	3	1.88	0.21	2.41	0.15	2.38	0.16	2.21	0.16	2.39	0.15	2.34	0.14
	6	2.1	0.19	2.63	0.13	2.60	0.15	2.42	0.15	2.60	0.13	2.55	0.13
	9	2.31	0.17	2.83	0.12	2.81	0.13	2.62	0.15	2.76	0.11	2.76	0.12
	12	2.51	0.16	2.99	0.11	2.99	0.12	2.82	0.14	2.88	0.10	2.94	0.12
Exhibi-tion	-3	1.35	0.26	1.75	0.18	1.72	0.18	1.67	0.20	1.73	0.18	1.72	0.18
	0	1.54	0.23	2.01	0.17	1.98	0.17	1.88	0.19	1.99	0.16	1.98	0.17
	3	1.76	0.22	2.25	0.15	2.22	0.15	2.09	0.17	2.23	0.15	2.20	0.15
	6	1.98	0.2	2.48	0.12	2.46	0.13	2.30	0.16	2.45	0.12	2.41	0.13
	9	2.19	0.18	2.68	0.11	2.68	0.12	2.51	0.15	2.64	0.10	2.59	0.11
	12	2.41	0.16	2.87	0.10	2.88	0.10	2.71	0.13	2.80	0.09	2.77	0.11
Triplet	-3	1.37	0.23	1.55	0.18	1.53	0.18	1.48	0.19	1.53	0.17	1.52	0.18
	0	1.56	0.21	1.84	0.16	1.81	0.17	1.73	0.19	1.82	0.17	1.81	0.17
	3	1.77	0.2	2.12	0.15	2.09	0.16	1.96	0.18	2.10	0.15	2.08	0.15
	6	2.09	0.18	2.37	0.13	2.34	0.14	2.19	0.16	2.35	0.13	2.30	0.13
	9	2.3	0.17	2.60	0.12	2.58	0.12	2.41	0.15	2.56	0.11	2.51	0.12
	12	2.51	0.15	2.80	0.10	2.80	0.11	2.62	0.14	2.73	0.10	2.70	0.11

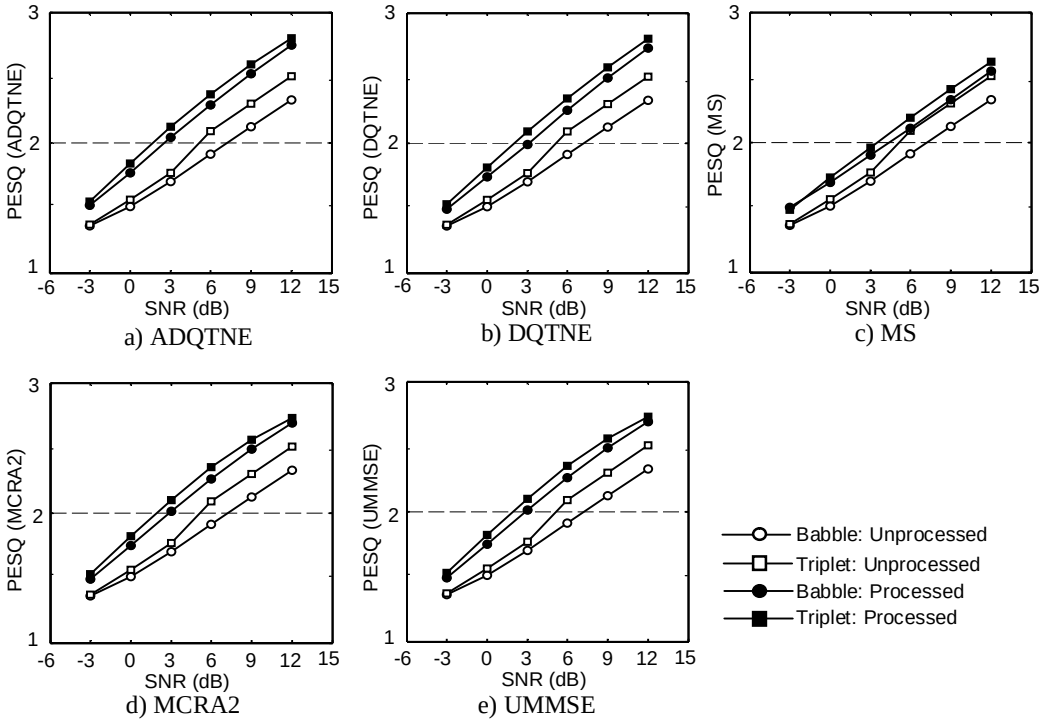


Figure 4.18 PESQ scores as a function of SNR for babble and triplet noise: (a) ADQTNE, (b) DQTNE, (c) MS, (d) MCRA2, and (e) UMMSE.

processed and unprocessed signals for a score of 2 (generally considered as acceptable for speech). For babble, the SNR advantages using ADQTNE, DQTNE, MCRA2, UMMSE, and MS were 4.8, 4.3, 4.3, 4.2, and 2.8 dB, respectively. For triplet noise, the SNR advantage obtained using ADQTNE, MCRA2, DQTNE, UMMSE, and MS were 4.0, 3.3, 3.1, 3.1, and 1.6 dB, respectively. Across different noises, the SNR advantages using ADQTNE, DQTNE, MCRA2, UMMSE, and MS ranged 4−11, 3−10, 3−10, 3−10, and 2−9 dB. Thus, DQTNE, MCRA2, and UMMSE provided almost the same SNR advantage and ADQTNE may be considered as the best technique.

The standard deviation of the scores for unprocessed speech, across noises and SNRs, were 0.14−0.26, indicating that the degrading effect of noise varied significantly across the speech segments used for testing. Therefore, the means and standard deviations of the improvements in the scores were calculated for a paired comparison across the test segments for each of the noise and SNR combinations. These results are given in Table 4.5. The mean improvements for all the techniques were statistically significant at $p < 0.001$ (for one-tailed t-test). The improvement obtained using ADQTNE, DQTNE, MS, MCRA2, and UMMSE are in the range 0.10−0.68, 0.10−0.64, 0.10−0.46, 0.06−0.66, 0.02−0.65, with ADQTNE showing

Table 4.5 Improvements in PESQ scores for speech enhanced using the ADQTNE, DQTNE, MS [71], MCRA2 [81], and UMMSE [85] over unprocessed noisy speech (Mean: mean improvement in PESQ scores, S.D.: standard deviation of improvement in PESQ scores, No. of test segments = 120).

Noise	SNR	Δ PESQ									
		ADQTNE-Est.		DQTNE-Est.		MS-Est.		MCRA2-Est.		UMMSE-Est.	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
White	-3	0.24	0.12	0.23	0.11	0.16	0.08	0.25	0.11	0.22	0.11
	0	0.36	0.14	0.34	0.12	0.26	0.09	0.36	0.14	0.34	0.13
	3	0.49	0.14	0.45	0.14	0.36	0.10	0.48	0.15	0.48	0.14
	6	0.60	0.14	0.55	0.14	0.43	0.09	0.59	0.15	0.58	0.13
	9	0.66	0.13	0.61	0.13	0.45	0.08	0.65	0.13	0.64	0.12
	12	0.68	0.11	0.64	0.11	0.46	0.06	0.66	0.12	0.65	0.10
Babble	-3	0.16	0.17	0.13	0.16	0.14	0.15	0.14	0.17	0.12	0.19
	0	0.26	0.14	0.23	0.14	0.18	0.11	0.24	0.15	0.23	0.16
	3	0.33	0.13	0.29	0.13	0.19	0.10	0.30	0.14	0.30	0.15
	6	0.38	0.12	0.34	0.12	0.20	0.08	0.35	0.12	0.33	0.14
	9	0.41	0.11	0.38	0.11	0.21	0.07	0.37	0.11	0.34	0.13
	12	0.41	0.10	0.40	0.10	0.21	0.05	0.35	0.10	0.33	0.12
Street	-3	0.47	0.15	0.46	0.14	0.29	0.10	0.47	0.14	0.44	0.16
	0	0.50	0.13	0.48	0.12	0.29	0.09	0.49	0.12	0.44	0.14
	3	0.50	0.11	0.49	0.11	0.28	0.07	0.47	0.11	0.42	0.12
	6	0.47	0.10	0.47	0.09	0.27	0.05	0.41	0.10	0.39	0.11
	9	0.43	0.09	0.44	0.08	0.27	0.04	0.32	0.10	0.37	0.09
	12	0.37	0.09	0.40	0.09	0.26	0.04	0.20	0.10	0.33	0.09
Station	-3	0.52	0.14	0.50	0.13	0.31	0.12	0.50	0.14	0.48	0.16
	0	0.55	0.13	0.53	0.12	0.32	0.11	0.53	0.13	0.47	0.15
	3	0.55	0.12	0.54	0.11	0.32	0.09	0.52	0.12	0.44	0.14
	6	0.52	0.11	0.52	0.10	0.32	0.08	0.46	0.11	0.39	0.13
	9	0.47	0.11	0.48	0.09	0.30	0.06	0.37	0.11	0.36	0.11
	12	0.41	0.10	0.43	0.09	0.29	0.04	0.25	0.11	0.33	0.10
Restaurant	-3	0.10	0.18	0.10	0.16	0.10	0.12	0.06	0.18	0.02	0.20
	0	0.22	0.15	0.21	0.14	0.16	0.10	0.19	0.15	0.14	0.16
	3	0.32	0.12	0.30	0.12	0.22	0.09	0.29	0.12	0.25	0.14
	6	0.37	0.11	0.35	0.11	0.24	0.07	0.34	0.11	0.30	0.13
	9	0.40	0.10	0.39	0.09	0.25	0.06	0.35	0.10	0.33	0.11
	12	0.40	0.09	0.40	0.08	0.25	0.04	0.33	0.09	0.34	0.10
Lobby	-3	0.27	0.15	0.24	0.14	0.18	0.12	0.25	0.15	0.24	0.19
	0	0.33	0.13	0.30	0.12	0.21	0.10	0.32	0.13	0.30	0.16
	3	0.37	0.12	0.34	0.11	0.21	0.09	0.35	0.12	0.32	0.14
	6	0.38	0.11	0.36	0.11	0.20	0.08	0.36	0.11	0.31	0.13
	9	0.37	0.11	0.37	0.10	0.20	0.06	0.33	0.11	0.29	0.13
	12	0.36	0.10	0.37	0.10	0.20	0.05	0.28	0.11	0.26	0.12
Exhibition	-3	0.20	0.17	0.19	0.16	0.14	0.15	0.18	0.17	0.17	0.19
	0	0.30	0.15	0.27	0.14	0.18	0.12	0.28	0.15	0.27	0.17
	3	0.36	0.14	0.33	0.14	0.20	0.11	0.34	0.14	0.32	0.15
	6	0.39	0.13	0.37	0.13	0.21	0.09	0.37	0.13	0.33	0.15
	9	0.40	0.12	0.39	0.12	0.22	0.08	0.37	0.12	0.32	0.14
	12	0.39	0.12	0.39	0.11	0.22	0.06	0.33	0.11	0.30	0.13
Triplet	-3	0.28	0.15	0.24	0.14	0.17	0.12	0.22	0.15	0.27	0.16
	0	0.38	0.13	0.34	0.12	0.22	0.10	0.32	0.14	0.37	0.15
	3	0.45	0.12	0.41	0.12	0.25	0.09	0.39	0.12	0.41	0.13
	6	0.41	0.12	0.40	0.11	0.21	0.06	0.37	0.12	0.35	0.13
	9	0.41	0.11	0.41	0.11	0.21	0.06	0.35	0.11	0.33	0.13
	12	0.38	0.10	0.40	0.10	0.21	0.04	0.30	0.11	0.31	0.12

the highest improvement in PESQ scores under most of the processing conditions. The plots of mean improvements in the PESQ scores vs SNR are shown in Figure 4.19. It can be seen that considering all noise and SNR combinations, the improvements are highest for ADQTNE, and the other techniques can be ranked as DQTNE, MCRA2, UMMSE, and MS. Thus, the ranking of the techniques based on improvement in the scores and SNR advantage both show ADQTNE as the best technique, closely followed by DQTNE.

4.6 Implementation for Real-Time Processing and Test Results

4.6.1 Implementation for real-time processing

The DQTNE technique with generalized spectral subtraction [66] is implemented for real-time processing on a low-power DSP chip. The 16-bit fixed-point processor TI/TMS320C5515 [98] is selected for this purpose. It has several features, including DMA-based I/O and on-chip hardware for 8 to 1024-point FFT, making it particularly suited for implementing the denoising technique for real-time processing. It has a maximum clock rate of 120 MHz. The implementation was carried out using DSP board 'eZdsp' [99] with codec TLV320AIC3204 [100] supporting 16/20/24/32-bit stereo ADC and DAC with sampling frequency of 8–192 kHz. The implementation uses one channel of the codec, with 16-bit quantization and 10 kHz sampling. TI's 'CCStudio, ver. 4.0' was used as the development environment for programming in C. The DSP chip and the DSP board are the same as used for real-time processing of the sliding-band compression technique presented in the third chapter.

Figure 4.20 shows a block diagram of the implementation. It has two main blocks (marked by dotted outlines). The audio codec has an ADC and a DAC. The digital signal processor comprises the input/output (I/O) and data buffering block based on direct memory access (DMA) and the processing block for speech enhancement using noise estimation and noise suppression. The analog input signal is converted into digital samples by the ADC of the audio codec at the selected sampling frequency. The digital samples are buffered by the I/O block and applied as the input to the processing block. The processed output samples from the processing block are buffered by the I/O and data-buffering block and are applied as the input to DAC of the audio codec, which generates the analog output signal. The processing block uses DQTNE for noise estimation. Noise suppression is carried out using generalized spectral subtraction [66], [115] with magnitude subtraction, subtraction factor of 2, and noise floor factor of 0.001. The processing steps are implemented with due care to avoid overflows.

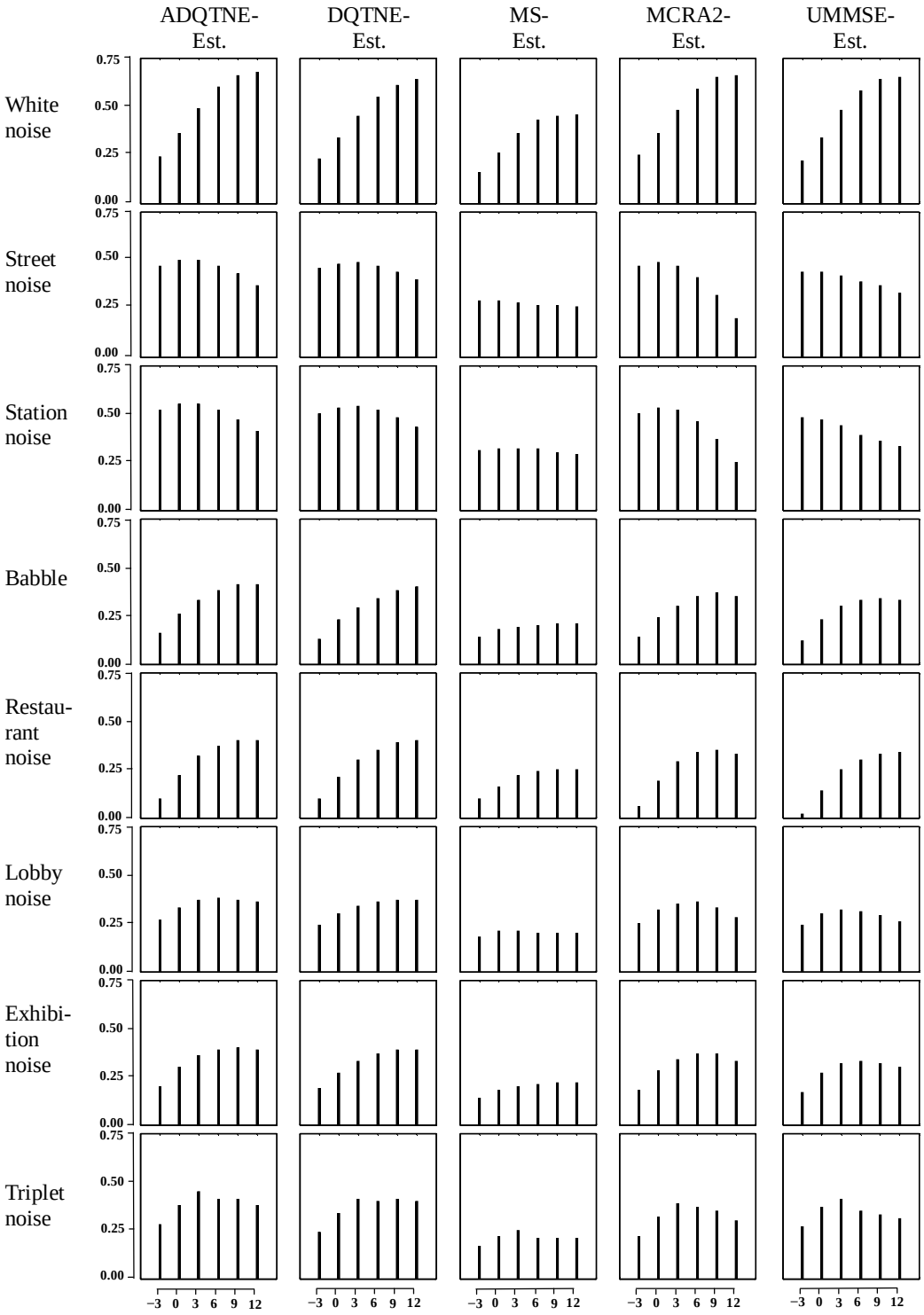


Figure 4.19 Improvements in PESQ scores (0−0.75, on y-axis) vs SNR (−3, 0, 3, 6, 9, 12 dB, on x-axis) for speech enhanced using ADQTNE, DQTNE, MS [71], MCRA2 [81], and UMMSE [85].

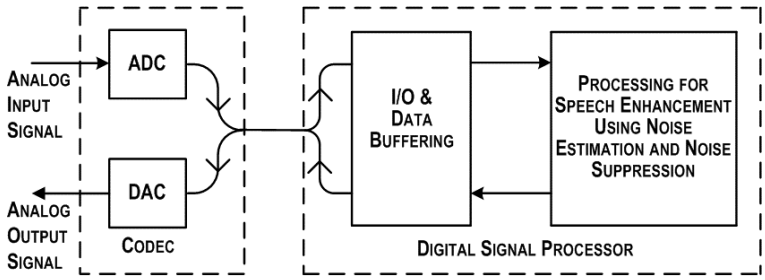


Figure 4.20 Implementation of speech enhancement on the DSP board.

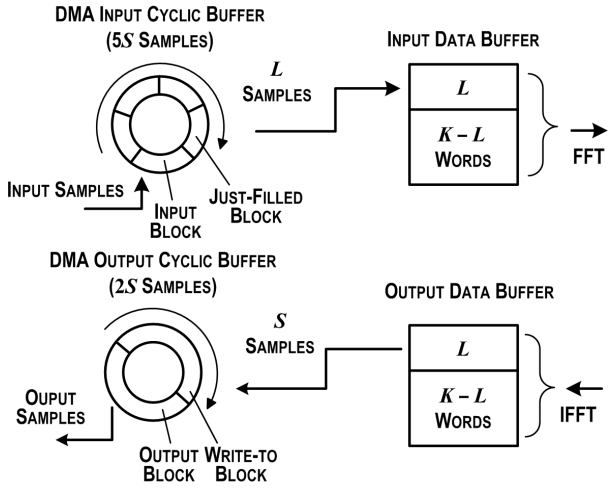


Figure 4.21 Data transfer and buffering on the DSP board ($S = L/4$).

Figure 4.21 shows the input, output, data transfer, and buffering operations devised for an efficient realization of the processing with 75% overlap and zero padding. It uses L-sample analysis window and K-point FFT ($L = 256$, $K = 512$). The input digital samples are read in using a 5-block DMA input cyclic buffer and the processed samples are written out using a 2-block DMA output cyclic buffer, with S-word blocks and with S as $L/4$. Cyclic pointers are used to keep a track of the current input block, just-filled input block, current output block, and write-to output block. The pointers are initialized to 0, 4, 0, and 1, respectively and are incremented at every DMA interrupt generated when a block gets filled. The DMA-mediated reading of the input digital samples into the current input block and writing of the output digital samples from the current output block are continued. Input window with L samples is formed using the samples of the just-filled block and the previous three blocks. These L samples are windowed with a window of length L and are copied to the input data buffer. These samples padded with $K–L$ zero-valued samples serve as input for processing. The

spectral samples obtained from the processing are stored in the output data buffer. The *S* samples are copied in the write-to block of the 2-block DMA output cyclic buffer.

4.6.2 Test results for real-time speech enhancement

The experimental set-up comprised the DSP board and two notebook PCs with sound cards. The speech signal with added noise was output from the sound card of a PC and applied as input to the codec of the DSP board. The output from codec of the DSP board was acquired through the sound card of the other PC. Two machines were used to reduce the noise caused by ground loops. The real-time implementation was evaluated using informal listening, visual examination of spectrograms, and objective evaluation using PESQ. The evaluation was carried out using speech mixed with white, babble, car, street, and train noises at different SNRs. Informal listening showed that the output of the real-time processing was perceptually similar to the corresponding output of offline processing.

Figure 4.22 shows an example of processing showing the noise-free speech, noisy speech with white noise at SNR of 3 dB, output from offline processing, and output from real-time processing. The spectrograms of the enhanced speech outputs from the two types of processing show a close match. The match between outputs from the real-time processing and the output from offline processing was also confirmed by high PESQ scores (greater than 3.5) for real-time processing with offline processing as the reference, indicating that the processing artifacts due to fixed-point processing were not significant.

The audio latency was measured as described in Section 3.4.2 in Chapter 3. It was found to be approximately 36 ms and may be considered as acceptable for use of the processing in the hearing aids along with lipreading.

An empirical estimation of the processor capacity used for implementing the proposed denoising technique was carried out as described in Section 3.4.2 in Chapter 3. For comparison, the processing was also implemented without noise estimation (zero-valued spectral samples for the estimated noise and the code for noise estimation bypassed). The minimum clock frequencies needed for processing with bypassing of noise estimation and DQTNE were 38 and 50 MHz, respectively, indicating a requirement of approximately 32% and 41% of the processor capacity. Thus, the results show that the proposed DQTNE technique can be used for real-time speech enhancement. As the proposed processing needs only 41% of the available capacity, the rest can be used in implementing other processing as needed for a hearing aid. Implementation of DQTNE and ADQTNE with GA-based noise suppression as described in Section 4.5 was found to be not feasible using this processor.

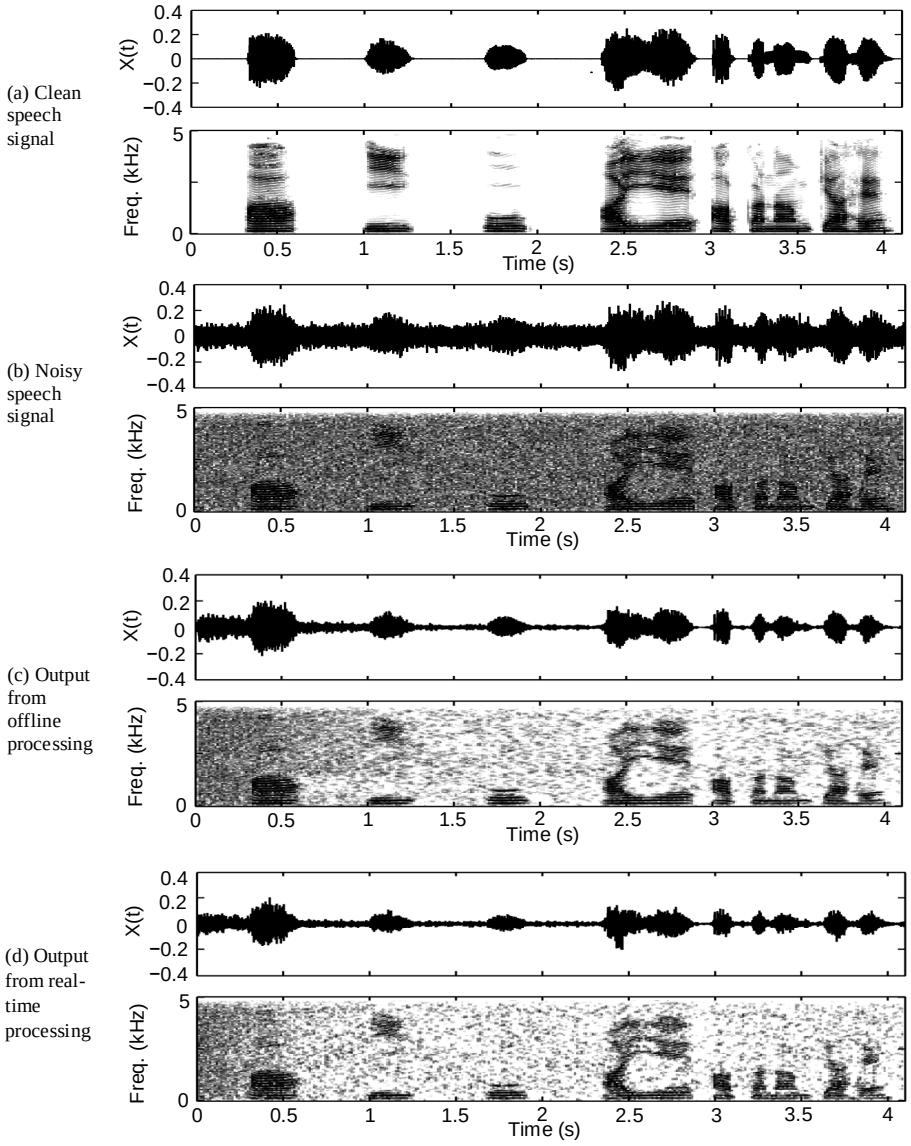


Figure 4.22 Processing of the sequence ("-/a/-/i/-/u/– "aayiye aap kaa naam kyaa hai?" – "Where were you a year ago?"", from a male speaker) with white noise at SNR of 3 dB: signals and spectrograms.

These techniques have been implemented and tested for satisfactory real-time processing as part of a smartphone app [102] as described in Appendix C.

4.7 Discussion

In the preceding sections, we have presented investigations for developing (i) the technique DQTRE for dynamic tracking of quantiles of a data stream without prior knowledge of the distribution of the data and without storage and sorting of the past samples; (ii) the technique DQTNE, using the dynamic quantile tracking technique, for approximately tracking the

quantiles of the spectral samples of the noisy speech spectrum; (iii) the technique ADQTNE using adaptive quantiles for improved estimation of nonstationary noises; (iv) single-input speech enhancement using the quantile-based noise estimation and noise suppression using geometric approach to spectral subtraction; and (v) real-time implementation of the speech enhancement technique.

The technique DQTRE (dynamic quantile tracking using range estimation) has been developed for approximately tracking a quantile of a data stream, without prior knowledge of the distribution of the data and without storage and sorting of the past samples. In this technique, the quantile is estimated recursively by applying an increment, calculated as a fraction of the range, such that the estimated quantile converges to the sample quantile. The range is dynamically estimated using first-order recursive relations for peak and valley detection. The proposed technique provides a trade-off between variance and adaptivity of the estimation. It is suitable for sample-by-sample or window-based tracking of quantiles of non-stationary data. Its memory and computational requirements are independent of the number of possible data values. The technique was tested using synthetic and real data with different distributions and compared with some of the techniques having similar features and reported earlier. As compared to the technique with fast adaptivity (EWSA [108]), DQTRE showed much lower variance during stationary segments and an acceptable adaptivity during transitions. It has significantly lower memory and computational requirements as compared with the low-variance techniques (P2 [109], SSA [110]). Due to its low memory and computational requirements, DQTRE may be considered as suitable for real-time quantile tracking of multiple variables using a DSP chip.

The technique DQTNE (dynamic quantile tracking based noise estimation) uses DQTRE for tracking the quantiles of the spectral samples of the noisy speech spectrum for noise spectrum estimation. It has a very low memory requirement and computational complexity, as compared with earlier quantile-based noise estimation techniques. It permits the use of a different quantile for each sub-band without processing overheads. The technique was compared with the MS [71], MCRA2 [81], and UMMSE [85] techniques, which have an acceptable computational complexity for real-time processing. Its application for speech enhancement and evaluation by informal listening showed that DQTNE had lower speech distortion in the output than MCRA2 and UMMSE. It resulted in less residual noise than MS but more than MCRA2 and UMMSE. The technique ADQTNE (adaptive dynamic quantile tracking based noise estimation) was developed for improving the noise tracking for speech degraded with nonstationary noise and variable SNR. In this technique, the quantile function for each spectral sample is estimated by dynamically tracking multiple quantiles, using an

adaptive convergence factor based on an estimate of the speech presence probability. The quantile where the quantile function has the lowest slope is used as the adaptive quantile representing the noise for each spectral sample. The computational complexity of this technique is increased due to the use of an adaptive quantile that requires tracking of multiple quantiles. In terms of low computational complexity, the techniques are ranked as DQTNE, MCRA2, ADQTNE, MS, and UMMSE, with DQTNE having significantly lower computational complexity than the other techniques.

Evaluation of the noise tracking by the different noise estimation techniques was carried out using sentences from GRID database degraded with noises from the NOISEX and AURORA databases at SNRs of 12, 9, 6, 3, 0, –3, and –6 dB as the test material. The visual examination of spectrograms indicated that ADQTNE resulted in lower overestimation and underestimation than other techniques, across the different noise and SNR combinations. For quantification of the errors in noise estimation, SREE (segmental relative estimation error) was used as an objective measure. Considering error measures for all SNRs and different types of noises, the techniques may be ranked as ADQTNE, DQTNE, MCRA2, MS, and UMMSE.

The proposed noise estimation techniques were used in combination with the GA-based spectral subtraction [90] for enhancement of speech degraded by background noise and evaluated using informal listening, visual inspection of the spectrograms, and objective evaluation using PESQ scores. Informal listening showed that none of the processing techniques resulted in noticeable roughness or musical noise. Considering residual noise and speech attenuation together, ADQTNE provided the highest quality output. The output of DQTNE was similar to that of ADQTNE, except for having a higher residual noise at low SNRs. Visual examination of the spectrograms of the enhanced outputs indicated that ADQTNE and DQTNE were able to track nonstationary noises and provided speech enhancement without noticeable speech attenuation. Both listening and spectrograms showed that MS resulted in significant residual noise and UMMSE resulted in speech attenuation.

In terms of PESQ scores for the enhanced speech across the noises and SNRs, ADQTNE had the highest scores and MS had the lowest scores. The DQTNE scores were lower than the ADQTNE scores for SNR below 6 dB and similar for higher SNRs. Considering the increase in PESQ scores after processing, the SNR advantages using ADQTNE, DQTNE, MCRA2, UMMSE, and MS ranged 4−11, 3−10, 3−10, 3−10, and 2−9 dB, respectively, across different noises. The mean improvements in PESQ scores using ADQTNE, DQTNE, MS, MCRA2, and UMMSE ranged 0.10−0.68, 0.10−0.64, 0.10−0.46, 0.06−0.66, 0.02−0.65, with ADQTNE showing the highest improvement in PESQ scores under most of the processing conditions.

Thus, the results indicated the suitability of ADQTNE for use in presence of stationary as well as nonstationary noises. DQTNE provided an acceptable noise estimation and with a very low computational complexity.

For real-time processing, the computation for each frame of the FFT-based analysis-synthesis should be completed within a window shift and the audio latency (sum of the algorithmic delay and the delay in the input and output buffering operations and filters) should be significantly lower than 120 ms to be acceptable for face-to-face conversation [180], as discussed earlier in Section 3.5. The speech enhancement using DQTNE was implemented on the 16-bit fixed-point processor TI/TMS320C5515 as used earlier for the dynamic range compression. The processing needed approximately 41% of the processing capacity and the audio latency was found to be approximately 36 ms. Implementation of speech enhancement using ADQTNE was found to be not feasible on this DSP chip.

Hearing aids are generally designed using ASICs (application specific integrated circuits) due to power and size constraints. Therefore, incorporation of a new processing technique in hearing aids and its field evaluation is prohibitively expensive. Use of smartphone-based application software (app) to implement a new processing technique permits its use and evaluation by a large number of users without incurring the expenses involved in the ASIC-based hearing aid development. The sliding-band dynamic range compression technique presented in Chapter 3 was implemented for real-time processing as a smartphone app using the Android-based handset Nexus 5X and it was found that the processing used only a small fraction of the processing capacity and acceptable audio latency [101]. Therefore, the feasibility of using ADQTNE for real-time speech enhancement was assessed by implementing it along with the dynamic range compression as a smartphone app for the same handset. This app, described in Appendix C, enables the setting of the processing parameters in an interactive and real-time mode using a graphical touch interface and makes the proposed processing techniques conveniently available to the hearing-impaired user. The audio latency of the app tested using the handset Nexus 5X was found to be 45 ms. For the frame length of 20 ms with 75% overlap, the algorithmic delay was 25 ms (1.25 times the frame length). The additional delay was due to audio input-output latency of the handset hardware, buffering operations in the OS, and delays in the anti-aliasing and smoothening filters. The implementation used less than 50% of the processor capacity. Implementation of the app on other smartphones and its use by a large number of hearing-impaired listeners is needed for real-life evaluation and further enhancement.

Left blank

Chapter 5

SUMMARY AND CONCLUSION

5.1 Introduction

The objective of the research was to develop signal-processing techniques for dynamic range compression and background noise suppression to enhance the performance of hearing aids used by listeners with sensorineural loss. In order to overcome the drawbacks of the existing techniques, investigations were carried out to develop (i) sliding-band dynamic range compression to compensate for frequency-dependent loudness recruitment and (ii) speech enhancement using dynamic quantile tracking for estimation of background noise. The first technique was developed with an aim to avoid distortions associated with the commonly employed single-band and multiband compression techniques. The second technique was developed for single-input speech enhancement by suppression of background noise, without voice activity detection for estimation of the noise spectrum. The techniques were developed with considerations for low memory and computational requirement for implementation in hearing aids, low audio latency for face-to-face communication, and low perceptible distortions. Implementations of the proposed techniques for offline processing were used for their evaluation and comparison with some of the existing techniques. Subsequently, the proposed techniques were implemented individually for real-time processing using a 16-bit fixed-point DSP chip. To enable the use and evaluation of these techniques by a large number of users without incurring the expenses involved in the ASIC-based hearing aid development, a smartphone app implementing the two techniques with interactive touch-controlled graphical user interface was also developed. The examples of the offline and real-time processing are available at [181]. The summary of the investigations, conclusions, and suggestions for further research are presented in the following sections.

5.2 Summary of the Investigations

The research reported in this thesis can be summarized as the following.

1) Development of the sliding-band dynamic range compression technique

A dynamic range compression technique, named as ‘sliding-band compression’ (SLBC), was developed to compensate for frequency-dependent loudness recruitment, without introducing the distortions generally associated with the single-band and multiband compressions. In this technique, the gain for each spectral sample is based on the short-time power in a band

centered at its frequency. The technique avoids the attenuation of high-frequency components due to the presence of strong low-frequency components, which may occur in single-band compression. Further, it avoids distortions in the shape of spectral resonances and discontinuities during the resonance transitions, which may occur in multiband compression. It can be used with settable attack and release times and compression functions in accordance with the selected hearing-aid fitting procedure. The technique and results of its evaluation are presented in Section 3.2 and Section 3.3, respectively, of the third chapter.

The proposed technique was implemented for offline processing and evaluated by comparing it with single-band compression and multiband compression techniques by examination of the spectrograms for undesirable level changes in the outputs for single-tone, two-tone, and speech inputs and by quantifying the deviations from the expected output levels for single-tone and two-tone inputs. Both evaluations showed that the sliding-band compression did not result in the distortions associated with the single-band or multiband compressions.

2) Implementation of the sliding-band compression for real-time processing using a fixed-point DSP chip

The proposed sliding-band compression involves level estimation and gain calculation for each spectral sample and thus has a higher computational requirement than the multiband compression. To assess its suitability for use in hearing aids, the technique was implemented for real-time processing using a 16-bit fixed-point DSP chip (TI/TMS320C5515). The implementation and test results are presented in Section 3.4 of the third chapter. The processed output from the real-time processing had a close match with that of the offline processing. The processing requires approximately 41% of the processing capacity of the chip and has an audio latency of approximately 36 ms.

3) Development of a technique for the dynamic tracking of quantiles of a data stream

A computationally efficient dynamic quantile tracking technique based on stochastic approximation was proposed for estimating the quantiles of a data stream. It does not require prior knowledge of the distribution of the data and does not involve storage and sorting of the past samples. In this technique, the quantile is estimated recursively by applying an increment, calculated as a fraction of the dynamically estimated range, such that the estimated quantile converges to the sample quantile. The technique is presented in Section 4.2 of the fourth chapter. A detailed description of the technique and the test results are provided in Appendix B. Evaluation, using synthetic and real data with different distributions, and comparison with the existing techniques showed the proposed technique to have lower

variance than the techniques with good adaptivity and much lower memory and computational requirements than the techniques with low variance. The results indicate its suitability for real-time quantile tracking of multiple variables using a DSP chip and hence it can be used for quantile-based noise estimation for real-time speech enhancement.

4) Development of the techniques for quantile based noise estimation

The dynamic quantile tracking technique was applied for the tracking of the quantiles of the spectral samples of the noisy speech spectrum for noise spectrum estimation without voice activity detection, resulting in the technique named as 'dynamic quantile tracking based noise estimation' (DQTNE). It permits use of a different fixed quantile for each sub-band without processing overheads. For improving the tracking of nonstationary noises, the technique named as 'adaptive dynamic quantile tracking based noise estimation' (ADQTNE) was developed. It uses an adaptive quantile and an adaptive convergence factor. It involves estimating a quantile function for each sub-band by dynamically tracking multiple quantiles and an estimate of the speech presence probability. The two techniques and results of their evaluation are presented in Section 4.3 and Section 4.4, respectively, of the fourth chapter.

The two proposed techniques were evaluated and compared with three earlier reported techniques (MS [71], MCRA2 [81], and UMMSE [85]). In terms of low computational complexity, the techniques are ranked as DQTNE, MCRA2, ADQTNE, MS, and UMMSE, with DQTNE having significantly lower computational complexity than the other techniques. Considering low error in noise tracking for different stationary and nonstationary noises, the techniques may be ranked as ADQTNE, DQTNE, MCRA2, MS, and UMMSE.

5) Evaluation of the proposed noise estimation techniques in a speech enhancement framework

The two proposed noise estimation techniques (DQTNE and ADQTNE) and three earlier reported techniques (MS, MCRA2, and UMMSE) were used in combination with the spectral subtraction based on the geometric approach [90] for enhancement of speech degraded by background noise. The evaluation was carried out, for input speech material degraded by different stationary and nonstationary noises and SNRs, using informal listening, examination of the spectrograms, and objective evaluation using PESQ scores. The implementation and the test results are presented in Section 4.5 of the fourth chapter.

None of the processing techniques resulted in noticeable roughness or musical noise in the processed outputs. The three evaluations resulted in similar ranking of the techniques, with ADQTNE generally providing better quality output than the other techniques. It was observed that MS generally resulted in residual noise, DQTNE occasionally resulted in residual noise,

UMMSE generally resulted in speech attenuation, and MCRA2 occasionally resulted in speech attenuation. Considering residual noise and speech attenuation together, ADQTNE provided the highest quality output. The output of DQTNE was similar to that of ADQTNE, except for having a higher residual noise at low SNRs and particularly for nonstationary noises. Considering the increase in PESQ scores after processing, the SNR advantages using ADQTNE, DQTNE, MCRA2, UMMSE, and MS ranged 4−11, 3−10, 3−10, 3−10, and 2−9 dB, respectively, across different noises. The mean improvements in PESQ scores using ADQTNE, DQTNE, MS, MCRA2, and UMMSE ranged 0.10−0.68, 0.10−0.64, 0.10−0.46, 0.06−0.66, 0.02−0.65, with ADQTNE showing the highest improvement in PESQ scores under most of the processing conditions.

6) Implementation of speech enhancement using DQTNE for real-time processing

The speech enhancement using DQTNE was implemented using a 16-bit fixed-point DSP chip (TI/TMS320C5515), as for the dynamic range compression. The implementation and the test results are presented in Section 4.6 of the fourth chapter. There were no noticeable differences between outputs from the offline processing and the real-time processing. The computation requirement and the audio latency were approximately the same as for the compression technique. The technique ADQTNE, with its computational complexity being significantly larger than that of DQTNE, could not be implemented on this chip.

7) Implementation of ADQTNE and SLBC as a smartphone app

A smartphone app with the signal processing for single-input speech enhancement using ADQTNE for noise estimation and SLBC for dynamic range compression was developed. It was tested using the Android-based handset Nexus 5X, which was selected for its low input-output delay and the capacity of its processor for running the audio apps. The implementation and test results are presented in Appendix C. The app has facility for setting the processing parameters in an interactive and real-time mode using a graphical touch interface. It was found that the app used less than 50% of the processor capacity and the audio latency was 45 ms.

5.3 Conclusions

A hearing aid is used to overcome the deficits associated with hearing loss. It employs frequency-selective amplification to improve the sound audibility and signal processing for dynamic range compression and noise reduction to improve the comfort and the speech intelligibility. Investigations were carried out to develop the signal processing techniques to improve the usefulness of the hearing aids for persons with sensorineural and mixed losses.

The first technique has been developed with an aim to avoid distortions associated with the commonly employed single-band and multiband compression techniques. The second technique has been developed for single-input speech enhancement by suppression of background noise, without voice activity detection for estimation of the noise spectrum. The techniques have been developed with considerations for low memory and computational requirements for implementation in hearing aids, low audio latency for face-to-face communication, and low perceptible distortions. For evaluation and comparison of the performances, the proposed techniques and some of the corresponding existing techniques have been implemented for offline processing. Evaluations using informal listening, examination of spectrograms, and objective measures showed the proposed techniques to provide better performances than the corresponding existing techniques.

Suitability of the proposed techniques for use in hearing aids has been verified by their implementation for real-time processing using a 16-bit fixed-point DSP chip. To enable the use and evaluation of these techniques by a large number of users, without incurring the expenses involved in the ASIC-based hearing aid development, a smartphone app implementing the two techniques with interactive touch-controlled graphical user interface has been developed.

The main conclusions from the research can be summarized as the following:

1) The technique developed for dynamic range compression and named as ‘sliding-band compression’ (SLBC) can be used to compensate for frequency-dependent loudness recruitment caused by sensorineural hearing loss without introducing the distortions generally associated with the single-band and multiband compressions. Although its computational requirement is higher than that of the multiband compression, it is suitable for implementation using currently available processors and with an acceptable audio latency.

2) The technique developed for quantile-based noise estimation for speech enhancement and named as ‘dynamic quantile tracking based noise estimation’ (DQTNE) can be used for an acceptable noise estimation and with a very low computational complexity. The technique named as ‘adaptive quantile tracking based noise estimation’ (ADQTNE) can be used for a better performance, particularly in case of nonstationary noise and varying SNR. The second technique has a much higher computational complexity and is not implementatble using currently available DSP chips. It can be implemented on the processors of currently available high-end smartphones and with an acceptable audio latency.

The two noise estimation techniques developed for single-input speech enhancement are based on a computationally efficient technique developed for tracking of quantiles of a data stream and named as ‘dynamic quantile tracking using range estimation’ (DQTRE).

Evaluation using synthetic and real data with different distributions has shown this technique to provide tracking with a low variance and high adaptivity. In addition to use in speech enhancement, the proposed quantile tracking technique may be suitable for use of order statistics in several signal processing and control applications.

5.4 Suggestions for Further Research

The signal processing for background noise suppression and dynamic range compression uses an analysis-synthesis method that modifies the spectral magnitude and retains the original phase. Discontinuities in the magnitude-phase relationship introduced by this method may partly offset the advantages of the processing. Use of the spectral phase reconstruction techniques, without a significant increase in the computational requirements of the processing, needs to be investigated. The noise estimation techniques have been applied for speech enhancement using spectral subtraction based on the geometric approach [90]. Speech enhancement using other noise suppression techniques needs to be investigated.

For evaluation of the techniques on the hearing-impaired subjects, the techniques for noise suppression and dynamic range compression need to be implemented as part of the processing in a hearing aid. The method and material for the proposed experiments should be in accordance with those reported in earlier studies for evaluation of compression and noise suppression. For evaluation of compression, nonsense CVC syllables, sentences from the Connected Speech Test [44], and speech material with large word-to-word level variations may be used as the test material in quiet and in presence of babble and speech-shaped noise. The listening tests should involve subjects with significant loudness recruitment. As the relationship between hearing loss and loudness recruitment varies from person to person, the tests should be conducted on a large number of listeners with moderate-to-severe sensorineural impairement. For evaluation of the noise suppression, the listening test for speech quality and intelligibility should be carried out with the listeners with normal hearing and those with moderate-to-severe sensorineural impairement. For these tests, sentences from IEEE database, TIMIT [79], or GRID database [111] with phonetically-balanced sentences having relatively low word-context predictability may be used along with noises from the NOISEX [74] and AURORA [93] databases. Subjective evaluation should involve a minimum of 10 normal hearing listeners and a large number of hearing impaired listeners with moderate-to-severe (flat and sloping) hearing loss.

Due to the power and size constraints, the hearing aids are generally based on ASICs, leading to prohibitive costs in development and testing of new processing techniques. Use of a smartphone-based app as a hearing aid is suggested as a low-cost alternative to hearing aids.

It can be used to provide user-configurable settings and a greater flexibility to the hearing aid users and developers. The smartphone app was developed, as part of our investigations, to assess the suitability of the proposed techniques for real-time processing. Feasibility of use of the app on commonly used smartphone handsets and availability of headsets with appropriate output levels need to be examined. The app needs to be revised, with inputs from audiologists and the users, for use by a large number of hearing-impaired listeners. Guidelines for setting the processing parameters of the app need to be developed and a study needs to be undertaken for its clinical evaluation and further enhancements.

Left blank

Appendix A

HEARING AID FITTING PROCEDURES

A.1 Introduction

Hearing aids process and present the input signal with frequency-dependent amplification and dynamic range compression to compensate for elevated hearing thresholds and reduced dynamic range associated with hearing loss. Fitting of a hearing aid involves selecting the amplification and compression characteristics most appropriate for the loss characteristics of the user's ear. It may be carried out using either a comparative or a prescriptive approach [5], [116]−[124]. In a comparative approach [117], [122], several hearing aids with different characteristics are compared by administering speech tests on the hearing aid user to select the hearing aid with the best performance. This approach may require testing an impractically large number of devices and the test material may affect the performance leading to low test-retest reliability. These problems are avoided in a prescriptive approach, which is more commonly used. Several prescriptive procedures [116], [119], [123], [124] have been developed for selection of the hearing aid characteristics, and they may be grouped according to (i) type of the audiometric data used for fitting, (ii) type of the amplification characteristics prescribed, and (iii) aim of the fitting procedure.

Based on the type of audiometric data used for fitting, the prescriptive procedures may be grouped into those based on threshold data, such as air conduction thresholds (AC) and bone conduction thresholds (BC), and those using supra-threshold loudness judgements, such as most comfortable level (MCL) and uncomfortable level (UCL). The prescriptive procedures that are based on the hearing thresholds include NAL [125], NAL-R [126], NAL-RP [127], CAM2 [128], POGO [129], POGO II [130], FIG6 [131], CAMEQ [132], NAL-NL1 [133], and NAL-NL2 [134]. The prescriptive procedures using the supra-threshold data, such as MCL, UCL, or full loudness scale, include Shapiro [135], LGOB [136], IHAFF [137], and DSL[i/o] [138].

The procedures may be grouped on the basis of the type of prescribed amplification characteristics into those for linear amplification and those for compression amplification. Procedures for linear amplification prescribe a fixed gain along with a maximum output level and they are suitable for conductive loss. The procedures for compression amplification prescribe different gains for different input levels and they are suitable for reduced dynamic range and loudness recruitment associated with sensorineural loss. Fitting procedures for

linear amplification include POGO, NAL, and DSL, whereas those for compression amplification include LGOB, IHAFF, FIG6, DSL[i/o], and DSLm[i/o] [139].

Based on the underlying theoretical rationale, the fitting procedures for compression amplification may be grouped into those for loudness normalization, loudness equalization, and intelligibility maximization. The loudness normalization procedures, such as LGOB, IHAFF, FIG6, DSL[i/o], and DSLm[i/o], prescribe the gains to make the loudness perceived by the hearing aid user to be same as that by a normal-hearing listener, maintaining the normal loudness relations between different frequency bands. Loudness equalization procedures, such as CAMEQ and CAM2, prescribe the gains to equalize loudness of the frequency bands in the averaged speech spectrum and to match the overall loudness to that perceived by a normal-hearing listener. The intelligibility maximization procedures, such as NAL-NL1 and NAL-NL2, aim to maximize speech intelligibility and to match the overall loudness to that perceived by a normal-hearing listener.

A review of some of the prescriptive procedures for linear amplification and compression amplification is given in the following sections.

A.2 POGO and POGO II Procedures

Knudsen and Jones [124] proposed a procedure for linear amplification that is based on mirroring the audiogram and it compensates each decibel of loss by a decibel of gain. For a listener with sensorineural loss, this procedure may provide an excessive gain at higher input levels and thus may cause uncomfortable loudness and speech distortion. Lybarger [123] reported that the appropriate gain for conversational speech to be audible and comfortable was approximately half of the loss and proposed the ‘half-gain rule’ in which each dB of loss at a frequency is compensated by 0.5 dB of gain. The procedure uses hearing threshold levels (HTL) as the input audiometric data to prescribe the frequency-dependent gain in dB as

$$[G(f)]_{\text{Lybarger}} = 0.5H(f) \qquad \text{(A.1)}$$

where f is the frequency and $H(f)$ is HTL in dB HL.

Many prescriptive procedures based on a modification of the half-gain rule have been proposed [125], [127], [129]. McCandless and Lyregaard [129] proposed the ‘prescription of gain and output’ (POGO) formula for linear amplification based on loudness normalization rationale, with a frequency-dependent gain correction factor in the low frequencies for reducing the upward spread of masking from low-frequency ambient noise and to avoid application of higher gains at low frequencies. The POGO formula prescribes the gain in dB for HTL up to 80 dB HL and is given as

$$[G(f)]_{\text{POGO}} = 0.5H(f) + k(f) \quad \text{(A.2)}$$

where $k(f)$ is a frequency-dependent gain correction, which is −10, −5, 0, 0, and 0 dB for 0.25, 0.50, 1, 2, and 4 kHz, respectively.

Schwartz et al. [130] proposed a revised formula, POGO II, for severe-to-profound hearing loss. This formula is same as the POGO formula for HTL up to 65 dB HL and every decibel of higher hearing loss is compensated by 1 dB of gain. The gain in dB is given as

$$[G(f)]_{\text{POGO II}} = \begin{cases} 0.5H(f) + k(f), & H(f) \le 65 \text{ dB HL} \\ 0.5H(f) + k(f) + 0.5(H(f) - 65), & H(f) > 65 \text{ dB HL} \end{cases} \quad \text{(A.3)}$$

A.3 NAL, NAL-R, and NAL-RP Procedures

The NAL procedure is a linear amplification prescriptive procedure from the National Acoustic Laboratory (Australia), proposed by Byrne and Tonnisson [125], with an aim to maximize speech intelligibility at a listening level preferred by the hearing aid user. It is based on the assumption that the intelligibility is maximized if all frequency bands of speech contribute equally to the loudness of the signal. The HTLs serve as the input audiometric data for this procedure. As a variation of the half-gain rule, every dB of loss is compensated by 0.46 dB of gain along with frequency-dependent corrections at 0.25, 0.50, 1, 1.50, 2, 4, and 8 kHz. The corrections comprise two sets of frequency-dependent adjustments. The first set provides adjustment for loudness differences in accordance with 60-phon equal-loudness contour, with corrections of −2, −4, 0, 0, −2, −7, and −8 dB at 0.25, 0.50, 1, 1.50, 2, 4, and 8 kHz, respectively. The corresponding corrections by the second set are −13, −9, 0, −1, 5, 5, and 2 dB and are provided to adjust for the critical-band levels of the long-term averaged power of the speech signal. Thus, the gain in dB is given as

$$[G(f)]_{\text{NAL}} = 0.46\,H(f) + k(f) \quad \text{(A.4)}$$

where the gain corrections $k(f)$ are −15, −13, 0, −1, 3, −2, and −6 dB at 0.25, 0.50, 1, 1.50, 2, 4, and 8 kHz, respectively. Byrne [140] later reported that the frequency response prescribed according to the NAL procedure results in insufficient gains at low frequencies. Byrne and Dillon [126] proposed the ‘NAL revised’ formula (NAL-R), with every dB of hearing loss compensated by about 1/3 dB of gain and the gain in dB given as

$$[G(f)]_{\text{NAL-R}} = 0.31\,H(f) + k(f) + 0.15H_{\text{3FA}} \quad \text{(A.5)}$$

Table A.1 NAL-RP: Profound correction PC, in dB, as a function of frequency and hearing threshold at 2 kHz (H(2 kHz)), adapted from [127].

H(2 kHz) in dB	PC(f)						
	f = 0.25 kHz	f = 0.50 kHz	f = 1 kHz	f = 2 kHz	f = 3 kHz	f = 4 kHz	f = 6 kHz
≤ 90	0	0	0	0	0	0	0
95	4	3	0	−2	−2	−2	−2
100	6	4	0	−3	−3	−3	−3
105	8	5	0	−5	−5	−5	−5
110	11	7	0	−6	−6	−6	−6
115	13	8	0	−8	−8	−8	−8
120	15	9	0	−9	−9	−9	−9

where H_{3FA} is the average of the HTLs at 0.50, 1, and 2 kHz. The formula prescribes gains at 0.25, 0.50, 0.75, 1, 1.50, 2, 3, 4, and 6 kHz, with the corresponding gain corrections $k(f)$ as −17, −8, −3, 1, 1, −1, −2, −2, and −2 dB, respectively. For listeners with a severe-to-profound hearing loss, Byrne et al. [127] proposed a formula known as 'NAL revised, profound' (NAL-RP). The gain in dB is given using a profound correction term as the following:

$$[G(f)]_{\text{NAL-RP}} = \begin{cases} 0.31\,H(f) + k(f) + 0.15H_{3FA} + \text{PC} & H_{3FA} \le 60 \text{ dB HL} \\ 0.31\,H(f) + k(f) + 0.35H_{3FA} - 12 + \text{PC} & H_{3FA} > 60 \text{ dB HL} \end{cases} \quad \text{(A.6)}$$

The profound correction term PC is a function of f and the HTL at 2 kHz, as given in Table A.1.

A.4 DSL Procedure

Seewald et al. [141] proposed a linear amplification prescriptive procedure called 'desired sensation level' (DSL) for fitting of pediatric hearing aids. The procedure is based on loudness normalization rationale and it uses hearing thresholds as the input audiometric data. It specifies the target (or desired) sensation level to make speech comfortably loud. The target sensation level decreases with increase in the hearing threshold, as a higher hearing threshold is generally associated with a reduced dynamic range. In the earlier versions of this procedure, the target sensation levels are specified as a function of frequency and hearing thresholds using tables. In DSL v3.1, these levels are obtained using a software. Unlike other linear prescriptive procedures, this procedure refers hearing thresholds, target sensation levels, and uncomfortable level to dB SPL in the ear canal for easier comparison and removes the age-related variation in the ear canal acoustics from the calculations. At each frequency, the average hearing threshold for normal hearing, the hearing threshold for the hearing-impaired listener, and the speech spectrum level for averaged speech spectrum at 70 dB SPL are used

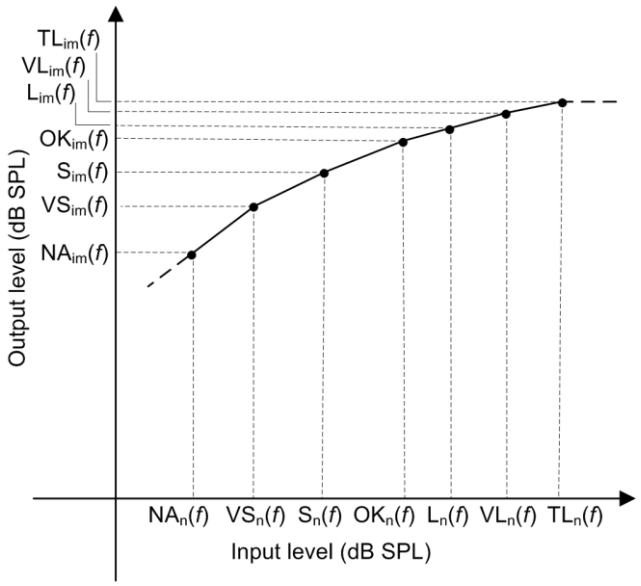


Figure A.1 LGOB procedure: Input-output relation using seven loudness categories with the input level for a normal-hearing listener (n) and the output level for the hearing-impaired listener.

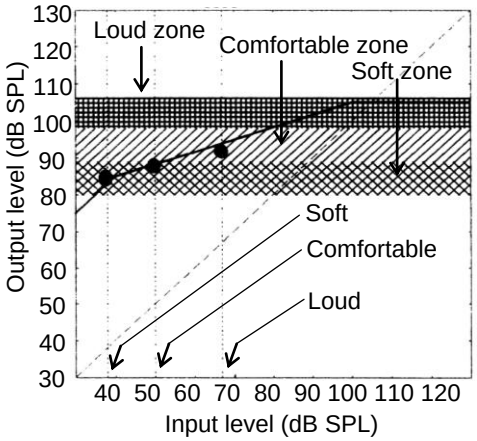
to obtain the target sensation level in dB SPL. The gain in dB at each frequency is calculated by subtracting the one-third octave band levels of speech at an overall level of 70 dB SPL from the target sensation level.

A.5 LGOB Procedure

Allen et al. [136] proposed a compression amplification prescriptive procedure called 'loudness growth in half-octave bands' (LGOB). The procedure is based on loudness normalization rationale and it uses the HTLs and supra-threshold loudness growth as the input audiometric data for obtaining the input-output relation. The loudness growth as a function of frequency and level is estimated by conducting a test using half-octave bands of noise, with center frequencies of 0.25, 0.50, 1, 2, and 4 kHz, as stimuli. The hearing-aid user categorizes the loudness of the stimuli, presented at 15 levels equispaced on a dB scale between the 'not audible' and 'too loud' levels, on a six-point loudness scale as 'very soft', 'soft', 'ok', 'loud', 'very loud', and 'too loud'. A multi-segment relation between the input level for a normal listener and output level for the hearing-impaired listener is obtained for each of categories on the loudness scale, as shown in Figure A.1, with the gain obtained as the difference between the output and input levels.

A.6 IHAFF Procedure

The 'independent hearing aid fitting forum' (IHAFF) protocol, described by Valente et al. [137], is a compression amplification prescriptive procedure, based on loudness normalization



Frequency (Hz)	Points on output vs input curve		
	Soft	Comfortable	Loud
250	0.01(s)	0.50(s)	0.87(s)
500	0.39(s)	0.90(s)	0.48(c)
1000	0.24(s)	0.85(s)	0.67(c)
2000	0.29(s)	0.82(s)	0.64(c)
3000	0.27(s)	0.82(s)	0.59(c)
4000	0.26(s)	0.71(s)	0.50(c)

Figure A.2 IHAFF procedure: An example of input-output curve (at 3 kHz) with two compression thresholds and two compression ratios; Table providing the vertical position of the three points corresponding to soft, average, and loud speech levels, on the output-input curve, as a fraction of soft zone (s) or comfortable zone (c), at different frequencies (adapted from [137]).

rationale. It uses the HTLs and supra-threshold loudness growth as the input audiometric data for obtaining the input-output relation. The loudness growth as a function of frequency and level is estimated by conducting loudness-scaling test, called as the contour test, using warble tones of 0.25, 0.50, 1, 2, 3, and 4 kHz as stimuli. The hearing-aid user categorizes the loudness of the stimuli, presented in 2 dB steps between the hearing threshold and 'uncomfortably loud' levels, on a seven-point loudness scale as 'very soft', 'soft', 'comfortable but slightly soft, 'comfortable, 'comfortable but slightly loud, 'loud but ok', and 'uncomfortably loud'. The responses are compared to the responses of a group of normal-hearing listeners and grouped using a software, based on 'visual input-output locator algorithm' [142], into three zones, as soft zone (very soft, soft, comfortable but slightly soft), comfortable zone (comfortable but slightly soft, comfortable, comfortable but slightly loud), and loud zone (comfortable but slightly loud, loud but ok, uncomfortably loud). An input-output curve with compression thresholds and compression ratios for loudness normalization is obtained at each frequency. An example input-output curve at 3 kHz is shown in Figure A.2. Three points corresponding to soft, average, and loud speech levels are placed on the input-output curve for obtaining the compression thresholds and compression ratios. For each of the three points, the horizontal position is the level of speech in the 1/3-octave bands when the complete speech signal is at soft, comfortable, and loud level for a normal-hearing listener and the vertical position is a fraction of soft zone (s) or comfortable zone (c). The compression ratio is obtained as the ratio of the level difference between the first and third points for the input to that for the output. The compression threshold for the input is selected in the range 40−45 dB SPL and the maximum output level is selected as the upper horizontal

line of the loud zone. The gain for the input up to 40 dB SPL is obtained by subtracting 40 dB from the very-soft output level. The values of gain below 40 dB SPL, compression threshold, compression ratio, and maximum output level are used to plot a compression function, which may be further adjusted to get a compression function closer to the three points.

A.7 FIG6 Procedure

The 'FIG6' procedure, described by Killion and Fikret-Pasa [131], is a compression amplification prescriptive procedure and gets its name from Figure 6 in [131]. It is based on loudness normalization rationale. It uses HTL $H(f)$ as the input audiometric data. In place of using the individual's loudness-growth measures, it is based on the loudness data averaged across a large number of hearing-impaired listeners with similar HTLs. It prescribes the gain for the input at 40, 60, and 95 dB SPL, corresponding to the soft level, MCL, and UCL, respectively, for a normal-hearing person. The prescribed gains at these three input levels are given [5] as the following:

$$[G(f)]_{\text{FIG6, 40 dB SPL}} = \begin{cases} 0, & H(f) < 20 \text{ dB HL} \\ H(f) - 20, & 20 \text{ dB HL} \le H(f) < 60 \text{ dB HL} \\ 0.5H(f) + 10, & H(f) \ge 60 \text{ dB HL} \end{cases} \quad \text{(A.7)}$$

$$[G(f)]_{\text{FIG6, 65 dB SPL}} = \begin{cases} 0, & H(f) < 20 \text{ dB HL} \\ 0.6(H(f) - 20), & 20 \text{ dB HL} \le H(f) < 60 \text{ dB HL} \\ 0.8H(f) - 23, & H(f) \ge 60 \text{ dB HL} \end{cases} \quad \text{(A.8)}$$

$$[G(f)]_{\text{FIG6, 95 dB SPL}} = \begin{cases} 0, & H(f) \le 40 \text{ dB HL} \\ 0.1(H(f) - 40)^{1.4}, & \text{otherwise} \end{cases} \quad \text{(A.9)}$$

The gains for other input levels are interpolated from the above three gains.

A.8 NAL-NL1 and NAL-NL2 Procedures

NAL-NL1 [133] is a compression amplification prescriptive procedure, aimed at maximizing the speech intelligibility and matching the loudness perceived by the hearing aid user to that by a normal-hearing listener. It uses only the HTLs as the input audiometric data. The procedure was developed using 52 audiograms representing different types and degrees of hearing loss, averaged speech spectrum, the loudness model proposed by Moore and Glasberg [143], and a modified form of the Speech Intelligibility Index (SII) [144] as the intelligibility model. For a given audiogram and input speech level, the amplified speech spectrum was input to the loudness and intelligibility models to calculate the loudness and intelligibility index. The gains for 1/3-octave bands were varied to maximize the speech intelligibility and

to match the predicted loudness to that for a normal-hearing person. The procedure was repeated for each of the audiograms and input levels. Curve fitting was used on the resulting gain functions for each of the audiograms and input levels to obtain a prescription formula that was implemented as a computer program for use in hearing aid fitting software.

In the NAL-NL2 procedure [134], the gain at each frequency was obtained in a manner similar to NAL-NL1 procedure, but using a modified form of SII [145] to avoid overestimation of intelligibility with increase in HTL. The procedure was repeated for a set of 240 audiograms covering a wide range of severity and slopes, each at seven speech levels from 40 to 100 dB SPL. Optimum gains for each audiogram and input level combination were obtained. In place of curve fitting, a three-layer neural network was trained using the audiograms and input levels as the inputs and the optimized gains as the target. The trained neural network is incorporated in a software [146] for hearing aid fitting.

A.9 DSL[i/o] and DSLm[i/o] Procedures

Desired sensation level input-output (DSL[i/o]) is a compression amplification prescriptive procedure proposed by Cornelisse et al. [138]. The procedure is based on loudness normalization rationale and it uses hearing thresholds as the input audiometric data. Like the DSL procedure for linear amplification (described in Section A.4), DSL[i/o] refers hearing thresholds, target sensation levels, and UCL to dB SPL in the ear canal. The procedure has two types of compression relations, a linear compression and a curvilinear compression.

For linear compression, the compression function at each frequency is obtained in three stages as linear gain, compression, and output limiting. It provides linear gain for inputs up to the compression threshold. In the compression region, a compression ratio is used to fit the dynamic range of input into the limited dynamic range of the hearing aid user. The output is limited to UCL_{im} in the limiting region. The compression function specifies the output signal level P_{OdB} in the ear canal as a function of the input level P_{IdB} in the sound field, the sound field to ear canal transform FE, the normal-hearing threshold TH_n, the normal-hearing uncomfortable level UCL_n, the hearing threshold of the hearing-impaired listener TH_{im}, and the uncomfortable level for the hearing-impaired listener UCL_{im}. The function is given as

$$P_{OdB}(f)_{\mathrm{DSL[i/o]\_lin}} = \begin{cases} P_{IdB}(f) + \mathrm{TH}_{im}(f) - (\mathrm{TH}_n(f) - \mathrm{FE}(f)), & P_{IdB}(f) < \mathrm{TH}_n(f) - \mathrm{FE}(f) \\ \mathrm{TH}_{im}(f) + [P_{IdB}(f) - (\mathrm{TH}_n(f) - \mathrm{FE}(f))]\dfrac{\mathrm{UCL}_{im}(f) - \mathrm{TH}_{im}(f)}{\mathrm{UCL}_{im}(f) - \mathrm{TH}_n(f)}, & \\ \qquad\qquad \mathrm{TH}_n(f) - \mathrm{FE}(f) \le P_{IdB}(f) < \mathrm{UCL}_{im}(f) - \mathrm{FE}(f) & \\ \mathrm{UCL}_{im}(f), \qquad P_{IdB}(f) \ge \mathrm{UCL}_{im}(f) - \mathrm{FE}(f) & \end{cases} \tag{A.10}$$

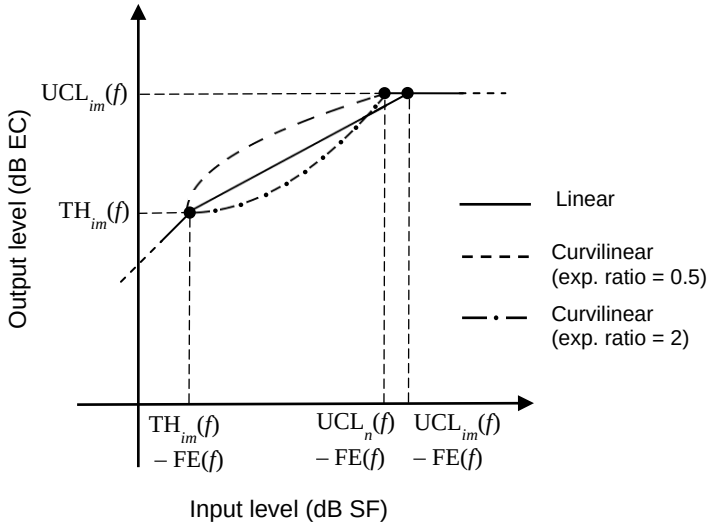


Figure A.3 DSL[i/o] procedure: Relation between input level in the sound field (dB SF) and output level in the listener's ear canal (dB EC) for linear compression and for curvilinear compression with loudness growth exponent ratio of 0.5 and 2.0 (adapted from [138]).

The compression function for the curvilinear compression also has three stages. The linear gain and the output limiting stages are the same as for the linear compression. For the compression stage, the function is curvilinear on dB scale. The function is given as

$$P_{OdB}(f)_{\mathrm{DSL[i/o]\_curv}} = \begin{cases} P_{IdB}(f) + \mathrm{TH}_{im}(f) - (\mathrm{TH}_n(f) - \mathrm{FE}(f)), & P_{IdB}(f) < \mathrm{TH}_n(f) - \mathrm{FE}(f) \\ \mathrm{TH}_{im}(f) + [\mathrm{UCL}_{im}(f) - \mathrm{TH}_{im}(f)] \left(\dfrac{P_{IdB}(f) - (\mathrm{TH}_n(f) - \mathrm{FE}(f))}{\mathrm{UCL}_n(f) - \mathrm{TH}_n(f)} \right)^{\frac{g_n}{g_{im}}}, & \\ \qquad\qquad \mathrm{TH}_n(f) - \mathrm{FE}(f) \le P_{IdB}(f) < \mathrm{UCL}_n(f) - \mathrm{FE}(f) & \\ \mathrm{UCL}_{im}(f), \qquad P_{IdB}(f) \ge \mathrm{UCL}_n(f) - \mathrm{FE}(f) & \end{cases} \tag{A.11}$$

where g_n is the exponent of the normal loudness growth function and g_{im} is the exponent of the hearing aid user's loudness growth function obtained from the corresponding loudness growth characteristics estimated using a loudness-scaling procedure. The loudness growth exponent ratio may be taken as 0.5 when $g_n < g_{im}$ and as 2 when $g_n > g_{im}$. Examples of the relation between the output level in the listener's ear canal (dB EC) and the input level in the sound field (dB SF) are shown in Figure A.3, for linear compression and for curvilinear compression with the loudness growth exponent ratios of 0.5 and 2.0.

Scollie et al. [139] proposed the multistage DSL[i/o] procedure, DSLm[i/o], with a four-stage compression function. It uses an input compression threshold higher than the normal hearing threshold. It provides limiting at high levels, wide dynamic range compression at intermediate levels, linear amplification below the compression threshold, and expansion for very low input levels. The compression threshold is kept higher than the normal-hearing

threshold to avoid over-amplification of the noise. The output-limiting threshold is set as 13 dB below UCL_{im}. It compensates for the conductive loss by increasing the UCL_{im} by one-fourth of the conduction component of the hearing loss.

A.10 Correction for Conductive Component of the Mixed Loss

The audiometric data used in the procedures described in Section A.2−A.9 are based on air conduction tests. With the HTLs measured using air conduction and bone conduction tests as $H(f)$ and $H_{BC}(f)$, respectively, the conductive component of the mixed loss is given by the air-bone gap (ABG) as

$$\mathrm{ABG}(f) = H(f) - H_{BC}(f) \tag{A.12}$$

Lybarger [123] proposed adding one-fourth of ABG as correction for the conductive component to the gain as obtained by the half-gain rule for linear amplification, i.e.

$$G(f)_{\mathrm{mixed\_Lybarger\_lin}} = 0.5\,H(f) + 0.25\,\mathrm{ABG}(f) \tag{A.13}$$

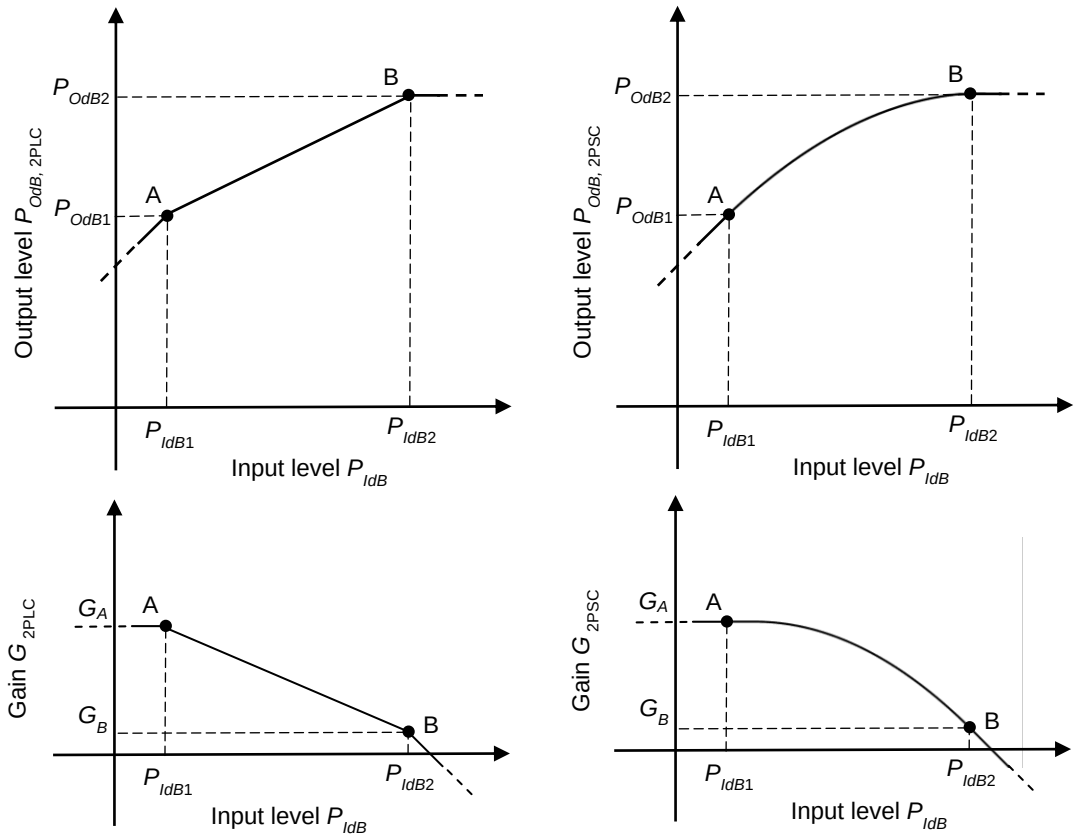
The right side of (A.13) can be also expressed as $0.5\,H_{BC}(f) + 0.75\,\mathrm{ABG}(f)$. Johnson [147] reported that these two expressions do not yield equivalent gain when the correction is applied for compression amplification. He proposed that the compression ratio should be based on the sensorineural component of the hearing loss, with the gain prescription formula given as

$$G(f)_{\mathrm{mixed\_Johnson\_comp}} = G(f)_{\mathrm{comp\_BC}} + 0.75\,\mathrm{ABG}(f) \tag{A.14}$$

where $G(f)_{\mathrm{comp\_BC}}$ is the gain prescription obtained using $H_{BC}(f)$ in place of $H(f)$.

A.11 A Proposed Prescriptive Procedure: Two-Point Smooth Compression

The prescriptive procedures for compression amplification described in Sections A.6−A.9 use thresholds or supra-threshold data for prescribing the gains. The threshold-based procedures, such as FIG6 and NAL-NL2, are based on the assumption that the loudness growth can be predicted from the audiograms. They are easy to apply, as they do not require supra-threshold data. However, the loudness of a sound as perceived by two listeners with similar audiograms may differ widely. The procedures based on supra-threshold data, such as LGOB, IHAFF, and DSLm[i/o], involve estimation of the loudness growth characteristics of the ear to be fitted with a hearing aid. However, measurement of the loudness growth characteristics using multiple loudness categories is a time consuming process and eliciting consistent responses in case of children and elderly is difficult. Further, the differences in the test stimuli (tones, narrowband noises of different bandwidths) may affect the loudness categorization of the stimuli [5].



a) Linear compression b) Curvilinear compression

Figure A.4 Proposed prescriptive procedure with (a) linear compression (2PLC) and (b) curvilinear compression (2PSC): Compression function relating the output and input levels; Gain as a function of the input level ($G_A = P_{OdB1} - P_{IdB1}$, $G_B = P_{OdB2} - P_{IdB2}$).

To overcome these difficulties, we propose a prescriptive procedure for compression amplification using a smooth compression function, with an estimation of the loudness growth characteristics using the ABG and either HTL or MCL data. The proposed procedure is devised for use with sliding-band or multiband compression, in which the gain at the high frequencies is not affected by the level of the low frequency components. Therefore, it does not involve a gain compensation based on averaged speech spectrum. As in the DSL[i/o] procedure, the compression function of the proposed procedure has three segments. The first segment provides linear amplification, the second segment provides compression amplification, and the third segment provides output limiting. The three segments are specified by two points, corresponding to the compression threshold and the output-limiting threshold. Two types of compression functions are proposed. The first function has a linear compression segment and the second one has a curvilinear compression segment selected for slope continuity. The first function is referred to as two-point linear compression (2PLC) and the second function is referred to as two-point smooth compression (2PSC). For a simplified

notation, the frequency dependence of the gains and levels is kept implicit in the subsequent description of the procedure.

The 2PLC function is a three-segment linear relation between the input level P_{IdB} and the output level P_{OdB} on a dB scale, as shown in Figure A.4(a), with the compression threshold and the output-limiting threshold marked as the point A (P_{IdB1}, P_{OdB1}) and the point B (P_{IdB2}, P_{OdB2}), respectively. The function provides linear amplification for $P_{IdB} \le P_{IdB1}$, compression amplification for $P_{IdB1} < P_{IdB} \le P_{IdB2}$, and output limiting for $P_{IdB2} < P_{IdB}$. The three-segment linear compression function is given as

$$P_{OdB,2\mathrm{PLC}} = \begin{cases} P_{IdB} + P_{OdB1} - P_{IdB1}, & P_{IdB} \le P_{IdB1} \\ P_{OdB1} + \dfrac{P_{OdB2} - P_{OdB1}}{P_{IdB2} - P_{IdB1}}(P_{IdB} - P_{IdB1}), & P_{IdB1} < P_{IdB} \le P_{IdB2} \\ P_{OdB2}, & P_{IdB2} < P_{IdB} \end{cases} \tag{A.15}$$

The compression ratio is the ratio of the change in the input level to the change in the output level. Its value for the compression amplification segment is given as

$$\mathrm{CR} = (P_{IdB2} - P_{IdB1}) / (P_{OdB2} - P_{OdB1}) \tag{A.16}$$

The values of the compression ratio for the linear amplification and output-limiting segments are 1 and ∞, respectively. The gain for amplification in dB, $G_{2\mathrm{PLC}} = P_{OdB,2\mathrm{PLC}} - P_{IdB}$, is given as

$$G_{2\mathrm{PLC}} = \begin{cases} P_{OdB1} - P_{IdB1}, & P_{IdB} \le P_{IdB1} \\ P_{OdB1} + \dfrac{P_{IdB} - P_{IdB1}}{\mathrm{CR}} - P_{IdB}, & P_{IdB1} < P_{IdB} \le P_{IdB2} \\ P_{OdB2} - P_{IdB}, & P_{IdB2} < P_{IdB} \end{cases} \tag{A.17}$$

A plot of the gain $G_{2\mathrm{PLC}}$ as a function of P_{IdB} is also shown below the compression function in Figure A.4(a).

The 2PSC function is a three-segment smooth curvilinear relation between P_{IdB} and P_{OdB}, as shown in Figure A.4(b), with the points A and B as in case of 2PLC function. The function provides linear amplification for $P_{IdB} \le P_{IdB1}$, compression amplification for $P_{IdB1} < P_{IdB} \le P_{IdB2}$, and output limiting for $P_{IdB2} < P_{IdB}$. The linear amplification and output limiting segments of this function are the same as those of 2PLC and given in (A.15). The curvilinear compression segment is selected as a power-law function

$$P_{OdB} = a_0 + a_1 \left(P_{IdB} + a_2 \right)^{a_3} \tag{A.18}$$

with the parameters a_0, a_1, a_2, and a_3 selected to provide a smooth transition between the segments. For continuity of the segments and slopes of the segments at the points A and B, the parameters should satisfy the following equations:

$$a_0 + a_1\left(P_{IdB} + a_2\right)^{a_3}\Big|_{P_{IdB}=P_{IdB1}} = P_{OdB1} \tag{A.19}$$

$$a_0 + a_1\left(P_{IdB} + a_2\right)^{a_3}\Big|_{P_{IdB}=P_{IdB2}} = P_{OdB2} \tag{A.20}$$

$$\frac{d}{dP_{IdB}}(a_0 + a_1\left(P_{IdB} + a_2\right)^{a_3})\Big|_{P_{IdB}=P_{IdB1}} = 1 \tag{A.21}$$

$$\frac{d}{dP_{IdB}}(a_0 + a_1\left(P_{IdB} + a_2\right)^{a_3})\Big|_{P_{IdB}=P_{IdB2}} = 0 \tag{A.22}$$

With the parameters obtained by solving (A.19)−(A.22), the three-segment curvilinear compression function is given as

$$P_{OdB,2\mathrm{PSC}} = \begin{cases} P_{IdB} + P_{OdB1} - P_{IdB1}, & P_{IdB} \le P_{IdB1} \\ P_{OdB2} - (P_{OdB2} - P_{OdB1})\left(\dfrac{P_{IdB2} - P_{IdB}}{P_{IdB2} - P_{IdB1}}\right)^{\mathrm{CR}}, & P_{IdB1} < P_{IdB} \le P_{IdB2} \\ P_{OdB2}, & P_{IdB2} < P_{IdB} \end{cases} \tag{A.23}$$

The gain for amplification in dB, $G_{2\mathrm{PSC}} = P_{OdB,2\mathrm{PSC}} - P_{IdB}$, is given as

$$G_{2\mathrm{PSC}} = \begin{cases} P_{OdB1} - P_{IdB1}, & P_{IdB} \le P_{IdB1} \\ P_{OdB2} + (P_{OdB2} - P_{OdB1})\left(\dfrac{P_{IdB2} - P_{IdB}}{P_{IdB2} - P_{IdB1}}\right)^{\mathrm{CR}} - P_{IdB}, & P_{IdB1} < P_{IdB} \le P_{IdB2} \\ P_{OdB2} - P_{IdB}, & P_{IdB2} < P_{IdB} \end{cases} \tag{A.24}$$

A plot of the gain $G_{2\mathrm{PSC}}$ as a function of P_{IdB} is also shown below the compression function in Figure A.4(b).

The input and output levels, in SPL, corresponding to the points A and B may be obtained using the audiometric data of a normal-hearing listener and the hearing-impaired listener, with the point A corresponding to HTL and the point B corresponding to UCL. The input and output levels for the point A are taken as HTL_n (hearing threshold for normal hearing, in SPL) and HTL_{im} (hearing threshold for the hearing-impaired listener, in SPL), respectively. The input and output levels for the point B are taken as UCL_n (uncomfortable level for normal hearing, in SPL) and UCL_{im} (uncomfortable level for the hearing-impaired listener, in SPL),

respectively. With these points, P_{IdB1}, P_{IdB2}, the maximum gain (gain for the linear amplification segment) $G_{\max}$, and CR are given as the following:

$$P_{IdB1,\mathrm{HTL}} = \mathrm{HTL}_n \tag{A.25}$$

$$P_{IdB2,\mathrm{HTL}} = \mathrm{UCL}_n \tag{A.26}$$

$$G_{\max,\mathrm{HTL}} = \mathrm{HTL}_{im} - \mathrm{HTL}_n \tag{A.27}$$

$$\mathrm{CR}_{\mathrm{HTL}} = (\mathrm{UCL}_n - \mathrm{HTL}_n) / (\mathrm{UCL}_{im} - \mathrm{HTL}_{im}) \tag{A.28}$$

It may be noted that the linear gain $G_{\max,\mathrm{HTL}}$ is the hearing threshold H in dB HL as obtained from the audiogram. Taking the dynamic range of a listener with normal hearing as 90 dB, $\mathrm{UCL}_n = 90 + \mathrm{HTL}_n$. The uncomfortable level for the hearing-impaired listener with a mixed loss may be obtained by adding the air-bone gap (ABG) to UCL_n. Therefore, $\mathrm{UCL}_{im} = \mathrm{UCL}_n + \mathrm{ABG}$. The dynamic range of the hearing-impaired listener can be given as $\mathrm{UCL}_{im} - \mathrm{HTL}_{im} = ((90 + \mathrm{HTL}_n) + \mathrm{ABG}) - \mathrm{HTL}_{im} = 90 + \mathrm{ABG} - H$. A compression amplification is provided for the input levels between HTL_n and UCL_n. With these values, the relations in (A.25)−(A.28) can be given as

$$P_{IdB1,\mathrm{HTL}} = \mathrm{HTL}_n \tag{A.29}$$

$$P_{IdB2,\mathrm{HTL}} = \mathrm{HTL}_n + 90 \tag{A.30}$$

$$G_{\max,\mathrm{HTL}} = H \tag{A.31}$$

$$\mathrm{CR}_{\mathrm{HTL}} = 90 / (90 + \mathrm{ABG} - H) \tag{A.32}$$

With the HTL-based thresholds, the gain for linear compression $G_{2\mathrm{PLC\_HTL}}$ and the gain for smooth compression $G_{2\mathrm{PSC\_HTL}}$ are given as the following:

$$G_{2\mathrm{PLC\_HTL}} = \begin{cases} H, & P_{IdB} \le \mathrm{HTL}_n \\ (\mathrm{HTL}_n + H) + \dfrac{(P_{IdB} - \mathrm{HTL}_n)}{\mathrm{CR}_{\mathrm{HTL}}} - P_{IdB}, & \mathrm{HTL}_n < P_{IdB} \le \mathrm{HTL}_n + 90 \\ \mathrm{HTL}_n + 90 + \mathrm{ABG} - P_{IdB}, & \mathrm{HTL}_n + 90 < P_{IdB} \end{cases} \tag{A.33}$$

$$G_{2\mathrm{PSC\_HTL}} = \begin{cases} H, & P_{IdB} \le \mathrm{HTL}_n \\ (\mathrm{HTL}_n + 90 + \mathrm{ABG}) - P_{IdB} + & \\ (90 + \mathrm{ABG} - H)\left(\dfrac{\mathrm{HTL}_n + 90 - P_{IdB}}{90}\right)^{\mathrm{CR}_{\mathrm{HTL}}}, & \mathrm{HTL}_n < P_{IdB} \le \mathrm{HTL}_n + 90 \\ \mathrm{HTL}_n + 90 + \mathrm{ABG} - P_{IdB}, & \mathrm{HTL}_n + 90 < P_{IdB} \end{cases} \tag{A.34}$$

For high intelligibility of the conversational speech, it may be more appropriate to set the point A in Figure A.4 corresponding to the most comfortable level (MCL), with $P_{IdB1} = \mathrm{MCL}_n$ and $P_{OdB1} = \mathrm{MCL}_{im}$. With this setting, $G_{\max}$, P_{IdB1}, P_{IdB2}, and CR are given as the following:

$$P_{IdB1,\mathrm{MCL}} = \mathrm{MCL}_n \tag{A.35}$$

$$P_{IdB2,\mathrm{MCL}} = \mathrm{UCL}_n \tag{A.36}$$

$$G_{\max,\mathrm{MCL}} = \mathrm{MCL}_{im} - \mathrm{MCL}_n \tag{A.37}$$

$$\mathrm{CR}_{\mathrm{MCL}} = (\mathrm{UCL}_n - \mathrm{MCL}_n) / (\mathrm{UCL}_{im} - \mathrm{MCL}_{im}) \tag{A.38}$$

Taking MCL at the center of the dynamic range of hearing, $\mathrm{MCL}_n = \mathrm{HTL}_n + 45$ and $\mathrm{MCL}_{im} = (\mathrm{HTL}_n + \mathrm{HTL}_{im} + \mathrm{ABG})/2 + 45$. With these values, the relations in (A.35)−(A.38) can be given as

$$P_{IdB1,\mathrm{MCL}} = \mathrm{HTL}_n + 45 \tag{A.39}$$

$$P_{IdB2,\mathrm{MCL}} = \mathrm{HTL}_n + 90 \tag{A.40}$$

$$G_{\max,\mathrm{MCL}} = (H + \mathrm{ABG}) / 2 \tag{A.41}$$

$$\mathrm{CR}_{\mathrm{MCL}} = 90 / (90 + \mathrm{ABG} - H) \tag{A.42}$$

With the MCL-based thresholds, the gain for linear compression $G_{2\mathrm{PLC\_MCL}}$ and the gain for smooth compression $G_{2\mathrm{PSC\_MCL}}$ are given as the following:

$$G_{2\mathrm{PLC\_MCL}} = \begin{cases} (H + \mathrm{ABG}) / 2, & P_{IdB} \le \mathrm{HTL}_n + 45 \\ (\mathrm{HTL}_n + \mathrm{HTL}_{im} + \mathrm{ABG}) / 2 + 45 & \\ \quad + \dfrac{P_{IdB} - \mathrm{HTL}_n - 45}{\mathrm{CR}_{\mathrm{MCL}}} - P_{IdB}, & \mathrm{HTL}_n + 45 < P_{IdB} \le \mathrm{HTL}_n + 90 \\ \mathrm{HTL}_n + 90 + \mathrm{ABG} - P_{IdB}, & \mathrm{HTL}_n + 90 < P_{IdB} \end{cases} \tag{A.43}$$

$$G_{2\mathrm{PSC\_MCL}} = \begin{cases} (H + \mathrm{ABG}) / 2, & P_{IdB} \le \mathrm{HTL}_n + 45 \\ \mathrm{HTL}_n + 90 + \mathrm{ABG} & \\ \quad + \dfrac{(90 + \mathrm{ABG} - H)}{2} \left(\dfrac{\mathrm{HTL}_n + 90 - P_{IdB}}{45} \right)^{\mathrm{CR}_{\mathrm{MCL}}} & \\ -P_{IdB}, & \mathrm{HTL}_n + 45 < P_{IdB} \le \mathrm{HTL}_n + 90 \\ \mathrm{HTL}_n + 90 + \mathrm{ABG} - P_{IdB}, & \mathrm{HTL}_n + 90 < P_{IdB} \end{cases} \tag{A.44}$$

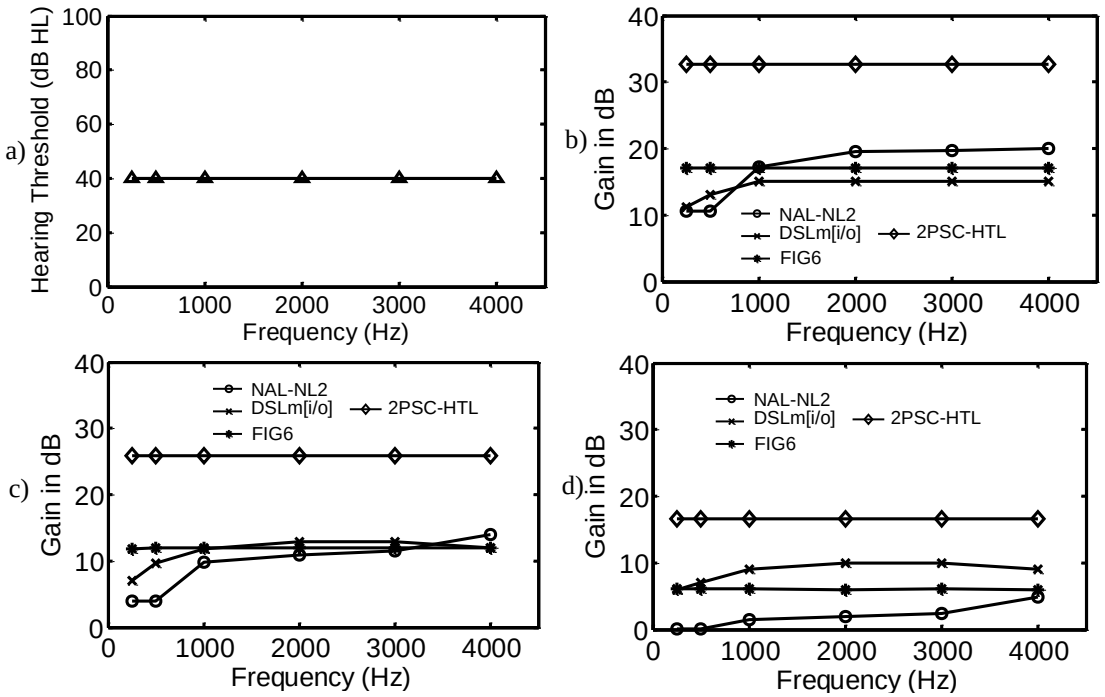


Figure A.5 Gains prescribed by FIG6, NAL-NL2, DSLm[i/o], and proposed prescriptive procedure 2PSC-HTL for a flat loss: (a) Hearing thresholds, (b) gains prescribed at 50 dB SPL input, (c) gains prescribed at 65 dB SPL input, and (d) gains prescribed at 80 dB SPL input.

It may be noted that changing the compression threshold from the HTL point (as given by the relations in (A.29)−(A.32)) to the MCL point and assuming the MCL point to be at the center of the dynamic range of hearing (as given by the relations in (A.39)−(A.42)) results in a lower gain for linear amplification. However, there is no change in CR. It is expected that use of the MCL_{im} value estimated by a loudness scaling test, if practical, will give a better fit of the hearing aid in terms of comfort and speech intelligibility.

A.12 Comparison of Gains Prescribed by Compression Amplification Procedures

For comparing gains prescribed by different procedures for compression amplification, the gains prescribed by FIG6, NAL-NL2, and DSLm[i/o] were obtained for an audiogram with flat loss and one with sloping loss. The gains as prescribed by these three procedures for the two audiograms at 0.25, 0.50, 1, 2, 3 and 4 kHz were obtained from Figures 10.12 and 10.13 in [5]. The gains prescribed by the proposed procedure with smooth compression and with the point A set as HTL, referred to as 2PSC-HTL, were also obtained for the same audiograms, using the gain as given in (A.34).

The audiogram with the flat loss and the corresponding prescribed gains for the input levels of 50, 65, and 80 dB SPL are shown in Figure A.5. The audiogram has a flat loss of 40 dB. For all the procedures, the gains are highest for the input level of 50 dB SPL and decrease

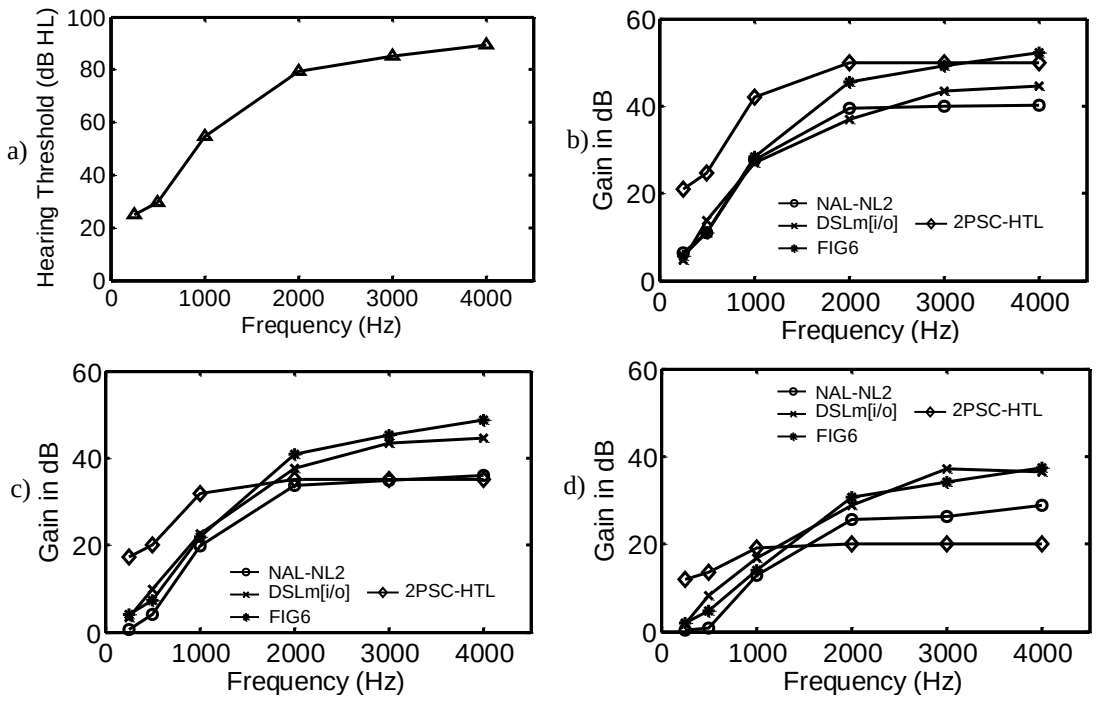


Figure A.6 Gains prescribed by FIG6, NAL-NL2, DSLm[i/o], and proposed prescriptive procedure 2PSC-HTL for a sloping loss: (a) Hearing thresholds, (b) gains prescribed at 50 dB SPL input, (c) gains prescribed at 65 dB SPL input, and (d) gains prescribed at 80 dB SPL input.

with increase in the input level. FIG6, based on the loudness normalization criterion, prescribes a flat gain. DSLm[i/o], based on loudness normalization criterion and considering the low-frequency dominance in the averaged speech spectrum, prescribes higher gains at higher frequencies. NAL-NL2, based on the intelligibility maximization criterion, prescribes low gain at low frequencies to avoid masking of high frequency components. The proposed procedure prescribes a flat gain higher than FIG6, NAL-NL2, and DSLm[i/o] for all input levels.

The audiogram with the sloping loss and the corresponding prescribed gains for the input levels of 50, 65, and 80 dB SPL are shown in Figure A.6. For all the procedures, the gains decrease with increase in the input level. The gains prescribed by FIG6 and DSLm[i/o] are similar. NAL-NL2 prescribes lower gains at high frequencies as higher hearing thresholds at these frequencies are generally associated with increased temporal and spectral masking. The gains prescribed by the proposed procedure are similar to those by the other three procedures at input level of 50 dB SPL. For higher input levels, the proposed procedure prescribes a lower gain at high frequencies where the thresholds are higher to avoid the output from becoming uncomfortably loud. Use of the proposed prescriptive procedure in hearing aids with compression amplification and its evaluation by a large number of hearing-impaired listeners is needed for real-life evaluation and further enhancement.

Left blank

Appendix B

A TECHNIQUE WITH LOW MEMORY AND COMPUTATIONAL REQUIREMENTS FOR DYNAMIC TRACKING OF QUANTILES

B.1 Introduction

Quantiles estimated from the observations or samples of a random variable are useful in monitoring the statistics of several types of data such as transaction records, data packets in a network, call durations, etc. They are also used in several applications such as query optimizers, data mining, risk assessment, regulatory action, workload distribution among processors in a parallel database system, and signal processing for noise suppression [13], [14], [148]–[152]. The p-quantile (or 100p-percentile) q of a variable X is given as

$$q = \min(x : \text{Prob}(X \leq x) \geq p) \tag{B.1}$$

For N samples of X forming the sequence $\{x(1), x(2), ..., x(N)\}$, the sequence is sorted in ascending order as $\{x((1)), x((2)), ..., x((N))\}$ and a point estimate of the p-quantile is obtained as

$$q = x((\lceil pN \rceil)) \tag{B.2}$$

This estimate from the order statistics of the samples is known as the sample quantile.

Sorting operations for finding the sample quantile require large memory space and multiple passes through the data. For applications involving a long sequence of stored data (e.g., disk-resident data), many techniques offering different trade-offs between the memory space and the number of passes are available [105], [153], [154]. For applications where approximate quantiles may be acceptable, many estimation techniques with guaranteed error bounds and using a memory for buffering a small fraction of the sequence length have been reported [148]–[151], [155]–[158]. Several techniques using finite-length buffers with specific data structures have been reported for updating quantile summaries of a data stream as its samples arrive [159]–[163]. The memory space and updating time in these techniques are functions of the accuracy, the number of samples in the data stream, and the number of possible values of the samples. Therefore, these techniques are not usable for online quantile tracking of a data stream with an unrestricted number of samples.

For online approximate calculation of quantiles of data streams, Jain and Chlamtac [109] proposed a heuristic algorithm with low memory requirement. It uses five markers, corresponding to minimum, p/2-quantile, p-quantile, (1+p)/2-quantile, and maximum, with

each marker stored as a point with the height as the quantile value and the horizontal location as the number of samples which are less than or equal to the quantile value. The quantile is recursively adjusted by applying increments calculated using a piecewise-parabolic formula and inversely proportional to the number of samples. As the successive samples have a decreasing effect on the estimate, the technique is not usable for non-stationary data.

A quantile estimation technique with low memory requirement and based on stochastic approximation [104] was proposed by Tierney [106]. For input sample $x(n)$, the estimate of p-quantile $\hat{q}(n)$ is recursively calculated as $\hat{q}(n) = \hat{q}(n-1) + d(n)$, with the increment $d(n)$ calculated as

$$d(n) = [p - I_{(-\infty,\hat{q}(n-1))}(x(n))]\, c(n) \tag{B.3}$$

The indicator function $I_{(\cdot)}$ and control sequence $c(n)$ in the above equation are given as the following:

$$I_{(-\infty,\hat{q}(n-1))}(x(n)) = \begin{cases} 1, & x(n) < \hat{q}(n-1) \\ 0, & \text{otherwise} \end{cases} \tag{B.4}$$

$$c(n) = 1/[n\, w(n)] \tag{B.5}$$

$$w(n) = \max(\hat{f}(n), \hat{f}(0)/\sqrt{n}) \tag{B.6}$$

where $\hat{f}$ is the recursively estimated probability density at q. The initial estimates of q and $\hat{f}$ are obtained from a set of initial samples. The memory requirement does not change with the sequence length and the estimated value for stationary data converges to the sample quantile. The technique involves additional calculation for estimating the probability density, with a lower bound set on it to prevent the estimation from becoming unstable. Möller *et al.* [107] investigated statistical properties of the stochastic approximation for quantile estimation using different control sequences for reducing the variance of the estimate for stationary data and improving the adaptivity of the estimate for non-stationary data. It was shown that the estimate $\hat{q}(n)$ converges in L_1 to p-quantile q, i.e., $\lim_{n\to\infty} E(\hat{q}(n)) = q$, with a finite variance for $c(n) \in (0, 1/M)$, with M as the upper bound of the density function f. Further, $\hat{q}(n)$ converges almost surely to q for the control sequence $c(n) = e/n$, and that the asymptotic variance of the estimator is minimum if $e = 1/f$. They proposed recursive histogram and slope methods to estimate f and showed that adaptivity and variance are contrary factors.

Amiri and Thiam [110] proposed a quantile estimation technique using smooth stochastic approximation by replacing the indicator function in (B.3) by the smooth function

$H((\hat{q}(n-1)-x(n))/b(n))$, with $H(z) = 1/(1+\exp(-z))$ and $b(n)$ as the bandwidth sequence. The control sequence $c(n)$ as in (B.5) is calculated using

$$w(n) = \max(\mu, \min(\hat{f}_n, \nu \ln(n+1)) \tag{B.7}$$

where μ and ν are positive constants. The density function $\hat{f}$ is recursively calculated as

$$\hat{f}(n) = (1-n^{-1})\hat{f}(n-1)+[K((\hat{q}(n-1)-x(n)/h(n))]/[nh(n)] \tag{B.8}$$

Using $K(\bullet)$ as a positive bounded kernel function and $h(n)$ as the bandwidth sequence. The estimate using Gaussian function as the kernel function and $b(n)=h(n)=an^{-\xi}$ with $a>0$ and $1/5<\xi<1/2$ converges almost surely. An initial density estimate is obtained using the first n_0 samples as

$$\hat{f}(n_0) = [\sum_{i=1}^{n_0} K((\hat{q}(i-1)-x(i)/\tilde{h}(i))]/[n_0\tilde{h}(i)] \tag{B.9}$$

with $\tilde{h}(i) = 1.06\,\sigma_i i^{-1/5}$ as the initial bandwidth sequence and σ_i as the standard deviation of the preceding i samples. Calculation of Gaussian kernel has relatively high computational requirement and the estimation is sensitive to initial values of the density and the bandwidth sequence.

In the methods using iterative calculation of the quantile using an increment weighted by $1/n$ [106], [109], [110], the estimated values may change by a reordering of the samples. For dynamic tracking of quantiles, Chen *et al.* [108] proposed an exponentially weighted stochastic approximation by replacing $1/n$ in the control sequence by a constant factor. In this method, the probability density is calculated recursively as

$$\hat{f}(n) = (1-\xi)\hat{f}(n-1)+\frac{\xi\, I_{(-\infty,2\sqrt{2}\,r(n-1))}(x(n)-\hat{q}(n-1))}{2\sqrt{2}\,r(n-1)} \tag{B.10}$$

where $\xi = 0.05$ and $r(n)$ is the inter-quartile range obtained as the difference of the current estimates of 0.75- and 0.25-quantiles. The initial quantile and density estimates are obtained using first 10 samples. Cao *et al.* [164] proposed a stochastic approximation technique for tracking multiple quantiles of a data stream for a set of probabilities p_i, maintaining the monotonicity of the estimated quantiles. It has a high computational requirement as it involves estimating multiple probability densities.

We present a technique for dynamic tracking of quantiles with low memory and computational requirements for use in applications involving real-time estimation of quantiles of a data stream. The quantile is estimated recursively by applying an increment, calculated as a fraction of the range, such that the estimated quantile converges to the sample quantile. The range is dynamically estimated using first-order recursive relations for peak and valley

detection. The technique provides a trade-off between variance and adaptivity of the estimation. The memory and computational requirements of the technique are independent of the number of samples in the data stream and the number of possible values of the samples. It is tested using synthetic and real data with different distributions and compared with some of the techniques having similar features and reported earlier. The proposed technique for dynamic tracking of quantiles is described in the second section. The data used for testing are described in the third section. The test results along with discussion are presented in the fourth section, followed by the conclusion in the last section.

B.2 Dynamic Tracking of Quantiles

The quantile is dynamically estimated as the input sample of the data stream arrives by applying an increment Δ_+ or a decrement Δ_- on the previous estimate. The values of Δ_+ and Δ_- are calculated as appropriate fractions of the range of the input samples such that the estimate after a sufficiently large number of input samples converges to the sample quantile. As the underlying distribution of the data is unknown, the range also needs to be dynamically estimated. The method uses a step-size control factor for a trade-off between variance and adaptivity of the estimation. For a more controlled trade-off, a technique with two-stage dynamic tracking is presented in which the increments and decrements are fractions of a range segment bracketing the quantile. The techniques using range estimation and range segment estimation are presented in the following two subsections, followed by a comparison of the computational and storage requirements in the third subsection.

B.2.1 Dynamic quantile tracking using range estimation (DQTRE)

For input sample $x(n)$, the estimate of p-quantile is calculated recursively as

$$\hat{q}(n) = \hat{q}(n-1) + d(n) \tag{B.11}$$

where the increment $d(n)$ is given as

$$d(n) = \begin{cases} \Delta_+, & x(n) \geq \hat{q}(n-1) \\ -\Delta_-, & \text{otherwise} \end{cases} \tag{B.12}$$

The values of Δ_+ and Δ_- should be such that the quantile estimate converges to the sample quantile and sum of the increments approaches zero, i.e., $\sum d(n) \approx 0$. For stationary data and sufficiently large number of input samples N, $d(n)$ is expected to be $-\Delta_-$ for pN samples and Δ_+ for $(1-p)N$ samples. Therefore, we should have

$$(1-p)N\Delta_+ - pN\Delta_- = 0 \tag{B.13}$$

which results in

$$\Delta_+ / \Delta_- = p / (1-p) \tag{B.14}$$

Therefore, Δ_+ and Δ_- may be selected as

$$\Delta_+ = \lambda p R \tag{B.15}$$

$$\Delta_- = \lambda (1-p) R \tag{B.16}$$

where R is the range (difference between the maximum and minimum values) and λ is the step-size control factor. It can be shown, as in [107], that $\lim_{n\to\infty} E(\hat{q}(n)) = q$, if $0 < \lambda < 1/(MR)$, where M is the peak of the density function. Near convergence, the peak-to-peak ripple δ in the estimated values is $\Delta_+ + \Delta_-$ and therefore it is given as

$$\delta = \lambda R \tag{B.17}$$

During tracking, the number of steps needed for the estimated value to change from its initial value q_{in} to its final value q_{fin} is given as to final value

$$s = \max\{(q_{\text{fin}} - q_{\text{in}}) / \Delta_+,\ (q_{\text{in}} - q_{\text{fin}}) / \Delta_-\} \tag{B.18}$$

Since $(|q_{\text{fin}} - q_{\text{in}}|)_{\max} = R$, the number of steps is given as

$$s = \max\left\{\frac{1}{\lambda p},\ \frac{1}{\lambda(1-p)}\right\} \tag{B.19}$$

It may be noted that s becomes very large for very low or high values of p. The value of λ is selected to ensure convergence and for an appropriate trade-off between δ and s which are related to variance and adaptivity, respectively.

The range is estimated using dynamic peak and valley detectors. The peak estimate $\hat{P}(n)$ and the valley estimate $\hat{V}(n)$ are updated without any restriction on their polarity, using the following first-order recursive relations:

$$\hat{P}(n) = \begin{cases} \alpha\hat{P}(n-1) + (1-\alpha)x(n), & x(n) \geq \hat{P}(n-1) \\ \beta\hat{P}(n-1) + (1-\beta)\hat{V}(n-1), & \text{otherwise} \end{cases} \tag{B.20}$$

$$\hat{V}(n) = \begin{cases} \alpha\hat{V}(n-1) + (1-\alpha)x(n), & x(n) \leq \hat{V}(n-1) \\ \beta\hat{V}(n-1) + (1-\beta)\hat{P}(n-1), & \text{otherwise} \end{cases} \tag{B.21}$$

and the range is tracked as

$$\hat{R}(n) = \hat{P}(n) - \hat{V}(n) \tag{B.22}$$

The coefficients α and β are selected in the range [0, 1] to control the rise and fall rates of the range estimation. An over-estimation of the range results in increased variance in the quantile estimation and an under-estimation of the range results in lower adaptivity. Considering adaptivity to be more important for applications involving non-stationary data, we select a

small α to provide fast response to an increase in the range and a large β for a slow response to a decrease in the range.

Let us consider the data to be stationary for sequence length greater than L samples, with the successive peaks (and the successive valleys) separated by at the most L samples. Let the peak and valley values be P and V, respectively. Further, let the peak detector output be $\hat{P}_1$ at the input peak and be $\hat{P}_2$ just before it. These values can be obtained using the recursive relation in (B.20) and (B.21), as $\hat{P}_1 = \alpha\hat{P}_2 + (1-\alpha)P$ and $\hat{P}_2 \approx \beta^L\hat{P}_1 + (1-\beta^L)V$. The peak-to-peak ripple in the peak estimation is given as

$$\hat{P}_1 - \hat{P}_2 = (1-\alpha)\frac{1-\beta^L}{1-\alpha\beta^L}(P-V) \tag{B.23}$$

With the valley detector output as $\hat{V}_1$ at the input valley and as $\hat{V}_2$ just before it, the peak-to-peak ripple in the valley estimation $\hat{V}_1 - \hat{V}_2$ can be shown to be the same. Therefore, the peak-to-peak ripple in the range estimation is given as

$$(\hat{P}_1 - \hat{V}_1) - (\hat{P}_2 - \hat{V}_2) = 2(1-\alpha)\frac{1-\beta^L}{1-\alpha\beta^L}(P-V) \tag{B.24}$$

With the range $R = P - V$, the peak-to-peak ripple as a fraction of R is given as

$$r = 2(1-\alpha)(1-\beta^L)/(1-\alpha\beta^L) \tag{B.25}$$

where the data may be considered as stationary for sequence length greater than L samples. We select $\alpha << 1$ for fast rise and $\beta = (1-\alpha)^{1/L}$ for keeping the L-sample fall error equal to 1-sample rise error, resulting in $r \approx 2\alpha$.

With the range $\hat{R}(n)$ tracked as (B.22), the dynamic quantile tracking as given by (B.11), (B.12), (B.15), and (B.16) can be rewritten as the following:

$$\hat{q}(n) = \begin{cases} \hat{q}(n-1) + \lambda p\hat{R}(n), & x(n) \geq \hat{q}(n-1) \\ \hat{q}(n-1) - \lambda(1-p)\hat{R}(n), & \text{otherwise} \end{cases} \tag{B.26}$$

The technique, comprising the computation steps as given by (B.20), (B.21), (B.22), and (B.26) and using sample delay operations, is shown as a block diagram in Figure B.1. The estimated quantile may be low-pass filtered for reducing the ripple.

The technique may be used for obtaining multiple quantile values $(q_1, q_2, \ldots q_M)$ of the input data stream for a set of probabilities $(p_1, p_2, \ldots p_M)$, using a quantile estimator for each probability

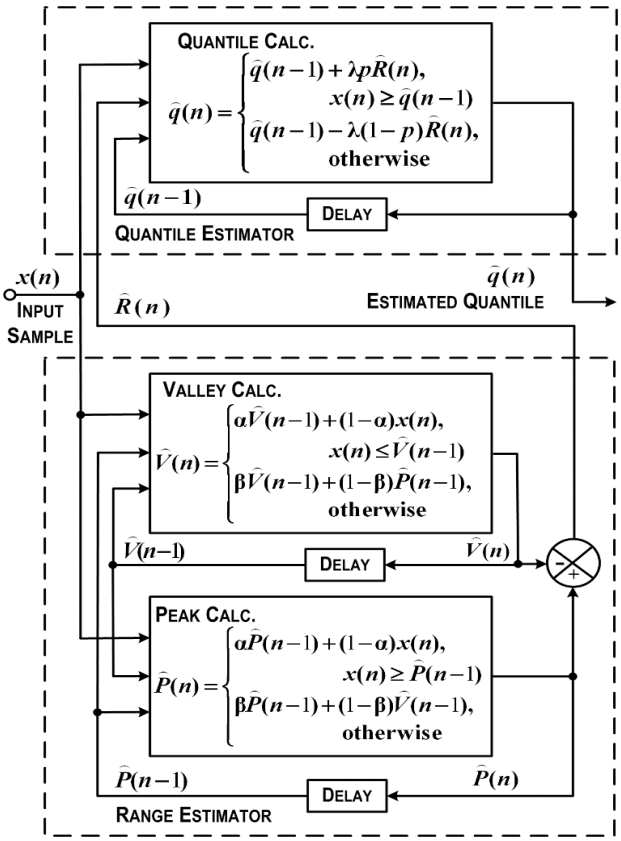


Figure B.1 Dynamic quantile tracking using range estimation.

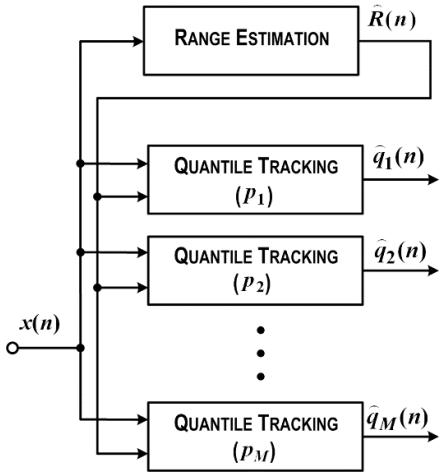


Figure B.2 Dynamic quantile tracking of multiple quantiles.

and a common range estimator as shown in Figure B.2. The estimated quantiles may not be a monotonic function of probability, particularly for small differences between successive probabilities. The deviations from monotonicity at any sample starts getting corrected at the

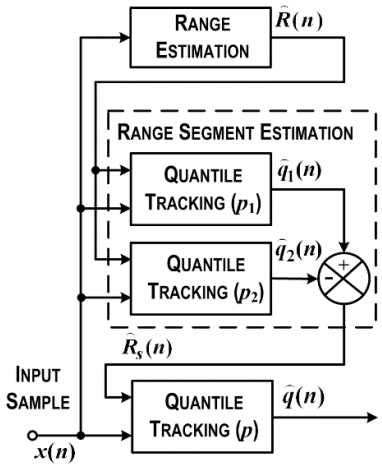


Figure B.3 Dynamic quantile tracking using range segment estimation.

next sample and the deviations have an upper bound of λR. The estimated quantile values may be processed to correspond to a monotonic cumulative distribution function.

B.2.2 Dynamic quantile tracking using range segment estimation (DQTRSE)

The ripple in quantile estimation, particularly in the vicinity of the peaks in the distribution, can be reduced by using a two-stage dynamic quantile tracking in which the increment is a fraction of a range segment bracketing the *p*-quantile. The technique is shown as a block diagram in Figure B.3. Two probabilities bracketing p and with the bracketing interval p_b are selected as the following:

$$p_1 = (1 - p_b)\, p \tag{B.27}$$

$$p_2 = (1 - p_b)\, p + p_b \tag{B.28}$$

In the first stage, the p_1-quantile and the p_2-quantile are estimated using DQTRE. The range segment is estimated as the difference between the estimates of p_1-quantile and p_2-quantile with a lower bound as the following:

$$\hat{R}_s(n) = \max((\hat{q}_2(n) - \hat{q}_1(n)),\ lp_b\hat{R}(n)) \tag{B.29}$$

Where l is the segment-limiting fraction selected to avoid too small a segment which may lead to poor adaptivity. In the second stage, the range segment estimated in the first stage is used to obtain Δ_+ and Δ_- as the following:

$$\Delta_+ = \lambda p\hat{R}_s(n) / p_b \tag{B.30}$$

$$\Delta_- = \lambda(1 - p)\hat{R}_s(n) / p_b \tag{B.31}$$

where λ is the step-size control factor as described earlier. The *p*-quantile is estimated as the following

$$\hat{q}(n) = \begin{cases} \hat{q}(n-1) + \lambda p \hat{R}_s(n) / p_b, & x(n) \geq \hat{q}(n-1) \\ \hat{q}(n-1) - \lambda(1-p)\hat{R}_s(n) / p_b, & \text{otherwise} \end{cases} \quad \text{(B.32)}$$

B.2.3 Comparison of computational and storage requirements

The computational and storage requirements of the proposed technique are compared with those of some of the techniques reported earlier for quantile estimation of data streams: heuristic algorithm using piecewise-parabolic formula (P2) by Jain and Chlamtac [109], smooth stochastic approximation (SSA) technique by Amiri and Thiam [110], and exponentially weighted stochastic approximation (EWSA) technique by Chen *et al.* [108]. The storage and computational requirements of these techniques are independent of the number of samples in the data stream, the number of possible values of the samples, and the type of distribution.

For examining the relative suitability of the techniques for online quantile estimation, we compare the computational operations per sample after initialization and the storage requirement. The computational steps and associated operations, in terms of the numbers of addition (subtraction), multiplication (division), comparison, and other operations, are shown in Table B.1. The total numbers of operations (sum of the operations across the steps) and storage requirement (in terms of the number of variables and constants involved in the recursive calculations) are shown in Table B.2. The computational requirement of SSA in terms of basic arithmetic operations is very high as it involves power, exponent, and log operations. Use of look-up table for reducing the arithmetic operations will significantly increase the memory requirement. The computational requirement of P2 is much lower than that of SSA but significantly higher than that of the other three. The computational requirements of DQTRSE and EWSA are almost comparable and about 40% higher than that of DQTRE. The storage requirements are highest for P2 and lowest for DQTRE.

The computational operations needed for initialization also need to be examined as these may contribute to the overall resource requirement, particularly for dedicated hardware or resource-constrained implementations. P2 requires 5 input samples and their sorting for initialization. SSA uses a relatively much larger number of input samples and computational operations. EWSA uses typically 10 input samples and calculation of inter-quartile range and probability density. The initial values of recursion variables of DQTRE and DQTRSE may be set as zero and hence these techniques do not need any additional resources for initialization.

Table B.1 Computation steps and number of operations in quantile estimation using the P2 [109], SSA [110], EWSA [108], DQTRE, and DQTRSE techniques.

Technique	Computation steps and number of operations
P2	1) Updating of desired marker positions: 4 additions. 2) Updating of current marker positions: ≤ 4 additions, ≤ 3 comparisons. 3) Calculation of shift in marker positions: 3 additions. 4) Parabolic interpolation: 27 additions, 18 multiplications. 5) Arranging the marker values in non-decreasing order: 4 comparisons. 6) Linear interpolation: ≤ 9 additions, ≤ 3 multiplications. 7) Post-shift updating of current marker positions: 4 additions.
SSA	1) Calculation of weights: 1 log, 1 multiplication, 2 comparisons. 2) Calculation of smooth function: 2 additions, 2 multiplications, 1 power. 3) Calculation of bandwidth: 1 multiplication, 1 power. 4) Density calculation: 3 additions, 8 multiplications, 2 powers. 5) Quantile calculation: 2 additions, 3 multiplications.
EWSA	1) Inter-quartile range calculation: 3 additions, 5 multiplications, 2 comparisons. 2) Density calculation: 2 additions, 3 multiplications, 1 comparison. 3) Quantile calculation: 1 addition, 2 multiplications, 1 comparison.
DQTRE	1) Peak and valley calculation: ≤ 2 comparisons, 2 additions, 4 multiplications. 2) Range calculation: 1 addition. 3) Quantile calculation: 1 addition, 1 multiplication, 1 comparison.
DQTRSE	1) Peak and valley calculation: ≤ 2 comparisons, 2 additions, 4 multiplications. 2) Range calculation: 1 addition. 3) $p1$- and $p2$-quantile calculation: 2 additions, 2 multiplications, 2 comparisons. 4) Range segment calculation: 1 addition, 1 multiplication, 1 comparison. 3) Quantile tracking: 1 addition, 1 multiplication, 1 comparison.

Table B.2 Number of operations per sample and storage for quantile estimation using the P2 [109], SSA [110], EWSA [108], DQTRE, and DQTRSE techniques.

Technique	No. of operations per sample						Storage	
	Addition	Multi-plication	Compa-rison	Power	Expo-nent	Log	No. of Variables	No. of Constants
P2	51	21	7	-	-	-	20	3
SSA	7	15	2	2	2	1	9	4
EWSA	6	10	4	-	-	-	9	4
DQTRE	4	5	3	-	-	-	7	4
DQTRSE	7	8	6	-	-	-	12	9

As the two proposed techniques have low memory and computational requirements and do not need any additional resources for initialization, they can be used for real-time tracking of multiple quantiles of multiple variables using a microcontroller, DSP chip, FPGA, or ASIC, enabling the use of order statistics in signal processing and control applications.

B.3 Comparison of computational and storage requirements

The techniques were tested using synthetic and real data with different distributions. The synthetic stationary data consisted of sequences of random numbers with several symmetrical and asymmetrical density functions. For synthetic non-stationary data, the model parameters

were varied as a function of sample number. The real data were obtained from audio waveforms.

B.3.1 Synthetic Stationary Data

To evaluate the performance of quantile estimation for data with different underlying distributions, sequences of random numbers were generated corresponding to uniform, Gaussian, exponential, and Gaussian mixture probability density functions, as described below.

a) Uniform distribution with range of $[x_1, x_2]$

$$f(x) = (u(x-x_1)-u(x-x_2))/(x_2-x_1), \quad x_2 > x_1$$

with $x_1 = 0$, $x_2 = 1$. It has mean $\mu = 1/2$ and standard deviation $\sigma = 1/\sqrt{12}$.

b) Gaussian distribution

$$f(x) = (1/(\sqrt{2\pi}\sigma))\exp(-(x-\mu)^2/(2\sigma^2))$$

with $\mu = 1/2$ and $\sigma = 1/4$.

c) Exponential distribution

$$f(x) = \mu \exp(-\mu x)u(x)$$

with $\mu = 1/6$. It has $\sigma = 1/6$.

d) Mixture of two Gaussian distributions

$$f(x) = (k_1/(\sqrt{2\pi}\sigma_1))\exp(-(x-\mu_1)^2/(2\sigma_1^2)) + (k_2/(\sqrt{2\pi}\sigma_2))\exp(-(x-\mu_2)^2/(2\sigma_2^2))$$

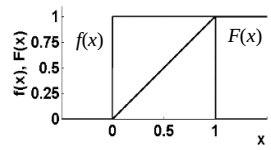
with $k_1 = k_2 = 0.5$, $\mu_1 = 0.3$, $\mu_2 = 0.7$, $\sigma_1 = \sigma_2 = 0.1$.

e) Mixture of two Gaussian distributions with $k_1 = 1/3$, $k_2 = 2/3$, $\mu_1 = 0.3$, $\mu_2 = 0.7$, $\sigma_1 = \sigma_2 = 0.1$.

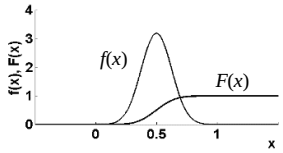
f) Mixture of two Gaussian distributions with $k_1 = 2/3$, $k_2 = 1/3$, $\mu_1 = 0.3$, $\mu_2 = 0.7$, $\sigma_1 = \sigma_2 = 0.1$.

The probability density functions $f(x)$ and cumulative distribution functions $F(x)$ of the test sequences are shown in Figure B.4.

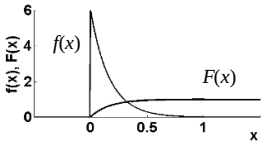
The sequences were generated using MATLAB, deleting the sample values outside [0, 1] in case of Gaussian and exponential distributions. Calculation of the order statistics for sequences of different lengths showed that the synthesized data may be considered as stationary for sequence lengths greater than 250 samples, with standard deviations of 1–3% in the estimated quantiles.



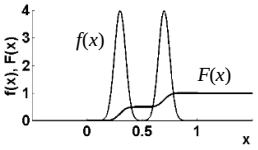
(a) Uniform distribution, range: [0, 1].



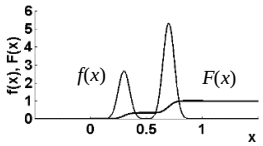
(b) Gaussian distribution, $\mu = 0.5$, $\sigma = 0.125$.



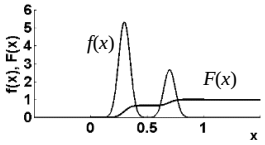
(c) Exponential distribution, $\mu = 1/6$.



(d) Mixture of two Gaussian distributions, $k_1 = k_2 = 0.5$, $\mu_1 = 0.3$, $\mu_2 = 0.7$, $\sigma_1 = \sigma_2 = 0.05$.



(e) Mixture of two Gaussian distributions, $k_1 = 1/3$, $k_2 = 2/3$, $\mu_1 = 0.3$, $\mu_2 = 0.7$, $\sigma_1 = \sigma_2 = 0.05$.



(f) Mixture of two Gaussian distributions, $k_1 = 2/3$, $k_2 = 1/3$, $\mu_1 = 0.3$, $\mu_2 = 0.7$, $\sigma_1 = \sigma_2 = 0.05$.

Figure B.4 Plots of probability density function $f(x)$ and cumulative distribution function $F(x)$ for the synthetic stationary data sequences with range of [0, 1], with $f(x)$ as light trace and $F(x)$ as dark trace.

B.3.2 Synthetic Non-stationary Data

Random number sequences with uniform distribution and the following time-varying parameters were generated:

a) Range changed from [0, 1] to [0.25, 0.75] and back to [0, 1], i.e., $\mu = 1/2$, and σ changed from $1/\sqrt{12}$ to $1/2\sqrt{12}$ and back to $1/\sqrt{12}$.

b) Range changed from [0, 0.5] to [0.5, 1] and back to [0, 0.5], i.e., μ changed from 1/4 to 3/4 and back to 1/4, and $\sigma = 1/2\sqrt{12}$.

B.3.3 Real Data

Broadband audio noise sampled at 10 kHz and amplitude modulated, with a scale factor changed from 1 to 2 and back to 1, was used for testing the dynamic response of the quantile tracking. Babble noise (from AURORA database [93]) and speech signal (three vowels and a sentence recorded from a male speaker), with the sampling frequency of 10 kHz, were used as two other examples of real data. The babble noise and the speech signal are non-stationary data which may be treated as stationary for 20 – 30 ms segments, i.e., 200 – 300 samples. The three test sequences were normalized to have the same root-mean-square value.

B.4 Results

For applying the proposed quantile tracking techniques on different types of test data and comparing the performances with some of the earlier techniques, we need to select the processing parameters. For DQTRE, we need to select the values of the step-size control factor λ, rise-time control coefficient α, and fall-time control coefficient β. For convergence

of the estimate to the sample quantile, $0 < \lambda < 1/(MR)$, where M is the peak of the probability density function. For a balanced trade-off between variance and adaptivity, we should select $\lambda \approx 1/L$, where L is the sequence length for which the data may be considered as stationary. For the test data described in the previous section, $1/(MR)$ was less than $1/6$ and L was found to be 200–300 samples. Hence we selected $\lambda = 1/256$. For range estimation, α was selected as 0.1 for one-sample rise of 90%, and β was correspondingly selected as $((1-\alpha)^{1/L}) \approx 0.9996$. For DQTRSE, we need to also select the probability bracketing interval p_b and the segment-limiting fraction l. DQTRSE with $p_b = 1$ is the same as DQTRE. A low value of p_b is needed to reduce the variance of the estimate in case of data with peaky distributions. Empirical investigations using the test data with the distributions as shown in Figure B.4 showed that the combination of $p_b = 1/10$ and $l = 1/4$ provided an acceptable trade-off between the requirements of variance and adaptivity and hence these values may be used for most of the distributions.

The quantile values estimated by the DQTRE and DQTRSE techniques were compared with the sample quantiles (SQ) as the reference values and also with those obtained using the P2, SSA, and EWSA techniques, after implementing them in accordance with their descriptions in [109], [110], and [108], respectively, and using the following parameters:

P2: Number of initialization samples = 5;

SSA: Number of initialization samples = 25, $\mu = 0.01$, $\nu = 1$, and $b(n) = h(n) = n^{-1/5}$;

EWSA: Number of initialization samples = 10, $e = 0.01$;

DQTRE: $\lambda = 1/256$, $\alpha = 0.1$, and $\beta = 0.9996$;

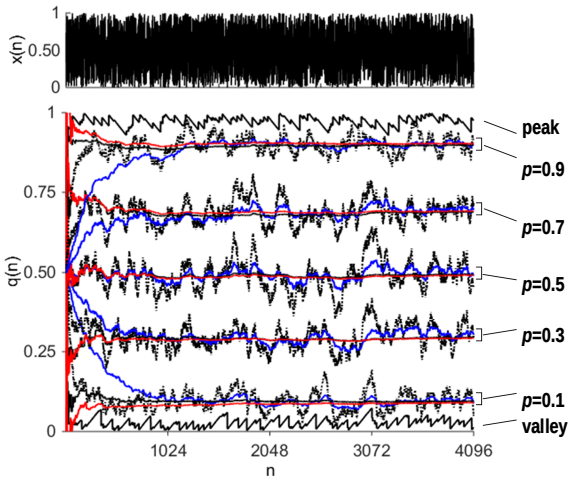
DQTRSE: $\lambda = 1/256$, α=0.1, $\beta = 0.9996$, $p_b = 0.1$, and $l = 0.25$.

The same parameters were used for all types of test data.

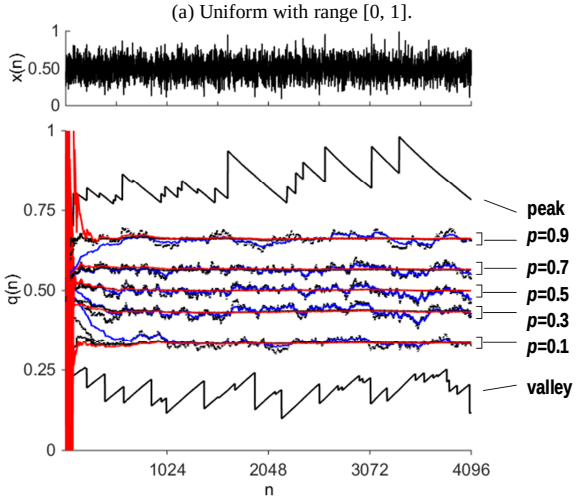
B.4.1 Results for Synthetic Stationary Data

For the data with different distributions as shown in Figure B.4, the sample-by-sample quantile estimates were obtained for p = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 using P2, SSA, EWSA, DQTRE, and DQTRSE and compared with the corresponding SQ values. The SQ values stabilized after about 250 samples. Considering all the test p values, the quantiles estimated by all the techniques converged after about 1000 samples.

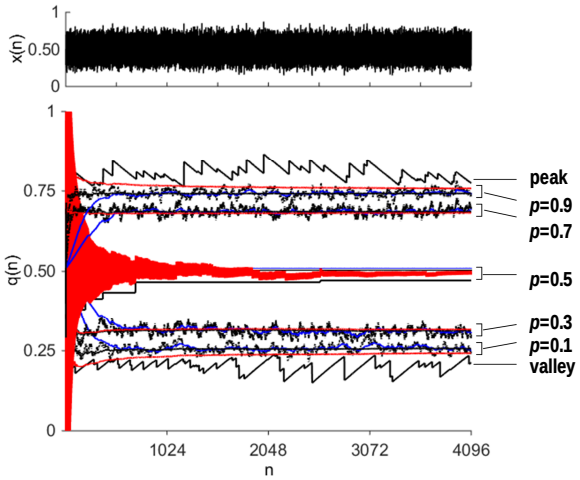
Plots of the estimated quantiles, for some of the p values, of the sequence with uniform distribution (Figure B.4a) are shown in Figure B.5a and those for the Gaussian distribution



(a) Uniform with range [0, 1].



(b) Gaussian with μ = 0. 5 and σ = 0.125



(c) Gaussian mixture with $k_1 = k_2 = 0.5$, $\mu_1 = 0.3$, $\mu_2 = 0.7$, $\sigma_1 = \sigma_2 = 0.05$.

Figure B.5 Quantile tracking for synthetic stationary data with different distributions: (a) uniform, (b) Gaussian, and (c) Gaussian mixture. SQ: solid black trace, SSA: red trace, EWSA: dotted black trace, DQTRE: blue trace.

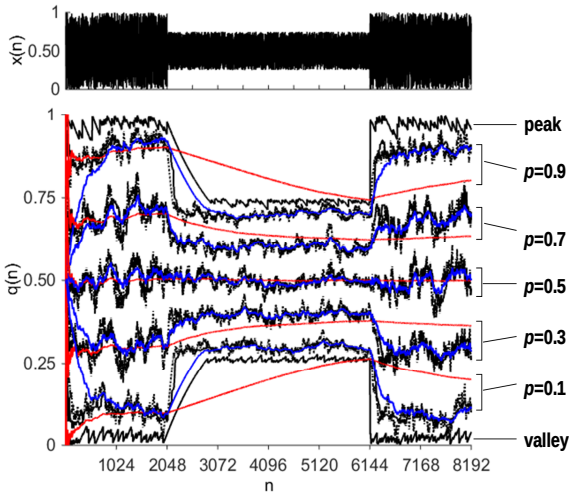
Table B.3 Bias ε (% of range) and standard deviation σ (% of range) in quantile estimation of synthetic stationary data (number of randomizations = 100).

Technique	Distribution											
	Uniform (Fig. B.4a)		Gaussian (Fig. B.4b)		Exponential (Fig. B.4c)		Gaussian Mixture 1 (Fig. B.4d)		Gaussian Mixture 2 (Fig. B.4e)		Gaussian Mixture 3 (Fig. B.4f)	
	ε	σ	ε	σ	ε	σ	ε	σ	ε	σ	ε	σ
P2	0.13	0.14	0.16	0.79	0.06	0.21	3.81	6.64	6.33	6.13	7.47	6.84
SSA	0.33	0.10	0.56	0.13	0.88	0.41	3.86	1.44	3.20	1.46	4.58	1.88
EWSA	0.74	4.19	0.17	1.54	0.50	2.79	4.46	7.41	6.99	7.38	4.55	6.35
DQTRE	0.27	1.95	0.11	1.06	0.42	1.59	4.18	1.00	5.06	1.49	1.95	1.31
DQTRSE	0.25	1.91	0.10	0.85	0.19	1.38	4.20	1.06	5.14	1.48	1.96	1.39

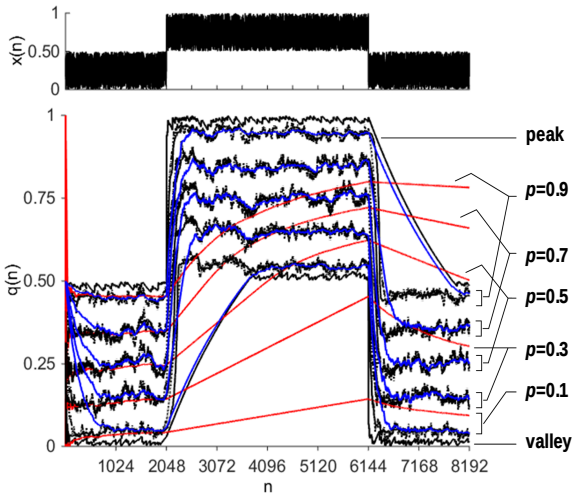
(Figure B.4b) are shown in Figure B.5b. Plots for P2 and DQTRSE were almost identical to those for SSA and DQTRE, respectively, and hence are not included in the figure. In all cases, the estimates converge to the corresponding SQ values. The ripples in the estimates from P2 and SSA after convergence were lower than those from the other three techniques. Ripples were highest for EWSA. For other distributions (Figure B.4c–Figure B.4f) also, P2 and SSA gave lowest ripples. But the biases in the estimates from P2 and SSA were comparable to those from the other techniques. It is particularly visible in the low-density segments in the plots of the quantile estimates for Gaussian mixture distribution with equal weights in Figure B.5c.

The SQ values of a sequence remain unchanged if the samples in the sequence are re-ordered. Hence it is desirable that the estimated quantiles do not exhibit a significant variability with re-ordering of the samples. Randomized re-ordering of the samples in each of the synthetic data sequences was used to obtain 100 scrambled sequences and quantiles were estimated using P2, SSA, EWSA, DQTRE, and DQTRSE. Three error indicators were calculated: (i) bias (absolute value of the mean error with reference to the SQ values), (ii) standard deviation, and (iii) peak-to-peak ripple in the estimate (difference of maximum and minimum values). The peak-to-peak ripple was found to be generally less than 6 times the corresponding standard deviation. For a comparison, the maximum bias and the maximum standard deviation of the quantile estimates for p values in the range of 0.1 – 0.9 are shown in Table B3.

For uniform, Gaussian, and exponential distributions, all techniques give biases well below 1%. The P2 and SSA techniques have lowest standard deviations and EWSA has highest standard deviations. Considering bias and standard deviation both, the performance of the techniques can be rank ordered as P2, SSA, DQTRSE, DTQRE, EWSA. For Gaussian mixtures, the bias in all the estimations increases, with the degradation being more



(a) Range changed from [0, 1] to [0.25, 0.75] and back to [0, 1].



(b) Range changed from [0, 0.5] to [0.5, 1] and back to [0, 0.5].

Figure B.6 Quantile tracking for synthetic dynamic data with uniform distribution: (a) range changed from [0, 1] to [0.25, 0.75] and back to [0, 1] and (b) range changed from [0, 0.5] to [0.5, 1] and back to [0, 0.5]. SQ: solid black trace, SSA: red trace, EWSA: dotted black trace, DQTRE: blue trace.

pronounced in case of P2, SSA, and EWSA. For these distributions, the standard deviations of DQTRE and DQTRSE are almost the same and generally lower than those of the other techniques. Thus the results show that the estimations obtained by the proposed techniques are relatively less affected by the distribution type.

B.4.2 Results for Synthetic Non-stationary Data

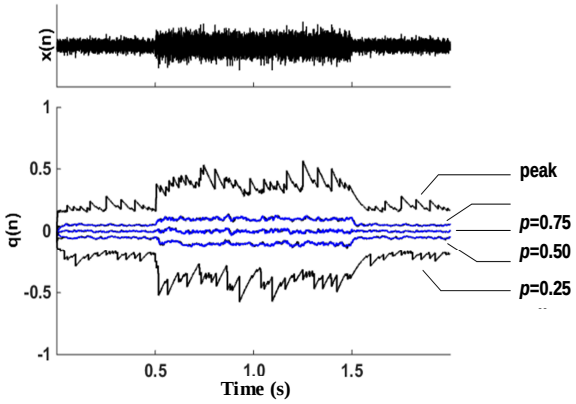
The dynamic tracking of the quantiles was examined for the test data consisting of random number sequences with uniform distribution and pulsed changes in the range. The results for the range changed from [0, 1] to [0.25, 0.75] and back to [0, 1] are shown in Figure B.6a and

those for the range changed from [0, 0.5] to [0.5, 1] and back to [0, 0.5] are shown in Figure B.6b. The SQ values at each sample were obtained by sorting the preceding 256 samples. The plots for P2, not shown in the figure, were almost identical to those for SSA. The plots for DQTRSE, not shown in the figure, were almost similar to those for DQTRE except that they exhibited slower convergence and smaller ripples. The results showed that P2 and SSA are not usable for tracking the quantiles of non-stationary data. As compared with DQTRE and DQTRSE, EWSA has faster adaptivity to changes, but it has larger ripples.

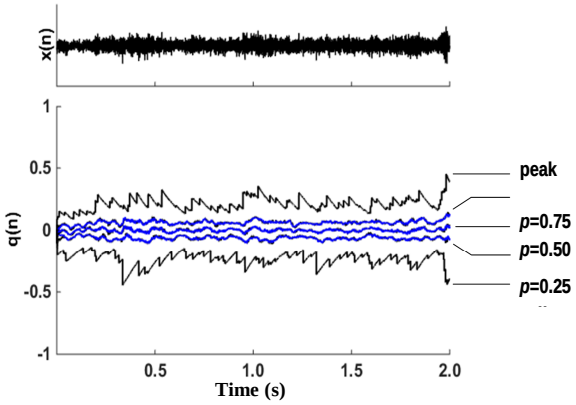
The P2 and SSA techniques use a weight inversely proportional to the sample number. This feature leads to low ripples for stationary data but decreasing adaptivity in case of non-stationary data. Due to the requirement of keeping track of the sample number, they can be used only for sequences of finite length. Their use may be extended for dynamic quantile tracking by segmenting the input sequence by a moving window and carrying out the quantile estimation for each segment. This window-based processing provides quantile estimates decimated at the rate of window shifting and adds to the computational and data buffering requirements. As EWSA, DQTRE, and DQTRSE do not keep a track of the sample number, they can be used for quantile tracking without any restriction on the sequence length.

B.4.3 Results for Real Data

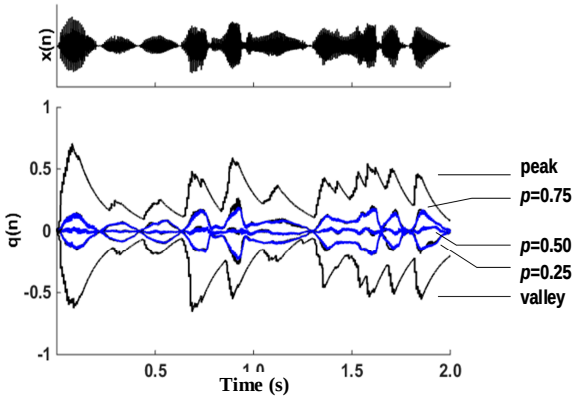
Sample-by-sample tracking of the quantiles was carried out using SQ, EWSA, DQTRE, and DQTRSE for the three types of real data as described earlier. Plots of the SQ and DQTRE estimates for p of 0.25, 0.5, and 0.75 are given in Figure B.7 and these show satisfactory quantile tracking. Plots for EWSA and DQTRSE have significant overlap with DQTRE and these are not shown in the figure. Root-mean-square error (RMSE) of the estimate with reference to SQ values as a percentage of the root-mean-square value of the sequence was calculated as an indicator of the error in tracking. These values are given in Table 4. For broadband noise and babble noise, the RMSE is highest for EWSA for all p values. RMSE values for DQTRE and DQTRSE are comparable and are lower than those for EWSA. For the data corresponding to the speech signal, having faster changes than the data corresponding to the broadband and babble noise, the RMSE of DQTRSE for $p = 0.25$ is higher than that of EWSA. It indicates a slower adaptivity of DQTRSE. A similar result is seen for $p = 0.75$. For all three data, DQTRSE generally gives lowest RMSE at $p = 0.5$, indicating its suitability for tracking quantiles of data near the peaks of the distribution. Considering all p values, DQTRE performs better than EWSA and DQTRSE.



(a) White noise with pulsed change in amplitude



(b) Babble noise



(c) Speech signal

Figure B.7 Quantile tracking using SQ and DQTRE for real data: (a) white noise with pulsed change in amplitude, (b) babble noise, and (c) speech signal. SQ: black trace, DQTRE: blue trace.**Figure B.4**

Table B.4 RMSE of sample-by-sample quantile tracking of real data (number of data samples = 20,000).

Technique	RMSE (% of the RMS value)								
	Test data: White noise with pulsed amplitude			Test data: Babble noise			Test data: Speech signal		
	p = 0.25	p = 0.50	p = 0.75	p = 0.25	p = 0.50	p = 0.75	p = 0.25	p = 0.50	p = 0.75
EWSA	12.0	12.2	12.2	14.7	15.1	16.5	17.1	13.8	18.5
DQTRE	6.6	6.5	6.4	8.4	7.6	8.6	12.6	7.5	11.5
DQTRSE	7.3	4.6	5.3	9.6	8.5	10.7	22.5	6.4	22.1

B.5 Conclusion

A technique for dynamic tracking of quantiles with low memory and computational requirements for use in applications involving real-time estimation of quantiles of a data stream has been presented. The quantile is estimated recursively by applying an increment, selected as a fraction of the estimated range, such that the estimated quantile converges to the sample quantile. It is suitable for online tracking of multiple quantiles with an upper bound on deviation from monotonicity. The technique has been tested using synthetic stationary data with several symmetric and asymmetric density functions, synthetic non-stationary data with time-varying mean and standard deviation, and real data streams with different distributions. It has been compared with some of the techniques reported earlier for quantile tracking of data streams. As compared with low-variance techniques such as P2 and SSA, it has low memory and computation requirements. As it does not keep a track of the sample number, it is suitable for sample-by-sample or window-based tracking of quantiles of non-stationary data and can be used without any restriction on the sequence length. As compared to technique with fast adaptivity such as EWSA, it gives much lower variance during stationary segments and an acceptable adaptivity during transitions.

The proposed DQTRE technique has a step-size control factor for a trade-off between variance and adaptivity of the estimation. For a more controlled trade-off, the DQTRSE technique uses a two-stage dynamic tracking with the step size as a fraction of a range segment bracketing the quantile. For data streams with peaky distributions, DQTRSE may be preferable over DQTRE.

Due to its low memory and computational requirements, the proposed technique can be used for real-time quantile tracking of multiple variables using a single processor. As the technique does not need additional resources for initialization, it is suited for implementation on a microcontroller, DSP chip, FPGA, or ASIC. We have used DQTRE for quantile-based noise estimation for real-time processing of speech signal using spectral subtraction for

suppression of background noise, involving dynamic quantile tracking of 256 spectral samples, using a low-power fixed-point DSP chip for use in hearing aids [115]. Other applications of the proposed techniques for use of order statistics in signal processing and control applications need to be investigated.

Appendix C

IMPLEMENTATION OF DIGITAL HEARING AID AS A SMARTPHONE APPLICATION

C.1 Introduction

Hearing aids are designed using ASICs (application specific integrated circuits) due to power and size constraints. Therefore, incorporation of a new processing technique in hearing aids and its field evaluation is prohibitively expensive. Use of smartphone-based application software (app) to customize and remotely configure settings on hearing aids provide greater flexibility to hearing aid users and developers. Many hearing aid manufacturers (GN ReSound, Phonak, Unitron, Siemens, etc) provide apps to control hearing aids using an Android or iOS smartphone. This type of app helps the hearing aid user in personalizing the listening experience by adjustment of settings during use of the device and avoids repeated visits to an audiology clinic. The smartphone-based apps may also be used for development and testing of signal processing techniques for hearing aids. Hearing aid apps (e.g. 'Petralex', 'uSound', 'Q+', and 'BioAid' for Android/iOS, 'Mimi', 'EnhancedEars' for iOS, and "Hearing Aid with Replay" and "Ear Assist" for Android) [165]–[170] provide users with moderate sensorineural hearing loss a low-cost alternative for hearing aids. In addition to providing frequency-selective gain and multiband dynamic range compression, they also offer the flexibility of creating and storing sound profiles specific to the user's hearing loss characteristics. However, they do not allow users to set the processing parameters in an interactive and real-time mode.

Ambrose et al. [171] have described a single in-ear audio coupling that can be used with a hearing aid and other devices (e.g. smartphone, MP3 player). A combination of the in-ear audio coupling with the smartphone performs the function of hearing aid. The speech input from the microphone of the smartphone is processed by the processor of the smartphone and the processed output is given to the in-ear audio coupling to serve as a hearing aid. The software application on the smartphone allows setting of the hearing loss profile. Neumann et al. [172] have described a device with two software modules for outputting a hearing loss compensated signal. The first module either routes the audio signal to the output of the device for normal hearing listeners or routes the audio signal to the input of the second software module. The second module processes the audio signal for hearing loss compensation. The processing parameters are input to the second module through a GUI (graphical user interface) or through a server connected through the internet.

Rader et al. [173] have described a personal communication device comprising a transmitter/receiver coupled to a communication medium for transmitting/receiving audio signals, control circuitry that controls transmission/reception and processing of call and audio signals, a speaker, and a microphone. The control circuitry uses a hearing loss profile or preferred hearing profile of the user for processing the audio signals. The hearing profile may be obtained from a remote server or through the user interface of the device. The device also has a provision for hearing test. Lang et al. [174] have described a device for increasing the intelligibility of speech signals in mobile communication, wherein the acoustic parameters of the speech are modified in the frequency domain to conform to the listener's hearing profile, which may be selected from a menu of predetermined profiles or may be entered through the user interface.

Camp [175] has described a device with a processor with software for conducting a hearing test to determine hearing profile of the listener, process the audio signals in accordance with the hearing profile, and output the processed signals through an earphone. Mouline [176] has described a device for processing the audio to compensate for frequency-dependent hearing loss, with a facility for storing the hearing loss profiles. Foo and Hughes [177] have described a method for updating a hearing loss profile stored in a hearing aid through a data link between the hearing aid and a hearing aid profile service. Westermann et al. [178] have described a system for managing hearing aid with the hearing loss profile set through the internet. Westergaard and Maretti [179] have described a method of personalizing a hearing aid by setting the processing parameters in accordance with the audiogram input from a server and further fine-tuning by an audiologist.

Thus, several devices have been reported for realizing hearing aids to compensate for the frequency-dependent hearing profile of the listener. These devices do not provide suppression of the background noise, which severely degrades the speech perception by listeners with sensorineural hearing impairment and does not permit setting of the processing parameters by the listener in an interactive and real-time mode. There is, therefore, a need to mitigate the disadvantages associated with the existing devices, by devising a hearing aid with processing for suppressing the background noise and a real-time interactive user interface for setting the processing parameters. An implementation of a smartphone app with signal processing for speech enhancement using an adaptive dynamic quantile tracking based noise estimation, as presented in the fourth chapter, and sliding-band dynamic range compression, as presented in the third chapter, is presented here. Signal processing for speech enhancement and dynamic range compression is described in Section C.2. Implementation of the hearing aid app is

described in Section C.3. The test results are presented in Section C.4, followed by the conclusion in the last section.

C.3 Signal Processing

The implementation provides signal processing for (i) speech enhancement using adaptive dynamic quantile tracking based noise estimation and (ii) sliding-band dynamic range compression, using a DFT-based analysis-synthesis. The two processing techniques are described briefly in the following subsections.

C.3.1 Speech enhancement using adaptive dynamic quantile tracking based noise estimation

Single-channel speech enhancement along with the spectral subtraction based on geometric approach (GA) [90], as presented in Section 4.5 of the fourth chapter, is used for suppression of background noise. The processing comprises windowing, FFT calculation, magnitude spectrum calculation, noise spectrum estimation, SNR-dependent gain calculation, enhanced complex spectrum calculation, IFFT calculation, and resynthesis using overlap-add.

The adaptive dynamic quantile tracking technique, as presented in Section 4.3.2 of the fourth chapter is used for noise estimation. It involves estimation of a quantile function for each spectral sample, by dynamically tracking multiple quantiles for a set of evenly spaced probabilities. Each quantile is updated recursively, without storage and sorting of past spectral samples, using the dynamic quantile tracking technique, as presented in Section 4.2 of the fourth chapter. The adaptive quantile representing the noise is obtained by finding the quantile where the quantile function has the lowest slope, which approximately corresponds to the peak of the probability density function of the noisy signal. The quantile for lowest slope is located as the quantile at which the difference between adjacent quantiles is minimum. The processing parameters for noise suppression are set as $\lambda = 1/256$ (fixed value in place of the adaptive λ), $\tau_p = \tau_v = 0.1$ and $\sigma_p = \sigma_v = (0.9)^{1/1024}$. The quantile function is estimated by tracking eight quantiles corresponding to p as 0.25, 0.30, 0.35,...., and 0.65. These p values are used for locating the adaptive quantile, because of the observation that a quantile corresponding to a lower p resulted in significant underestimation and that to a higher p resulted in significant overestimation.

C.3.2 Sliding-band dynamic range compression

Sliding-band dynamic range compression, as presented in Section 3.2 of the third chapter, is used to compensate for increased hearing thresholds and reduced dynamic range. It comprises the steps of short-time spectral analysis, frequency and level dependent spectral modification,

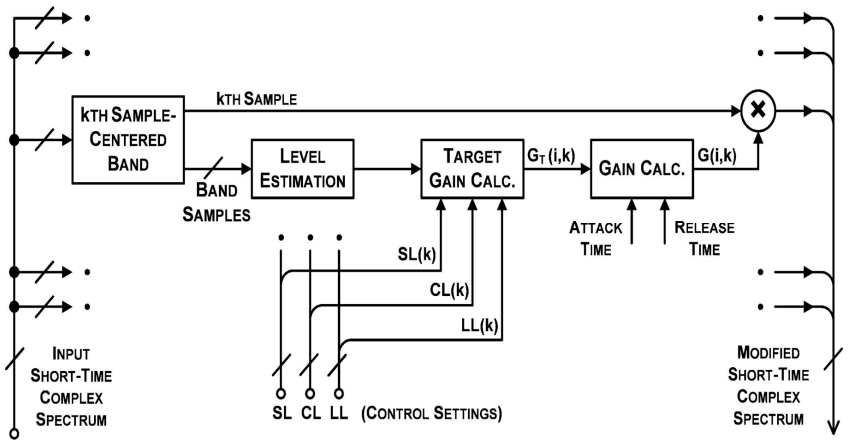


Figure C.1 Spectral modification for compensation of increased hearing thresholds and decreased dynamic range using sliding-band dynamic range compression.

and signal resynthesis. Block diagram of the spectral modification is shown in Figure C.1. It uses a frequency-dependent gain function calculated dynamically from the short-time power spectrum of the signal. The gain for each spectral sample is calculated based on the short-time power in a band centered at its frequency. The bandwidth is selected as auditory critical bandwidth. The time-varying power in the band is used to calculate a target gain for its center frequency. The gain applied to the *k*th spectral sample in the *n*th frame is obtained using the target gain and the values of attack and release times.

In the app-based implementation of sliding-band compression, the target gain is calculated on the basis of a compression function using the desired levels for 'soft', 'comfortable', and 'loud' sounds (referred to as SL, CL, LL, respectively). These levels are obtained as user inputs through the graphical user interface of the app. The compression function, with a piecewise linear three-segment relation between input level $P_{IdB}(n, k)$ and the output level $P_{OdB}(n, k)$ on a dB scale is shown in Figure C.2. It is specified by the values of $P_{OdBSL}(k)$, $P_{OdBCL}(k)$, and $P_{OdBLL}(k)$, which are the output signal levels corresponding to soft, comfortable, and loud sounds, respectively, for the hearing aid user and by the values of $P_{IdBSL}(k)$ and $P_{IdBLL}(k)$, which are the input signal levels corresponding to soft and loud sounds, respectively, for a normal-hearing listener. The relationship is defined in three regions with the compression ratio as 'CR = 1', 'CR > 1', and 'CR = ∞' in the first, second, and third region respectively. With $G_{LdB}(k) = P_{OdBSL}(k) - P_{IdBSL}(k)$, the compression ratio $\mathrm{CR}(k)$ in the second region is given as

$$\mathrm{CR}(k) = \frac{P_{IdBLL}(k) - P_{OdBCL}(k) + G_{LdB}(k)}{P_{OdBLL}(k) - P_{OdBCL}(k)} \tag{C.1}$$

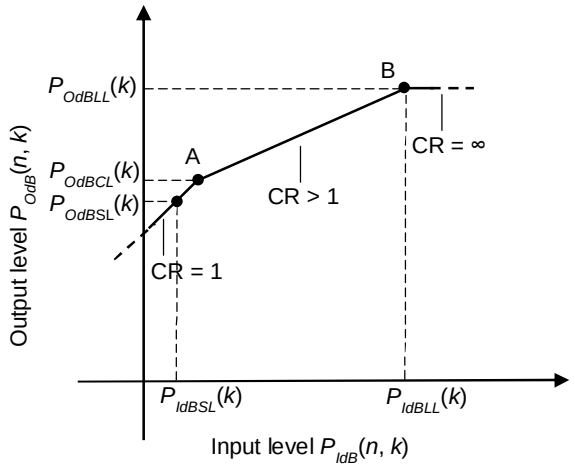


Figure C.2 Compression function relating the output level (dB) and input level (dB) and for *n*th frame and band centered at *k*th spectral sample.

The compression function used in this app is similar to the two-point linear compression (2PLC) of the proposed prescriptive procedure for mixed loss in Section A.11 of Appendix A. The 2PLC function is a three-segment linear relation, specified by the input and output levels for two points A and B, corresponding to the compression threshold and the output-limiting threshold, respectively, which are obtained from the audiometric data. In the current implementation, the compression parameters may be specified based on supra-threshold labeling of the levels as 'soft', 'comfortable', and 'loud' sounds. The loud sound corresponds to point B, with a fixed input level and settable output level. The comfortable level corresponds to point A, with a settable output level. The input level for point A is not explicitly set. The gain for the linear segment below point A is obtained from the settable output level and fixed input level for soft sounds.

The target gain for the *k*th spectral sample in the *n*th frame in the three regions is given as

$$G_{TdB}(n,k) = \begin{cases} G_{LdB}(k), & P_{IdB}(n,k) < P_{OdBCL}(k) - G_{LdB}(k) \\ \dfrac{G_{LdB}(k) - \{P_{IdB}(n,k) - P_{OdBCL}(k)\}\{\mathrm{CR}(k) - 1\}}{\mathrm{CR}(k)}, & P_{OdBCL}(k) - G_{LdB}(k) \le P_{IdB}(n,k) \le P_{IdBLL}(k) \\ P_{OdBLL}(k) - P_{IdB}(n,k), & P_{IdB}(n,k) > P_{IdBLL}(k) \end{cases} \quad \text{(C.2)}$$

The gain is obtained using G_{TdB} and the attack and release times set as 5 ms and 75 ms, respectively, as in (3.2) of the third Chapter.

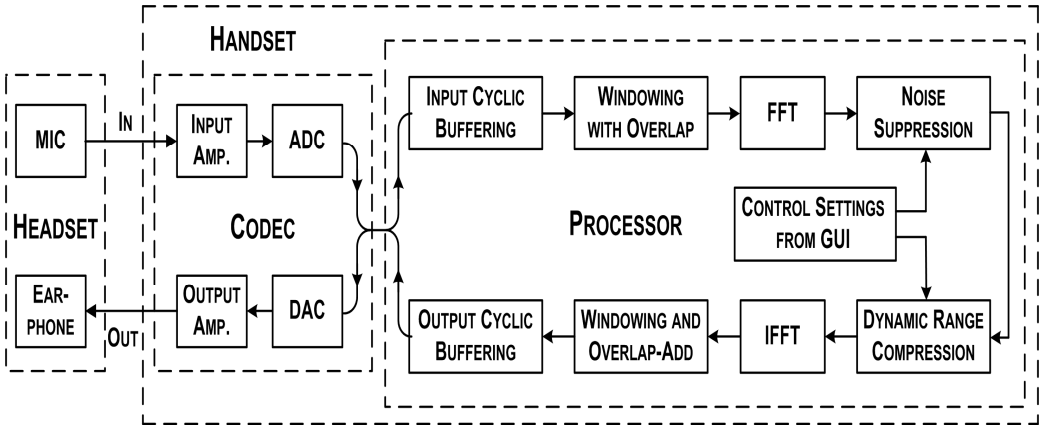


Figure C.3 Implementation of hearing aid app with noise suppression and dynamic range compensation.

C.3 App Implementation

The smartphone app for real-time processing has been developed and tested using 'Nexus 5X' with Android 7.1 Nougat OS due to its relatively small audio I/O delay and high processing capability. It has a touch-controlled graphical interface, enabling the user to adjust the processing parameters in an interactive and real-time mode. In addition to the facility for setting the processing parameters for dynamic range compression, there is a provision for the parameters for additional processing blocks of the future versions.

Figure C.3 shows a block diagram of the implementation of the hearing aid app. The setup comprises a handset with its headset. The headset consists of a microphone and a pair of earphones with associated wires and switching. The handset consists of the codec, the processor, and the touch screen for the user interface. The input signal acquired from the microphone of the headset is amplified and is converted to digital samples by ADC of the codec. These samples are buffered and processed by the processor. The resulting samples are output through DAC of the codec and amplified. The resulting signal is output through the earphones of the headset. The input samples acquired in an S-word buffer and the previous samples stored in a $3S$-word buffer form the L-sample input window for FFT-based analysis-synthesis. The processing for noise suppression and dynamic range compression is carried out using K-point FFT of the input window. The K-point IFFT of the modified complex spectrum is calculated and the output signal is re-synthesized using overlap-add. The analysis-synthesis uses 20-ms frames with 75% frame overlap and 1024-point FFT. The processing is carried out using sampling frequency = 24 kHz, L = 480, S = 120, and K = 1024. The dynamic range of

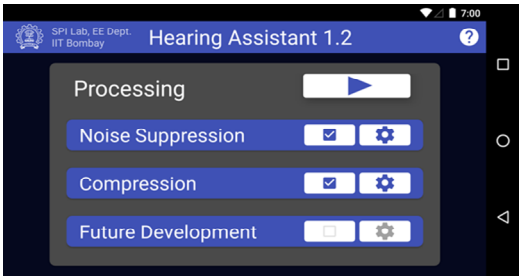


Figure C.4 Screenshot of the home screen of the app.

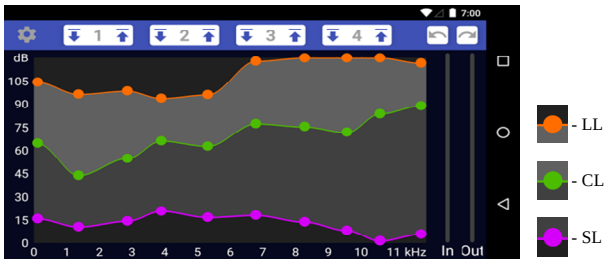


Figure C.5 Screenshot of the settings screen for sliding-band dynamic range compression.

normal hearing is taken as 120 dB in the implementation. The program was written using a combination of C++ and Java, with Android Studio 2.3.0 as the development environment.

The screenshot of the home screen of the app is shown in Figure C.4. The play/stop button is for control of the output. All processing modules have individual on/off and 'settings' buttons. The on/off button can be used for toggling the processing and the settings button can be used for setting the processing parameters graphically. Figure C.5 shows a screenshot of the 'settings' screen for dynamic range compression module with graphical controls for the SL, CL, and LL values. The UI consists of three touch-controlled curves to set the values of SL, CL, and LL across frequencies. Control points called as thumbs are provided to adjust the curves. Each curve consists of 10 thumbs. Provision is provided to store and retrieve up to 4 parameter settings. The UI also consists of undo and redo button to access recent thumb movements. The implementation enables the user to adjust the processing parameters in an interactive and real-time mode, to save them as one of the profiles, or to select the most appropriate profile from the saved ones.

C.4 Test Results

The processing modules were tested on the handset model 'Nexus 5X' with Android 7.1 OS. The evaluation was carried out using the headset of the handset for speech input through its microphone and audio output through its earphone. Informal listening was used for subjective

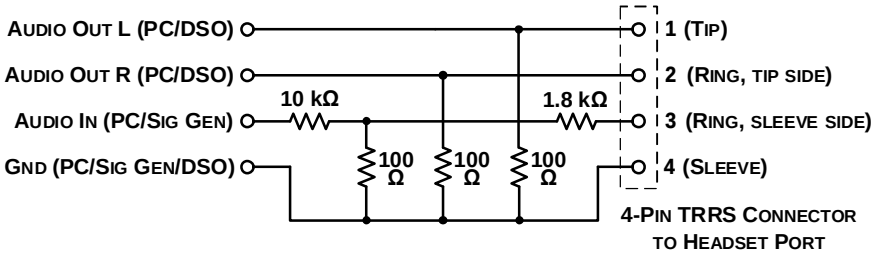


Figure C.6 Audio interface to the 4-pin TRRS headset port of the mobile handset.

evaluation. The experimental set-up comprised the smartphone handset, a 4-pin TRRS connector with an audio interface to the headset port of the handset, a notebook PC with sound card for generating the test input signals, and the sound card 'Focusrite Scarlett 2i2' interfaced to the notebook PC over the USB port to acquire the processed output. This set-up for presentation and acquisition of the signals was used to reduce the noise due to ground loops and other pickups. The audio interface to the 4-pin TRRS headset port of the mobile handset is shown in Figure C.6. It has a resistive attenuator for attenuating the input audio signal to a level compatible with the microphone signal level and output resistance of 1.8 kΩ for it to be recognized as an external microphone. The two output channels have 100 Ω load resistances. The input audio signal can be from a PC sound card or function generator.

The total audio latency was measured by applying a 1 kHz tone burst of 200 ms from a function generator as the input and observing the delay from onset of the input tone burst to the corresponding onset in the output, using a digital storage oscilloscope. The time taken for computation per frame was measured using 'Android device monitor', a profiling tool for Android OS. The test results for two processing modules are given in the following subsections.

C.4.1 Results for noise suppression

Informal listening and objective evaluation using the perceptual evaluation of speech quality (PESQ) measure [94] were used for evaluation of the noise suppression module. The processing was tested using the 30 sentences from NOIZEUS database [92] added with noises from AURORA database [93]. Airport, babble, car, street, and station noises from AURORA database and white noise were added at SNR of 15, 12, 9, 6, 3, and 0, and –3 dB to form noisy speech. Informal listening indicated no audible roughness or musical noise in the processed outputs. Table C.1 shows the PESQ improvement for different noises for 0, 3, and 6 dB SNR. It can be seen that the improvements in PESQ scores were in the range 0.17−0.35. Improvements were highest for the airport noise and lowest for the car noise.

Table C.1 PESQ scores for unprocessed speech and improvement in scores by noise suppression for different types of noises and SNRs (test material: 30 sentences from NOIZEUS database [92]).

Airport noise					Babble noise				
	Unproc. Score		Proc. Impr.			Unproc. Score		Proc. Impr.	
SNR	Mean	S. D.	Mean	S. D.	SNR	Mean	S. D.	Mean	S. D.
6 dB	2.21	0.15	0.29	0.16	6 dB	1.96	0.13	0.26	0.14
3 dB	2.01	0.17	0.33	0.16	3 dB	1.78	0.15	0.23	0.20
0 dB	1.81	0.18	0.35	0.18	0 dB	1.61	0.19	0.17	0.24
Car noise					Street noise				
	Unproc. Score		Proc. Impr.			Unproc. Score		Proc. Impr.	
SNR	Mean	S. D.	Mean	S. D.	SNR	Mean	S. D.	Mean	S. D.
6 dB	2.28	0.15	0.20	0.13	6 dB	2.28	0.15	0.27	0.15
3 dB	2.09	0.16	0.22	0.15	3 dB	2.08	0.17	0.30	0.16
0 dB	1.90	0.17	0.21	0.18	0 dB	1.86	0.19	0.35	0.17
Station noise					White noise				
	Unproc. Score		Proc. Impr.			Unproc. Score		Proc. Impr.	
SNR	Mean	S. D.	Mean	S. D.	SNR	Mean	S. D.	Mean	S. D.
6 dB	2.52	0.14	0.22	0.14	6 dB	1.89	0.15	0.32	0.26
3 dB	2.33	0.15	0.27	0.14	3 dB	1.73	0.18	0.26	0.24
0 dB	1.92	0.19	0.34	0.17	0 dB	1.57	0.19	0.18	0.26

C.4.2 Results for dynamic range compression

Informal listening was used for evaluation of the dynamic range compression module. An example of dynamic range compression with amplitude-modulated input is shown in Figure C.7. Input is an amplitude-modulated tone of 1 kHz and processing parameters are set as shown in Figure C.7 (b) with a compression ratio of 2. The processing gives higher gains at lower values of the input level. Spikes in the amplitude envelope of the output signal in response to step changes in the amplitude envelope of the input signal, as seen in the figure, are typical of the dynamic range compression with a finite frame shift and can be eliminated by using one-sample frame shift but with a significantly increased computational load.

The app was further tested for speech modulated with different types of amplitude envelopes. An example of the processing is shown in Figure C.8, for an amplitude modulated concatenation of speech signals. The input consists of three isolated vowels, a Hindi sentence, and an English sentence, (-/a/-/i/-/u/-“aaiye aap ka naam kya hai?” – “where were you a year ago?”). Informal listening showed that the processing, for speech, music, and environmental sounds with large level variation as inputs, resulted in outputs with the desired compression and without perceptible distortions.

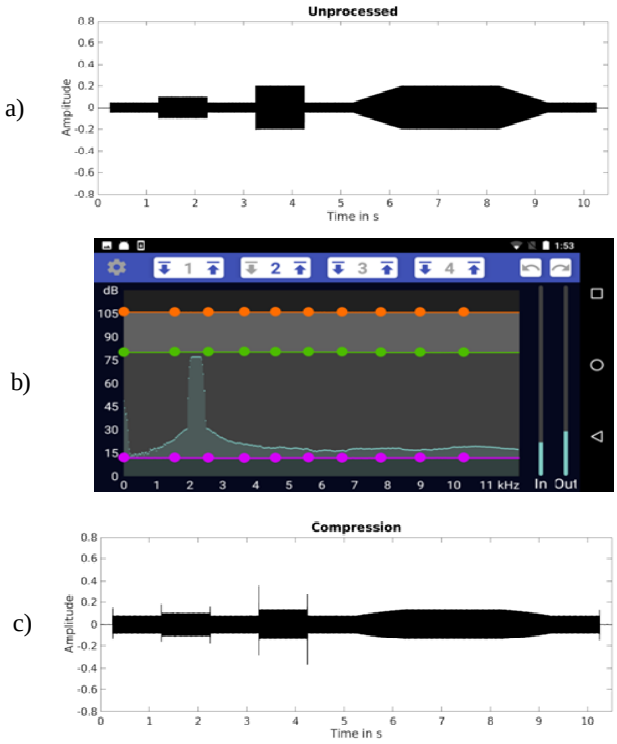


Figure C.7 Example of processing for dynamic range compression: (a) input signal of amplitude modulated tone of 1 kHz, (b) GUI parameters set for constant gain of 12 dB and compression ratio of 2, (c) processed output.

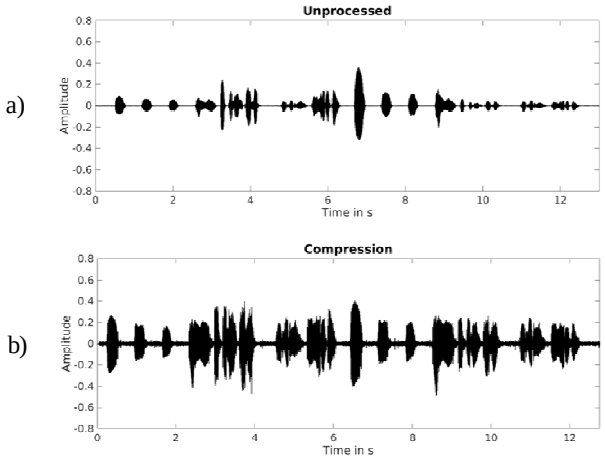


Figure C.8 Example of processing for dynamic range compression: (a) amplitude modulated speech and (b) processed speech with parameters as shown in Figure C.5.

C.4.3 Results for noise suppression and dynamic range compression

Informal listening showed that the outputs of the real-time processing and offline processing for different combinations of noise, SNR, and the modulating envelope of the speech signal,

were almost similar. The PESQ scores for the output of real-time processing with the output of offline processing as the reference were 4.4 or higher, showing that real-time processing did not introduce signal degradation with reference to offline processing.

C.4.4 Audio latency

The total audio latency of the application (signal delay comprising the algorithmic delay and input-output delay) was found to be approximately 45 ms. The algorithmic delay due to 20 ms frame length with 75% overlap corresponds to 25 ms (1.25 times the frame length). The additional delay is due to audio input-output latency of the handset hardware, buffering operations in the OS, and delays in the anti-aliasing and smoothening filters. The average time taken for computation per frame was measured as the difference of the average CPU time taken per frame shift with and without the processing module. The average CPU time taken per frame shift was found to be 1.3 ms for the compression module and 1.1 ms for the speech enhancement module. Thus the implementation required less than 50% of the processor capacity.

C.5 Conclusion

To enable the use of smartphone as a hearing aid, integration of the signal processing for dynamic quantile tracking based noise suppression and sliding-band dynamic range compression has been implemented using 'LG Nexus 5X' running 'Android 7.1'. The processing parameters can be set by the user in an interactive and real-time mode using a graphical touch interface. The audio latency of the app is 45 ms, which is much less than the detectability threshold of 125 ms for audio-visual delay [180] and hence may be considered as acceptable for a hearing aid during face-to-face conversation. In the current version of the app, the processing for speech enhancement using adaptive dynamic quantile tracking based noise estimation uses a fixed convergence factor. It needs be modified for incorporating the speech presence probability dependent convergence factor for quantile tracking, as described in Section 4.3.2 of the fourth chapter, and needs to be evaluated. Implementation of the app on other smartphones and its use by a large number of hearing-impaired listeners is needed for real-life evaluation and further enhancement.

Left blank

REFERENCES

[1] H. Levitt, J. M. Pickett, and R. A. Houde, Eds., *Senosry Aids for the Hearing Impaired*, New York: John Wiley & Sons, 1980, pp. 1–16.

[2] F. H. Silverman, *Speech, Language, and Hearing Disorders*, Needham, MA: Allyn & Bacon, 1995, pp. 14–38.

[3] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 6th ed. Leiden, The Netherlands: Brill, 2013, pp. 57–131.

[4] S. A. Gelfand, *Hearing*: *An Introduction to Psychological and Physiological Acoustics*, 5th ed. London, UK: Informa, 2010, pp. 51–102.

[5] H. Dillon, *Hearing Aids*, 2nd ed. Sydney, Australia: Boomerang, 2012.

[6] R. E. Sandlin, Ed., *Textbook of Hearing Aid Amplification*, 2nd ed. San Diego, CA: Singular, 2000, pp. 210–246.

[7] T. A. Ricketts, "Digital hearing aids: Current "state-of-the-art"," *The ASHA Leader*, vol. 6, no. 14, pp. 8–11, 2001, doi: 10.1044/leader.FTR2.06142001.8

[8] G. R. Popelka, B. C. J. Moore, R. R. Fay, and A. N. Popper, Eds., *Hearing Aids*, Basel, Switzerland: Springer, 2016, pp. 1–20.

[9] M. A. Stone, B. C. J. Moore, J. I. Alcántara, and B. R. Glasberg, "Comparison of different forms of compression using wearable digital hearing aids," *J. Acoust. Soc. Am.*, vol. 106, no. 6, pp. 3603–3619, 1999, doi: 10.1121/1.428213

[10] R. P. Lippmann, L. D. Braida, and N. I. Durlach, "Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss," *J. Acoust. Soc. Am.*, vol. 69, no. 2, pp. 524–534, 1981, doi: 10.1121/1.385375

[11] W. A. Dreschler, D. Eberhardt, and P. W. Melk, "The use of single-channel compression for the improvement of speech intelligibility," *Scandinavian Audiol.*, vol. 13, no. 4, pp. 231–236, 1984, doi: 10.3109/01050398409042131

[12] W. A. Dreschler, "Dynamic range reduction by peak clipping or compression and its effects on phoneme perception in hearing-impaired listeners," *Scandinavian Audiol.*, vol. 17, no. 1, pp. 45–51, 1988, doi: 10.3109/01050398809042179

[13] V. Stahl, A. Fisher, and R. Bipus, "Quantile based noise estimation for spectral subtraction and Wiener filtering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2010)*, Istanbul, Turkey, 2000, pp. 1875–1878, doi: 10.1109/ICASSP.2000.862122

[14] N. W. Evans and J. S. Mason, "Time-frequency quantile-based noise estimation," in *Proc. 11th Eur. Signal Process. Conf. (EUSIPCO 2002)*, Toulouse, France, 2002, pp. 539–542. [online]. Available: kom.aau.dk/group/04gr740/filer/Papers/Time-Frequency%20Quantile-Based%20Noise%20Estimation.pdf

[15] H. Bai and E. A. Wan, "Two-pass quantile based noise spectrum estimation," Center of Spoken Language Understanding, OGI School of Science and Engineering at OHSU, Hillsboro, Oregon, 2003. [online]. Available: citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.12.3528&rep=rep1&type=pdf

[16] V. K. Mai, D. Pastor, A. Aïssa-El-Bey, and R. Le-Bidan, "Robust estimation of non-stationary noise power spectrum for speech enhancement," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 4, pp. 670–682, 2015, doi: 10.1109/TASLP.2015.2401426

[17] E. M. Danaher, M. P. Wilson, and J. M. Pickett, "Backward and forward masking in listeners with severe sensorineural hearing loss," *Audiology*, vol. 17, no. 4, pp. 324–338, 1978, doi: 10.3109/00206097809101302

[18] L. L. Elliott, "Temporal and masking phenomena in persons with sensorineural hearing loss," *Audiology*, vol. 14, no. 4, pp. 336−353, 1975, doi: 10.3109/00206097509071748

[19] A. R. Jayan and P. C. Pandey, "Automated modification of consonant-vowel ratio of stops for improving speech intelligibility," *Int. J. Speech Technol.*, vol. 18, pp. 113–130, 2015, doi: 10.1007/s10772-014-9254-4

[20] T. G. Thomas, "Experimental evaluation of improvement in speech perception with consonant intensity and duration modification," Ph.D. thesis, Dept. Elect. Eng., IIT Bombay, Mumbai, India, 1996.

[21] R. M. Uchanski, S. S. Choi, L. D. Braida, C. M. Reed, and N. I. Durlach, "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," *J. Speech Hear. Res.*, vol. 39, no. 3, pp. 494–509, 1996, doi: 10.1044/jshr.3903.494

[22] N. E. Vaughan, I. Furukawa, N. Balasingam, M. Mortz, and S. A. Fausti, “Time-expanded speech and speech recognition in older adults,” *J. Rehabil. Res. Dev.*, vol. 39, no. 5, pp. 559–566, 2002. [online]. Available: www.rehab.research.va.gov/jour/02/39/5/pdf/Vaughan.pdf

[23] 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, doi: 10.3109/14992027.2011.642012

[24] P. N. Kulkarni, P. C. Pandey, and D. S. Jangamashetti, “Study of perceptual balance for designing comb filters for binaural dichotic presentation,” in *Proc. 20th Int. Congr. Acoust. (ICA 2010)*, Sydney, Australia, 2010, paper no. 556. [online]. Available: www.acoustics.asn.au/conference_proceedings/ICA2010/cdrom-ICA2010/papers/p556.pdf

[25] J. Yang, F. Luo, and A. Nehorai, “Spectral contrast enhancement: Algorithms and comparisons,” *Speech Commun.*, vol. 39, no. 1–2, pp. 33–46, 2003, doi: 10.1016/S0167-6393(02)00057-2

[26] Z. Ribic, J. Yang, and M. Latzel, “Adaptive spectral contrast enhancement based on masking effect for the hearing impaired,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 1996)*, Atlanta, GA, 1996, pp. 937–940, doi: 10.1109/ICASSP.1996.543276

[27] T. Arai, K. Yasu, and N. Hodoshima, “Effective speech processing for various impaired listeners,” in *Proc. 18th Int. Congr. Acoust. (ICA 2004)*, Kyoto, Japan, 2004, pp. 1389–1392. [online]. Available: splab.net/papers/2004/2004_09.pdf

[28] P. N. Kulkarni, P. C. Pandey, and D. S. Jangamashetti, “Multiband frequency compression for improving speech perception by listeners with moderate sensorineural hearing loss,” *Speech Commun.*, vol. 54, no. 3, pp. 341–350, 2012, doi: 10.1016/j.specom.2011.09.005

[29] S. Banerjee, *The Compression Handbook: An Overview of the Characteristics and Applications of Compression Amplification*, 3rd ed. Eden Prairie, MN: Starkey, 2011, pp. 1–60. [online]. Available: uk.starkeypro.com/pdfs/quicktips/Compression_Handbook.pdf

[30] J. Galaster, “Voice iQ - Multivariate benefits,” Technical paper, Starkey Hearing Technologies, Eden Prairie, MN, 2014. [online]. Available: starkeypro.com/pdfs/technical-papers/Voice_iQ-Multivariate_Benefits.pdf

[31] “ReSound Alera,” Product brochure, GN ReSound, Ballerup, Denmark. Accessed: Feb. 16, 2019. [online]. Available: www.gnhearing.se/~/media/DownloadLibrary/ReSound/Products/Alera/alera-product-overview.ashx

[32] “ReSound NoiseTracker II,” Technical white paper, GN ReSound, Ballerup, Denmark. Accessed: Feb. 16, 2019. [online]. Available: resoundpro.com/~/media/REFRESH/US/00-DOWNLOADS/Technology/noise-tracker-II-white-paper.ashx?la=en-US

[33] “Core processing: Dual-path compression system achieves best clarity, comfort and sound quality,” Technical paper, Phonak AG, Stäfa, Switzerland. Accessed: May 6, 2016. [online]. Available: www.phonakpro.com/content/dam/phonak/b2b/C_M_tools/Library/background_stories/en/BGS_CORE_processing_210x280_GB.pdf

[34] “Phonak Ambra: Product information,” Product brochure, Phonak AG, Stäfa, Switzerland. Accessed: May 6, 2016. [online]. Available: www.phonakpro.com/content/dam/phonakpro/gc_hq/en/products_solutions/hearing_aid/ambra/documents/product_information_ambra_027-0479.pdf

[35] “Beltone Legend,” Product brochure, Beltone, Glenview, Illinois. Accessed: May 6, 2016. [online]. Available: www.beltone.se/~/media/DownloadLibrary/Beltone/Beltone,-sp-,Product,-sp-,Page,-sp-,Downloads/Beltone,-sp-,Legend,-sp-,End,-sp-,User,-sp-,Brochure.ashx

[36] “Beltone Hearing Aids,” Product catalog, Beltone, Glenview, Illinois. Accessed: May 6, 2016. [online]. Available: www.beltone-hearing.com/en/hearing-solutions/hearing-aids#boostmax

[37] *Beltone Technology White Paper Series: 2010*, Beltone, Glenview, Illinois. 2010. Accessed: May 6, 2016. [online]. Available: www.beltone-hearing.com/~/media/DownloadLibrary/Beltone/Products/Beltone,-sp-,True/White-paper-M200476-GB.ashx?la=en

[38] *Specification of Hearing Aid Characteristics*, ANSI Standard S3.22-2003, American National Standards Institute, New York, 2003. [online]. Available: law.resource.org/pub/us/cfr/ibr/002/ansi.s3.22.2003.pdf

[39] M. C. Killion, “Compression: Distinctions,” *The Hearing Review*, vol. 3, no. 8, pp. 29–32, 1996. [online]. Available: www.etymotic.com/media/publications/erl-0032-1996.pdf

[40] L. D. Braida, N. I. Durlach, R. P. Lippmann, B. L. Hicks, W. M. Rabinowitz, and C. M. Reed, “Hearing aids–A review of past research on linear amplification, amplitude compression, and

frequency lowering," *J. Speech Hear. Disorders Suppl.*, ASHA Monograph 19, pp. 1−114, 1979. [online]. Available: www.asha.org/uploadedFiles/Monographs19.pdf

[41] B. C. J. Moore, "Psychoacoustics of cochlear hearing impairment and the design of hearing aids," in *Proc. 16th Int. Congr. Acoust. (ICA 1998)*, Seattle, Washington, pp. 2105−2108, 1998. [online]. Available: pdfs.semanticscholar.org/45c2/17f7bbe71c51eae112e10037fbad5aa428e6.pdf

[42] A. B. King and M. C. Martin, "Is AGC beneficial in hearing aids?," *Brit. J. Audiol.*, vol. 18, no. 1, pp. 31–38, 1984, doi: 10.3109/03005368409078926

[43] K. T. Boike and P. E. Souza, "Effect of compression ratio on speech recognition and speech-quality ratings with wide dynamic range compression amplification," *J. Speech Lang. Hear. Res.*, vol. 43, no. 2, pp. 456–468, 2000, doi: 10.1044/jslhr.4302.456

[44] R. M. Cox, G. C. Alexander, C. Gilmore, and K. M. Pusakulich, "Use of the connected speech test (CST) with hearing-impaired listeners," *Ear Hear.*, vol. 9, no. 4, pp. 198–207, 1988, doi: 10.1097/00003446-198808000-00005

[45] H. Dillon, "Compression? Yes, but for low or high frequencies, for low or high intensities, and with what response times?," *Ear Hear.*, vol. 17, no. 4, pp. 287–307, 1996. [online]. Available: journals.lww.com/ear-hearing/Citation/1996/08000/Tutorial_Compression__Yes,_But_for_Low_or_High.1.aspx

[46] P. E. Souza, "Effects of compression on speech acoustics, intelligibility, and sound quality," *Trends Amplif.*, vol. 6, no. 4, pp. 131–165, 2002, doi: 10.1177/108471380200600402

[47] J. M. Kates, "Principles of digital dynamic-range compression," *Trends Amplif.*, vol. 9, no. 2, pp. 45–76, 2005, doi: 10.1177/108471380500900202

[48] J. M. Kates, "Dynamic-range compression using digital frequency warping," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput. (ACSSC 2003)*, Pacific Grove, CA, 2003, pp. 715–719, doi: 10.1109/ACSSC.2003.1292007

[49] E. Villchur, "Signal processing to improve speech intelligibility in perceptive deafness," *J. Acoust. Soc. Am.*, vol. 53, no. 6, pp. 1646–1657, 1973, doi: 10.1121/1.1913514

[50] L. F. Shi and K. A. Doherthy, "Subjective and objective effects of fast and slow compression on the perception of reverberant speech in listeners with hearing loss," *J. Speech Lang. Hear. Res.*, vol. 51, no. 5, pp. 1328–1340, 2008, doi: 10.1044/1092-4388(2008/07-0196)

[51] J. M. Kates, "Understanding compression: Modeling the effects of dynamic-range compression in hearing aids," *Int. J. Audiol.*, vol. 49, no. 6, pp. 395–409, 2010, doi: 10.3109/14992020903426256

[52] P. E. Souza, L. Jenstad, and R. Folino, "Using multichannel wide-dynamic range compression in severely hearing-impaired listeners: Effects on speech recognition and quality," *Ear Hear.*, vol. 26, no. 2, pp. 120–131, 2005, doi: 10.1097/00003446-200504000-00002

[53] K. H. Arehart, J. M. Kates, and M. C. Anderson, "Effects of noise, nonlinear processing, and linear filtering on perceived speech quality," *Ear Hear.*, vol. 31, no. 3, pp. 420–436, 2010, doi: 10.1097/AUD.0b013e3181d3d4f3

[54] S. Gatehouse, G. Naylor, and C. Elberling, "Linear and nonlinear hearing aid fittings – 1. Patterns of benefit," *Int. J. Audiol.*, vol. 45, no. 3, pp. 130–152, 2006, doi: 10.1080/14992020500429518

[55] B. C. J. Moore, J. I. Alcántara, M. A. Stone, and B. R. Glasberg, "Use of a loudness model for hearing aid fitting II. Hearing aids with multi-channel compression," *Brit. J. Audiol.*, vol. 33, no. 3, pp. 157–170, 1999, doi: 10.3109/03005369909090095

[56] J. C. Tejero-Calado, J. C. Rutledge, and P. B. Nelson, "Preserving spectral contrast in amplitude compression for hearing aids," in *Proc. 23th Annual Conf. Eng. Medicine Biology Soc. (EMBS 2001)*, Istanbul, Turkey, 2001, pp. 1453−1456, doi: 10.1109/IEMBS.2001.1020477

[57] J. C. Rutledge, P. B. Nelson, J. C. Tejero-Calado, J. K. Chang, and R. R. Williams, "Performance of sinusoidal model based amplitude compression in fluctuating noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2010)*, Texas, Dallas, 2010, pp. 189−192, doi: 10.1109/ICASSP.2010.5496053

[58] R. F. Laurence, B. C. J. Moore, and B. R. Glasberg, "A comparison of behind-the-ear high-fidelity linear hearing aids and two-channel compression aids, in the laboratory and in everyday life," *Brit. J. Audiol.*, vol. 17, no. 1, pp. 31–48, 1983, doi: 10.3109/03005368309081480

[59] F. Asano, Y. Suzuki, T. Sone, S. Kakehata, M. Satake, K. Ohyama, T. Kobayashi, and T. Takasaka, "A digital hearing aid that compensates for sensorineural impaired listeners," in *Proc.*

IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 1991), Toronto, Ontario, Canada, 1991, pp. 3625–3628, doi: 10.1109/ICASSP.1991.151059

[60] J. M. Kates and K. H. Arehart, "Multichannel dynamic-range compression using digital frequency warping," *EURASIP J. Appl. Signal Process.*, vol. 18, pp. 3003–3014, 2005, doi: 10.1155/ASP.2005.3003

[61] E. Lindemann, "The continuous frequency dynamic range compressor," in *Proc. 1997 Workshop Appl. Signal Process. Audio Acoust. (WASPAA 1997)*, New Paltz, New York, 1997, pp. 1–4, doi: 10.1109/ASPAA.1997.625580

[62] E. C. Vickers, "Frequency domain multiband dynamics compressor with automatically adjusting frequency band boundary locations," U.S. Patent 8 903 109 B2, Dec. 2, 2014.

[63] A. H. Kamkar-Parsi and M. Bouchard, "Improved noise power spectrum density estimation for binaural hearing aids operating in a diffuse noise field environment," *IEEE Trans. Audio, Speech, Language Process.*, vol. 17, no. 4, pp. 521–533, 2009, doi: 10.1109/TASL.2008.2009017

[64] B. Cornelis, M. Moonen, and J. Wouters, "Performance analysis of multichannel Wiener filter-based noise reduction in hearing aids under second order statistics estimation errors," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 5, pp. 1368–1381, 2011, doi: 10.1109/TASL.2010.2090519

[65] 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, doi: 10.1109/TASSP.1979.1163209

[66] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 1979)*, Washington, DC, 1979, pp. 208–211, doi: 10.1109/ICASSP.1979.1170788

[67] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586–1604, 1979, doi: 10.1109/PROC.1979.11540

[68] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, 1984, doi: 10.1109/TASSP.1984.1164453

[69] H. Hirsch and C. Ehrlicher, "Noise estimation techniques for robust speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 1995)*, Detroit, MI, 1995, pp. 153–156, doi: 10.1109/ICASSP.1995.479387

[70] R. Martin, "Spectral subtraction based on minimum statistics," in *Proc. 6th Eur. Signal Process. Conf. (EUSIPCO 1994)*, Edinburgh, UK, 1994, pp. 1182−1185. [online]. Available: citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.472.3691&rep=rep1&type=pdf

[71] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, 2001, doi: 10.1109/89.928915

[72] G. Doblinger, "Computationally efficient speech enhancement by spectral minima tracking in subbands," *in Proc. EUROSPEECH 1995*, Madrid, Spain, 1995, pp. 1513–1516. [online]. Available: citeseerx.ist.psu.edu/viewdoc/citations;jsessionid=3DC896DD2DED19A3E9445AAD05F144A5?doi=10.1.1.47.4415

[73] P. Lockwood and J. Boudy, "Experiments with a nonlinear spectral subtractor (NSS), hidden Markov models and the projection, for robust speech recognition in cars," *Speech Commun.*, vol. 11, no. 2-3, pp. 215−228, 1992, doi: 10.1016/0167-6393(92)90016-Z

[74] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 13, no. 3, pp. 247–251, 1993, doi: 10.1016/0167-6393(93)90095-3

[75] R. Martin, "Bias compensation methods for minimum statistics noise power spectral density estimation," *Signal Process.*, vol. 86, no. 6, pp. 1215−1229, 2006, doi: 10.1016/j.sigpro.2005.07.037

[76] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 443−445, 1985, doi: 10.1109/TASSP.1985.1164550

[77] D. Malah, R. V. Cox, and A. J. Accardi, "Tracking speech-presence uncertainty to improve speech enhancement in non-stationary noise environments," in *Proc. IEEE Int. Conf. Acoust.,*

Speech, Signal Process. (ICASSP 1999), Phoenix, AZ, 1999, pp. 789–792, doi: 10.1109/ICASSP.1999.759789

[78] I. Cohen, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Process. Letters*, vol. 9, no. 1, pp. 12–15, 2002, doi: 10.1109/97.988717

[79] *IEEE Recommended Practice for Speech Quality Measurements*, IEEE Standard 297, The Institute of Electrical and Electronics Engineers, New York, 1969. [online]. Available: ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7405210

[80] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Process*., vol. 11, no. 5, pp. 466–475, 2003, doi: 10.1109/TSA.2003.811544

[81] S. Rangachari and P. C. Loizou, "A noise-estimation algorithm for highly non-stationary environments," *Speech Commun.*, vol. 48, no. 2, pp. 220–231, 2006, doi: 10.1016/j.specom.2005.08.005

[82] Y. Hu and P. C. Loizou, "Speech enhancement based on wavelet thresholding the multitaper spectrum," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 1, pp. 59–67, 2004, doi: 10.1109/TSA.2003.819949

[83] M. Nilsson, S. Soli, and J. A. Sullivan, "Development of hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.*, vol. 95, no. 2, pp. 1085–1099, 1994, doi: 10.1121/1.408469

[84] R. C. Hendriks, R. Heusdens, and J. Jensen, "MMSE based noise PSD tracking with low complexity," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2010)*, Dallas, TX, 2010, pp. 4266–4269, doi: 10.1109/ICASSP.2010.5495680

[85] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 4, pp. 1383–1393, 2012, doi: 10.1109/TASL.2011.2180896

[86] O. Cappe, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 345−349, 1994, doi: 10.1109/89.279283

[87] P. C. Loizou, *Speech Enhancement: Theory and Practice*, 2nd ed. New York: CRC, 2017. [online]. Available: www.crcpress.com/downloads/K14513/K14513_CD_Files.zip

[88] S. K. Waddi, P. C. Pandey, and N. Tiwari, "Speech enhancement using spectral subtraction and cascaded-median based noise estimation for hearing impaired listeners," in *Proc. 19th Nat. Conf. Commun. (NCC 2013)*, Delhi, India, 2013, paper no. 1569696063, doi: 10.1109/NCC.2013.6487989

[89] W. D. Voiers, A. D. Sharpley, and C. H. Helmsath, "Research on diagnostic evaluation of speech intelligibility," USAF Cambridge Res. Lab., Cambridge, MA, Contract AF19628-70-C-O182, Final Rep., Jan. 1973. [online]. Available: apps.dtic.mil/dtic/tr/fulltext/u2/755918.pdf

[90] Y. Lu and P. C. Loizou, "A geometric approach to spectral subtraction," *Speech Commun.*, vol. 50, no. 6, pp. 453−466, 2008, doi: 10.1016/j.specom.2008.01.003

[91] R. J. MacAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 2, pp. 137–145, 1980, doi: 10.1109/TASSP.1980.1163394

[92] Y. Hu and P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech Commun.*, vol. 49, no. 7–8, pp. 588–601, 2007, doi: 10.1016/j.specom.2006.12.006

[93] H. Hirsch and D. Pearce, "The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *Proc. ISCA Int. Workshop Autom. Speech Recognit. (ITRW ASR 2000)*, Paris, France, 2000, pp. 181–188. [online]. Available: dnt.kr.hsnr.de/aurora/download/asr2000_final_footer.pdf

[94] *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-To-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs*, Rec. ITU-T P.862, International Telecommunications Union, Geneva, Switzerland, 2001. [online]. Available: www.itu.int/rec/T-REC-P.862-200102-I/en

[95] E. Zwicker and E. Terhardt, "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," *J. Acoust. Soc. Am.*, vol. 68, no. 5, pp. 1523–1525, 1980, doi: 10.1121/1.385079

[96] D. W. Griffin and J. S. Lim, “Signal estimation from modified short-time Fourier transform,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 2, pp. 236–243, 1984, doi: 10.1109/TASSP.1984.1164317

[97] E. Zwicker, “Subdivision of the audible frequency range into critical bands (Freqenzgruppen),” *J. Acoust. Soc. Am.*, vol. 33, no. 2, pp. 248, 1961, doi: 10.1121/1.1908630

[98] “TMS320C5515 fixed-point digital signal processor,” Product data sheet, Texas Instruments, Inc., Dallas, TX, 2013. [online]. Available: www.ti.com/lit/ds/sprs645f/sprs645f.pdf

[99] “TMS320C5515 eZdsp USB stick,” Technical reference manual, Spectrum Digital, Inc., Stafford, TX, 2010. [online]. Available: support.spectrumdigital.com/boards/usbstk5515/reva/files/usbstk5515_TechRef_RevA.pdf

[100] “TLV320AIC3204 ultra low power stereo audio codec,” Texas Instruments, Inc., Dallas, TX, 2008. [online]. Available: www.ti.com/lit/ds/symlink/tlv320aic3204.pdf

[101] S. Sharma, N. Tiwari, and P. C. Pandey, “Implementation of a digital hearing aid with user-settable frequency response and sliding-band dynamic range compression as a smartphone app,” in *Proc. Int. Conf. Intell. Human Comp. Interaction (IHCI 2016)*, Pilani, India, 2016, pp. 173–186. [online]. Available: link.springer.com/content/pdf/10.1007/978-3-319-52503-7_14.pdf

[102] S. Sharma, N. Tiwari, and P. C. Pandey, “Implementation of digital hearing aid as a smartphone application,” in *Proc. INTERSPEECH 2018*, Hyderabad, India, 2018, pp. 1175–1179. [online]. Available: www.isca-speech.org/archive/Interspeech_2018/pdfs/2031.pdf

[103] N. Tiwari and P. C. Pandey, “A technique with low memory and computational requirements for dynamic tracking of quantiles,” *J. Signal Process. Syst.*, pp. 1–12, 2018, doi: 10.1007/s11265-017-1327-6

[104] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, 1951. [online]. Available: projecteuclid.org/download/pdf_1/euclid.aoms/1177729586

[105] J. I. Munro and M. S. Paterson, “Selection and sorting with limited storage,” *Theor. Comput. Science*, vol. 12, no. 3, pp. 315–323, 1980, doi: 10.1016/0304-3975(80)90061-4

[106] L. Tierney, "A space-efficient recursive procedure for estimating a quantile of an unknown distribution," *SIAM J. Sci. Statist. Comput.*, vol. 4, no. 4, pp. 706–711, 1983, doi: 10.1137/0904048

[107] E. Möller, G. Grieszbach, B. Shack, and H. White, "Statistical properties and control algorithms of recursive quantile estimators," *Biometrical J.*, vol. 42, no. 6, pp. 729–746, 2006, doi: 10.1002/1521-4036(200010)42:6<729::AID-BIMJ729>3.0.CO;2-W

[108] F. Chen, D. Lambert, and J. C. Pinheiro, “Incremental quantile estimation for massive tracking,” in *Proc. 6th ACM SIGKDD Int. Conf. Knowledge Discovery Data Mining (KDD 2000)*, Boston, MA, 2000, pp. 516–522, doi: 10.1145/347090.347195

[109] R. Jain and I. Chlamtac, “The P^2 algorithm for dynamic calculation of quantiles and histograms without storing observations,” *Commun. ACM Magazine*, vol. 28, no. 10, pp. 1076–1085, 1985. [online]. Available: www.cse.wustl.edu/~jain/papers/ftp/psqr.pdf

[110] A. Amiri and B. Thiam, “A smoothing stochastic algorithm for quantile estimation,” *Statist. Prob. Letters*, vol. 93, pp. 116–125, 2014, doi: 10.1016/j.spl.2014.06.016

[111] M. Cooke, J. Barker, S. Cunningham, and X. Shao, "An audio-visual corpus for speech perception and automatic speech recognition (L)," *J. Acoust. Soc. Am.*, vol. 120, no. 5, pp. 2421–2424, 2006, doi: 10.1121/1.2229005

[112] *Objective Measurement of Active Speech Level*, Rec. ITU-T P.56, International Telecommunications Union, Geneva, Switzerland, 1993. [online]. Available: www.itu.int/rec/T-REC-P.56

[113] T. Gerkmann and R. C. Hendriks, “Noise power spectral density estimation code,” Accessed: Feb 16, 2019. [online]. Available: www.inf.uni-hamburg.de/en/inst/ab/sp/publications/taslp12noisepsd.html

[114] Y. Lu and P. C. Loizou, “Speech enhancement code,” Accessed: Feb 16, 2019. [online]. Available: ecs.utdallas.edu/loizou/speech/GA_code.zip

[115] N. Tiwari and P. C. Pandey, “Speech enhancement using noise estimation based on dynamic quantile tracking for hearing impaired listeners,” in *Proc. Nat. Conf. Commun. 2015 (NCC 2015)*, Mumbai, India, 2015, paper no. 1570056299, doi: 10.1109/NCC.2015.7084849

[116] B. C. J. Moore, “Hearing loss in the elderly and its compensation with hearing aids,” *Gerontechnology*, vol. 1, no. 3, pp. 140–152, 2002, doi: 10.4017/gt.2001.01.03.003.00

[117] L. E. Humes, “Evolution of prescriptive fitting approaches,” *Am. J. Audiol.*, vol. 5, no. 2, pp. 19−23, 1996, doi: 10.1044/1059-0889.0502.19
[118] C. V. Palmer and G. A. Lindley IV, “Overview and rationale for prescriptive formulas for linear and nonlinear hearing aids,” in M. Valante, *Strategies for Selecting and Verifying Hearing Aid Fittings*, Chapter 1, pp. 1–22, Thieme Medical, 2002. [online]. Available: www.thieme.com/media/samples/pubid1013629716.pdf
[119] R. M. Cox, “A structured approach to hearing aid selection,” *Ear Hear.*, vol. 6, no. 5, pp. 226−239, 1985. [online]. Available: www.harlmemphis.org/files/5914/0269/2785/Structured_approach_to_hearing_aid_selection.pdf
[120] H. Davis, C. V. Hudgins, R. J. Marquis, R. H. Nicholas Jr., G. E. Peterson, D. A. Ross, and S. S. Stevens, “The selection of hearing aids,” *Layrngoscope*, vol. 56, no. 3, pp. 135–163, 1946, doi: 10.1002/lary.5540560302
[121] N. A. Watson and V. O. Knudsen, “Selective amplification in hearing aids,” *J. Acoust. Soc. Am.*, vol. 11, no. 4, pp. 406–419, 1940, doi: 10.1121/1.1916053
[122] B. E. Walden, D. M. Schwartz, D. L. Williams, L. L. Holum-Hardegen, and J. M. Crowley, “Test of the assumptions underlying comparative hearing aid evaluations,” *J. Speech Hear. Disord.*, vol. 48, no. 3, pp. 264–273, 1983, doi: 10.1044/jshd.4803.264
[123] S. F. Lybarger, *Simplified Fitting System for Hearing Aids*, Canonsburg, PA, Radio Ear Corp., 1963.
[124] V. O. Knudsen and I. H. Jones, “Artificial aids to hearing,” *Laryngoscope*, vol. 45, no. 1, pp. 48–69, 1935, doi: 10.1288/00005537-193501000-00003
[125] D. Byrne and W. Tonisson, “Selecting the gain of hearing aids for persons with sensorineural hearing impairments,” *Scandinavian Audiol.*, vol. 5, no. 2, pp. 51–59, 1976, doi: 10.3109/01050397609043095
[126] D. Byrne and H. Dillon, “The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid,” *Ear Hear.*, vol. 7, no. 4, pp. 257–265, 1986, doi: 10.1097/00003446-198608000-00007
[127] D. Byrne, A. Parkinson, and P. Newall, “Hearing aid gain and frequency response requirements for the severely/profoundly hearing impaired,” *Ear Hear.*, vol. 11, no. 1, pp. 40–49, 1990, doi: 10.1097/00003446-199002000-00009
[128] B. C. J. Moore, B. R. Glasberg, and M. A. Stone, “Development of a new method for deriving initial fittings for hearing aids with multi-channel compression: CAMEQ2-HF,” *Int. J. Audiol*, vol. 49, no. 3, pp. 216−227, 2010, doi: 10.3109/14992020903296746
[129] G. A. McCandless and P. E. Lyregaard, “Prescription of gain and output (POGO) for hearing aids,” *Hear. Instrum.*, vol. 34, no. 1, pp. 16–21, 1983.
[130] D. M. Schwartz, P. E. Lyregaard, and P. Lundh, “Hearing aid selection for severe to profound hearing loss,” *Hear. J.*, vol. 41, no. 2, pp. 13–17, 1988.
[131] M. C. Killion and S. Fikret-Pasa, “The 3 types of sensorineural hearing loss: Loudness and intelligibility considerations,” *Hear. J.*, vol. 46, no. 11, pp. 31–36, 1993. [online]. Available: www.etymotic.com/media/publications/erl-0025-1993.pdf
[132] B. C. J. Moore, “Use of a loudness model for hearing aid fitting. IV. Fitting hearing aids with multi-channel compression so as to restore ‘normal’ loudness for speech at different levels,” *Brit. J. Audiol.*, vol. 34, no. 3, pp. 165–177, 2000, doi: 10.3109/03005364000000126
[133] D. Byrne, H. Dillon, T. Ching, R. Katsch, and G. Keidser, “NAL-NL1 procedure for fitting nonlinear hearing aids: Characteristics and comparisons with other procedures,” *J. Am. Acad. Audiol.*, vol. 12, no. 1, pp. 37–51, 2001. [online]. Available: www.audiology.org/sites/default/files/journal/JAAA_12_01_04.pdf
[134] G. Keidser, H. Dillon, M. Flax, T. Ching, and S. Brewer, “The NAL-NL2 prescription procedure,” *Audiol. Res.*, vol. 1, no. 24, pp. 88–90, 2011, doi: 10.4081/audiores.2011.e24
[135] I. Shapiro, “Hearing aid fitting by prescription,” *Audiology*, vol. 15, no. 2, pp. 163–173, 1976.
[136] J. B. Allen, J. L. Hall, and P. S. Jeng, “Loudness growth in 1/2-octave bands (LGOB): A procedure for the assessment of loudness,” *J. Acoust. Soc. Am.*, vol. 88, no. 2, pp. 745–753, 1990, doi: 10.1121/1.399778

[137] M. Valente and D. VanVliet, “The Independent Hearing Aid Fitting Forum (IHAFF) protocol,” *Trends Amplif.*, vol. 2, no. 1, pp. 6–35, 1997, doi: 10.1177/108471389700200102
[138] L. E. Cornelisse, R. C. Seewald, and D. G. Jamieson, “The input/output formula: A theoretical approach to the fitting of personal amplification devices,” *J. Acoust. Soc. Am.*, vol. 97, no. 3, pp. 1854–1864, 1995, doi: 10.1121/1.412980
[139] S. D. Scollie, R. C. Seewald, L. E. Cornelisse, S. T. Moodie, M. P. Bagatto, D. Laurnagaray, S. Beaulac, and J. Pumford, “The desired sensation level multistage input/output algorithm,” *Trends Amplif.*, vol. 9, no. 4, pp. 159–197, 2005, doi: 10.1177/108471380500900403
[140] D. Byrne, “Effects of bandwidth and stimulus type on most comfortable loudness levels of hearing-impaired listeners,” *J. Acoust. Soc. Am.*, vol. 80, no. 2, pp. 484–494, 1986, doi: 10.1121/1.394044
[141] R. C. Seewald, M. Ross, and M. K. Spiro, “Selecting amplification characteristics for young hearing-impaired children,” *Ear Hear.*, vol. 6, no. 1, pp. 48–53, 1985. [online]. Available: insights.ovid.com/crossref?an=00003446-198501000-00013
[142] *The Visual Input/Output Locator Algorithm Program (VIOLA)*. (1999), Hearing Aid Research Lab. [online]. Available: www.harlmemphis.org/index.php/clinical-applications/viola/
[143] B. C. J. Moore and B. R. Glasberg, “A model of loudness perception applied to cochlear hearing loss,” *Auditory Neuroscience*, vol. 3, pp. 289–311, 1997. [online]. Available: www.mechanicsofhearing.org/mohdl/pdfs/AN/Moore-Glasberg-AudNeurosci-1997.pdf
[144] *Methods for the Calculation of the Articulation Index*, ANSI S3.5-1969, American National Standards Institute, New York, 1969.
[145] *Methods for the Calculation of the Speech Intelligibility Index*, ANSI S3.5-1997, American National Standards Institute, New York, 1997. [online]. Available: webstore.ansi.org/Standards/ASA/ANSIASAS31997R2017
[146] *NAL-NL2 Prescription Procedure*. (2011), National Acoustic Laboratories. [online]. Available: shop.nal.gov.au/epages/nal.sf/en_AU/?ObjectPath=/Shops/nal/Products/P4641
[147] E. E. Johonson, “Prescriptive amplification recommendations for hearing losses with a conductive component and their impact on the required maximum power output: An update with accompanying clinical explanation an international comparison of long-term average speech spectra,” *J. Am. Acad. Audiol.*, vol. 24, no. 6, pp. 452–460, 2013, doi: 10.3766/jaaa.24.6.2
[148] J. M. Chambers, D. A. James, D. Lambert, and S. V. Wiel, “Monitoring networked applications with incremental quantile estimation,” *Statist. Science*, vol. 21, no. 4, pp. 463–475, 2006, doi: 10.1214/088342306000000583
[149] K. Alsabti, S. Ranka, and V. Singh, “A one-pass algorithm for accurately estimating quantiles for disk-resident data,” in *Proc. 23rd Int. Conf. Very Large Data Bases (VLDB 1997)*, Athens, Greece, 1997, pp. 346–355. [online]. Available: citeseerx.ist.psu.edu/viewdoc/download? doi=10.1.1.96.708&rep=rep1&type=pdf
[150] R. Agrawal and A. Swami, “A one-pass space-efficient algorithm for finding quantiles,” in *Proc. 7th Int. Conf. Management Data (COMAD 1995)*, Pune, India, 1995, pp. 1–16. [online]. Available: citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.42.3631&rep=rep1&type=pdf
[151] G. S. Manku, S. Rajagopalan, and B. G. Lindsay, “Approximate medians and other quantiles in one pass and with limited memory,” in *Proc. 1998 ACM SIGMOD Int. Conf. Management Data (SIGMOD 1998)*, Seattle, WA, 1998, pp. 426–435, doi: 10.1145/276305.276342
[152] C. Ris and S. Dupont, “Assessing local noise level estimation methods: Application to noise robust ASR,” *Speech Commun.*, vol. 34, no. 1-2, pp. 141–158, 2001, doi: 10.1016/S0167-6393(00)00051-0
[153] S. Guha and A. McGregor, “Stream order and order statistics: Quantile estimation in random-order streams,” *SIAM J. Scientific Statist. Comput.*, vol. 38, no. 5, pp. 2044–2059, 2009, doi: 10.1137/07069328X
[154] S. Guha and A. McGregor, “Lower bounds for quantile estimation in random-order and multi-pass streaming,” in *Proc. 34th Int. Colloq. Automata, Languages, Programming (ICALP 2007)*, Wroclaw, Poland, 2007, pp. 704–715, doi: 10.1007/978-3-540-73420-8_61
[155] S. Guha, N. Koudas, and K. Shim, “Data-streams and histograms,” in *Proc. 33rd Annual ACM Symposium on Theory of Computing (STOC 2001)*, Hersonissos, Crete, Greece, 2001, pp. 471–475, doi: 10.1145/380752.380841

[156] G. S. Manku, S. Rajagopalan, and B. G. Lindsay, “Random sampling techniques for space efficient online computation of order statistics of large datasets,” in *Proc. 1999 ACM SIGMOD Int. Conf. Management Data (SIGMOD 1999)*, Philadelphia, PA, 1999, pp. 251–262, doi: 10.1145/304182.304204
[157] E. J. Chen and W. D. Kelton, “Estimating steady-state distribution via simulation-generated histograms,” *Comput. Operations Res.*, vol. 35, no. 4, pp. 1003–1016, 2008, doi: 10.1016/j.cor.2006.05.015
[158] E. J. Chen and W. D. Kelton, “Density estimation from correlated data,” *J. Simul.*, vol. 8, no. 4, pp. 281–292, 2014, doi: 10.1057/jos.2014.7
[159] M. Greenwald and S. Khanna, “Space-efficient online computation of quantile summaries,” in *Proc. 2001 ACM SIGMOD Int. Conf. Management Data (SIGMOD 2001)*, Santa Barbara, CA, 2001, pp. 58–66, doi:10.1145/376284.375670
[160] A. Arasu and G. S. Manku, “Approximate counts and quantiles over sliding windows,” in *Proc. 23rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS 2004)*, Paris, France, 2004, pp. 286–296, doi: 10.1145/1055558.1055598
[161] G. Cormode and S. Muthukrishnan, “An improved data stream summary: The count-min sketch and its applications,” *J. Algorithms*, vol. 55, no. 1, pp. 58–75, 2005, doi: 10.1016/j.jalgor.2003.12.001
[162] G. Cormode, F. Korn, S. Muthukrishnan, and D. Srivastava, “Effective computation of biased quantiles over data streams,” in *Proc. 21st Int. Conf. Data Engineering (ICDE 2005)*, Tokyo, Japan, 2005, pp. 20–31, doi: 10.1109/ICDE.2005.55
[163] L. Wang, G. Leo, K. Yi, and G. Cormode, “Quantiles over data streams: An experimental study,” in *Proc. 2013 ACM SIGMOD Int. Conf. Management Data (SIGMOD 2013)*, New York, 2013, pp. 737–748, doi: 10.1145/2463676.2465312
[164] J. Cao, L. E. Li, A. Chen, and T. Bu, “Incremental tracking of multiple quantiles for network monitoring in cellular networks,” in *Proc. 1st ACM Workshop on Mobile Internet through Cellular Networks (MINCET'09)*, Beijing, China, 2009, pp. 7–12, doi: 10.1145/1614255.1614258
[165] *Petralex Hearing Aid v3.3.1*. (2019), IT4You. [online]. Available: play.google.com/store/apps/details?id=com.it4you.petralex
[166] *Bio Aid v1.0.2*. (2012), Nick Clark. [online]. Available: itunes.apple.com/us/app/bioaid/id577764716?mt=8
[167] *Hearing Aid with Replay (Lite) v2.0.0*. (2014), Lamberg Solutions. [online]. Available: apkpure.com/hearing-aid-with-replay-lite/com.ls.soundamplifier
[168] *uSound (Hearing Assistant) v.3.0.14r*. (2019), S. A. Newbrick. [online]. Available: play.google.com/store/apps/details?id=com.newbrick.usound
[169] *Mimi Hearing Test v2.0.17*. (2018), Mimi Hearing Technologies GmbH. [online]. Available: itunes.apple.com/in/app/mimi-hearingtest/id932496645
[170] *Q+ Hearing Aid v2.0.0*. (2018), Quadio Devices. [online]. Available: play.google.com/store/apps/details?id=com.quadiodevices.qplus&hl=en_IN
[171] S. D. Ambrose, S. P. Gido, and R. B. Schulein, “Hearing device system and method,” U.S. Patent Appl. 20120057734A1, Mar. 8, 2012.
[172] J. Neumann, N. WackNun, M. Rodriguez, N. S. Grange, and J. Kinsbergen, “Consumer electronics device adapted for hearing loss compensation,” U.S. Patent Appl. 20150195661A1, Jul. 5, 2015.
[173] R. S. Rader, C. M. Madison, B. W. Edwards, S. Puria, and B. B. Johansen, “Sound enhancement for mobile phones and others products producing personalized audio for users,” U.S. Patent 7 529 545 B2, May 5, 2009.
[174] H. Lang, S. Jääskcläincn, S. Karjalainen, O. Aaltoncn, T. Kaikuranta, and P. Vuori, “Mobile station with audio signal adaptation to hearing characteristics of the user,” U.S. Patent 6 813 490 B1, Nov. 2, 2004.
[175] W. O. Camp Jr., “Mobile terminals including compensation for hearing impairment and methods and computer program products for operating the same,” U.S. Patent 7 613 314 B2, Nov. 3, 2009.
[176] A. Mouline, “Adaptation of audio data files based on personal hearing profiles,” U.S. Patent Appl. 20020068986A1, Jun. 6, 2002.

[177] E. W. Foo and G. F. Hughes, “Remotely updating a hearing and profile,” U.S. Patent 9 613 028 B2, Apr. 4, 2017.
[178] S. E. Westermann, S. V. Andersen, A. Westergaard, and N. E. B. Maretti, “System and method for managing a customizable configuration in a hearing aid,” Int. Publ. WO 2017/071757 Al, May 4, 2017.
[179] A. Westergaard and N. E. B. Maretti, “System and method for personalizing a hearing aid,” Int. Publ. WO 2017/028876 Al, Feb. 23, 2017.
[180] *Relative Timing of Sound and Vision for Broadcasting*, Rec. ITU-R BT.1359, Radiocommunication Sector of International Telecommunication Union, Geneva, Switzerland, 1998. [online]. Available: www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.1359-0-199802-S!!PDF-E.pdf
[181] N. Tiwari, “Dynamic range compression and noise suppression for use in hearing aids: Processing examples”, SPI Lab, EE Dept., IIT Bombay, Feb. 28, 2019, [online] Available: www.ee.iitb.ac.in/~spilab/material/nitya/thesis_processing_examples_2019feb28

Thesis Related Publications

Journal paper

1. N. Tiwari and P. C. Pandey, “A technique with low memory and computational requirements for dynamic tracking of quantiles,” *J. Signal Process. Systems*, 2018, doi: 10.1007/s11265-017-1327-6

Papers in conference proceedings

1. S. Sharma, N. Tiwari, P. C. Pandey, “Implementation of digital hearing aid as a smartphone application,” in *Proc. INTERSPEECH 2018*, Hyderabad, India, 2018, pp. 1175-1179.
2. N. Tiwari, S. Sharma, and P. C. Pandey, “Speech enhancement using noise estimation with adaptive dynamic quantile tracking for use in hearing aids,” in *Proc. 3DC INTERSPEECH 2017*, Stockholm, Sweden, 2017.
3. N. Tiwari, P. C. Pandey, and A. Sharma, “A smartphone app-based digital hearing aid with sliding-band dynamic range compression,” in *Proc. Nat. Conf. Commun. 2016 (NCC 2016)*, Guwahati, India, paper no. 1570227725.
4. S. Sharma, N. Tiwari, and P. C. Pandey, “Implementation of a digital hearing aid with user-settable frequency response and sliding-band dynamic range compression as a smartphone app,” in *Proc. Int. Conf. Intell. Human Comp. Interaction (IHCI 2016)*, Pilani, India, 2016, pp. 173–186.
5. N. Tiwari and P. C. Pandey, “Speech enhancement using noise estimation based on dynamic quantile tracking for hearing impaired listeners,” in *Proc. Nat. Conf. Commun. 2015 (NCC 2015)*, Mumbai, India, 2015, paper no. 1570056299.
6. N. Tiwari and P. C. Pandey, “A sliding-band dynamic range compression for use in hearing aids,” in *Proc. Nat. Conf. Commun. 2014 (NCC 2014)*, Kanpur, India, 2014, paper no. 1569847357.

Patents and patent applications

1. P. C. Pandey and N. Tiwari, "Method and system for suppressing noise in speech signals in hearing aids and speech communication devices," Indian Patent Application No. 640/MUM/2015, 26 Feb 2015, PCT Application No. PCT/IN2015/000183, 24 Apr 2015, US Patent No. US10032462B2, 24 July 2018.
2. P. C. Pandey and N. Tiwari, “Dynamic range compression with low distortion for use in hearing aids and audio systems,” Indian Patent Application No. 290/MUM/2014, PCT Application No. PCT/IN2015/000049, US Patent No. 9,672,834, 6 June 2017.
3. P. C. Pandey, N. Tiwari, and S. Sharma, "Personal communication device as a hearing aid with real-time interactive user interface," Indian Patent Application No. 201821032763, 31 Aug 2018.

Journal papers under review

1. N. Tiwari and P. C. Pandey, Speech enhancement using noise estimation with dynamic quantile tracking, *IEEE/ACM Trans. Audio, Speech Lang. Process.*, 2019.
2. N. Tiwari and P. C. Pandey, Sliding-band dynamic range compression for use in hearing aids, *Int. J. Speech Technol.*, 2019.

Left blank

Author's Resume

Nitya Tiwari: The author received the B.E. degree in electronics and telecommunication in 2010 from Shri G. S. Institute of Technology and Science, Indore, affiliated to Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, India and the M.Tech. degree in electrical engineering in 2012 from the Indian Institute of Technology Bombay, India, where she is currently a Ph.D. student. Her research interests include speech processing and digital signal processing.

Publications

Papers

1. N. Tiwari and P. C. Pandey, "A technique with low memory and computational requirements for dynamic tracking of quantiles," *J. Signal Processing Systems*, 2018, DOI: 10.1007/s11265-017-1327-6.
2. S. Sharma, N. Tiwari, P. C. Pandey, "Implementation of digital hearing aid as a smartphone application," in *Proc. INTERSPEECH 2018*, Hyderabad, India, 2018, pp. 1175-1179.
3. N. Tiwari, Saketh Sharma, and P. C. Pandey, "Speech enhancement using noise estimation with adaptive dynamic quantile tracking for use in hearing aids," in *Proc. 3DC INTERSPEECH 2017*, Stockholm, Sweden, 2017.
4. N. Tiwari, P. C. Pandey, and A. Sharma, "A smartphone app-based digital hearing aid with sliding-band dynamic range compression," in *Proc. Nat. Conf. Commun. 2016 (NCC 2016)*, Guwahati, India, paper no. 1570227725.
5. S. Sharma, N. Tiwari, and P. C. Pandey, "Implementation of a digital hearing aid with user-settable frequency response and sliding-band dynamic range compression as a smartphone app," in *Proc. Int. Conf. Intell. Human Comp. Interaction (IHCI 2016)*, Pilani, India, 2016, pp. 173–186.
6. N. Tiwari and P. C. Pandey, "Speech enhancement using noise estimation based on dynamic quantile tracking for hearing impaired listeners," in *Proc. Nat. Conf. Commun. 2015 (NCC 2015)*, Mumbai, India, 2015, paper no. 1570056299.
7. N. Tiwari and P. C. Pandey, "A sliding-band dynamic range compression for use in hearing aids," in *Proc. Nat. Conf. Commun. 2014 (NCC 2014)*, Kanpur, India, 2014, paper no. 1569847357.
8. R. Holani, P. C. Pandey, and N. Tiwari, "A JFET-based circuit for realizing a precision and linear floating voltage-controlled resistance," *Proc. IEEE Indicon 2014*, Pune, India, 2014, paper no. 1098.
9. N. Tiwari, S. K. Waddi, and P. C. Pandey, "Speech enhancement and multi-band frequency compression for suppression of noise and intraspeech spectral masking in hearing aids," *Proc. IEEE Indicon 2013*, Mumbai, 2013, paper no. 524.
10. N. Tiwari, P. C. Pandey, and P. N. Kulkarni, "Real-time implementation of multi-band frequency compression for listeners with moderate sensorineural impairment," *Proc. INTERSPEECH 2012*, Portland, Oregon, 2012, paper no. 689.

Patents and patent applications

1. P. C. Pandey and N. Tiwari, "Method and system for suppressing noise in speech signals in hearing aids and speech communication devices," Indian Patent Application No. 640/MUM/2015, 26 Feb 2015, PCT Application No. PCT/IN2015/000183, 24 Apr 2015, US Patent No. US10032462B2, 24 July 2018.
2. P. C. Pandey, A. R. Jayan, and N. Tiwari, "Method and system for consonant-vowel ratio modification for improving speech perception," Indian Patent Application No. 739/MUM/2014, 4 Mar 2014, PCT Application No. PCT/IN2015/000048, 27 Jan 2015, US Patent Application Publication No. US20160365099A1, 15 Dec 2016.
3. P. C. Pandey and N. Tiwari, "Dynamic range compression with low distortion for use in hearing aids and audio systems," Indian Patent Application No. 290/MUM/2014, PCT Application No. PCT/IN2015/000049, US Patent No. 9,672,834, 6 June 2017.
4. P. C. Pandey, N. Tiwari, and S. Sharma, "Personal communication device as a hearing aid with real-time interactive user interface," Indian Patent Application No. 201821032763, 31 Aug 2018.
5. P. C. Pandey, S. Debbarma, V. Marla, N. Tiwari, R. Holani and D. K. Sharma, "Continuously variable precision and linear floating resistor using metal-oxide-semiconductor field-effect transistors," Indian Patent Application No. 201821030404, 13 Aug 2018, PCT Application No. PCT/IN2018/050760, 16 Nov 2018.

Left blank

Acknowledgments

I would like to express my sincere gratitude and respect to my thesis supervisor Prof. P. C. Pandey, for his invaluable guidance, motivation, and support, which have made this work possible. I am thankful to him for encouraging me in taking up new challenges and motivating me. His curiosity, wisdom, and strive for perfection has been a source of inspiration for me. I am deeply thankful to Prof. P. Rao and Prof. V. Rajbabu, members of the research progress committee for their valuable suggestions, encouragement, and constructive criticism at various stages of my research work. I am very thankful to Prof. D. K. Sharma, Prof. V. M. Gadre, Prof. B. G. Fernandes, Prof. G. Kumar, Prof. V. Singh, Prof. A. Karandikar, Prof. A. Kumar, and all the professors who encouraged me to work harder and lifted my spirits whenever I felt low.

I would like to thank all members in the Signal Processing and Instrumentation Laboratory (SPI Lab), EE Dept., IIT Bombay for their support and encouragement. I am grateful to Saketh, Santosh, Nataraj, Dr. A. R. Jayan, Dr. P. N. Kulkarni, and Dr. M. S. Shah for sharing interesting discussions with me and providing support whenever needed. I am thankful to Hirak, Pramod, Prachir, Vikas, Shibam, Yogesh, Susmi, and Rani for the company, discussions, and for creating a joyful work environment. I would like to thank Mr. Vidyadhar Kamble for helping me in all the lab related issues. I would also like to thank my friends Mahima, Vinayak, Naveen, Siva, Pratigya, Gaurav, Sandeep, Mahendra, Shonal, Shoba, Debapratim, Shruti, and Pallavi for their support and encouragement and for making my Ph.D. journey memorable one. I am also thankful Mrs. Madhu, Mrs Vaishali, Mr. Santosh, Mr. Devidas, and to all the members of EE office for their help and support.

I am thankful to all my relatives for their unconditional love, encouragement, and support. I am grateful to my parents and to my in-laws for refilling me with energy and positivity whenever I felt exhausted and anxious. I am thankful to my brother whose jokes relieved my stress. I am indebted to my husband Nataraj who understood me and supported me throughout this journey. Ultimately, I am thankful to God for giving me the opportunity to learn and for being with me.

Nitya Tiwari

June 2019