MACQUARIE UNIVERSITY

DOCTORAL THESIS

Speech referenced dynamic compression limiting

Improving loudness comfort and acoustic safety

Michael John Amiel Fisher

Communications Engineer

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy at the

Department of Linguistics | Faculty of Human Sciences

'Speech referenced dynamic compression limiting:
Improving loudness comfort and acoustic safety'

Department of Linguistics | Faculty of Human Sciences
Macquarie University, Sydney, Australia

Supervisors
Adjunct Professor Harvey Dillon, PhD
Associate Professor Mridula Sharma, PhD
Emeritus Professor Philip Newall

Research conducted at the
National Acoustic Laboratories
Australian Hearing Hub
16 University Avenue
Macquarie University
NSW, 2109, AUSTRALIA

The development and experimental work described in this thesis was financially supported by the HEARing CRC, established and supported under the Australian Government's Co-operative Research Centres Program.

Intellectual property (IP) contained within this thesis relating to the system and method has been assigned to either HEARworks Pty Ltd or HearIP Pty Ltd. For information about licensing this IP contact Michael.Fisher@nal.gov.au.

Declaration

I, Michael John Amiel Fisher, declare that this thesis titled *Speech referenced dynamic compression limiting: Improving loudness comfort and acoustic safety* and the work described herein is entirely my work with the exceptions noted below.

- I gratefully acknowledge the contribution of my colleague at the National Acoustic Laboratories, Dr Nicky Chong-White. Dr Chong-White made a substantial contribution to the research, development and evaluation of the SRL MKI speech dominance detector and the SRL MKII voice dominance detector as well as C coding and objectively testing the SRL MKI scheme. Although the thesis does not go into this work at a detailed level, the speech dominance detection performance, and particularly the voice dominance detection performance, is critical to the practical success of the SRL scheme and Dr Chong-White has spent several years working on this.

- I gratefully acknowledge the contribution of my supervisor, Director of the National Acoustic Laboratories, Adjunct Professor Harvey Dillon, PhD. Professor Dillon has been my sounding board advising me with regards to the methods I have used and the views I have expressed in this thesis.

- I gratefully acknowledge the contribution of my colleague at the National Acoustic Laboratories, Cong Van Nguyen. Mr Nguyen coded the preliminary version for the MATLAB GUI used in the evaluation of the SRL MKI scheme and converted the SRL MKII MATLAB code of Dr Chong-White and myself into C code and verified its performance against the MATLAB code. The thesis does not describe this work other than to mention the code compatibility and high speed at which the scheme operates. This high performance is entirely due to Mr Nguyen's well-structured and efficient coding which will assist in the scheme's practical application.

I confirm that:

- This thesis has been submitted solely to Macquarie University for consideration for a doctoral degree.
- This thesis is less than 80,000 words.
- When I have drawn on the work of others, I have acknowledged this through the citations at the point of use and a reference in the bibliography.
- All the experimental work involving human subjects was approved by Australian Hearing's Human Research Ethics Committee and the approval documents are included in Appendix A of this thesis.

Signed: Date: 22 July 2016

Michael John Amiel Fisher

Abstract

This thesis presents the research, development and evaluation associated with a novel method of improving listening comfort and acoustic safety for people listening primarily to speech produced by electronic devices such as headsets, telephones, headphones, hearing aids, cochlear implants, level-dependent hearing protectors and public address systems.

The method involves a novel technique of amplitude limiting audio signals that convey speech. Time-varying, frequency-specific levels of speech generate a set of time-varying speech reference levels. The method limits the level of the audio to these speech reference levels and hence is called speech referenced limiting (SRL). In principle, SRL provides the greatest limiting of noise for the least limiting of speech, making it arguably the optimal method for limiting noise in speech systems.

Two schemes based on the method were developed, the SRL MKI and the SRL MKII schemes. The latter scheme was far superior, with the ability to estimate the speech loudness and power from frequency regions where speech was dominant, while ignoring frequency regions where it was not. It contained a novel method of determining the amount of additional control needed to correct for the loudness summation of noises with a bandwidth exceeding that of speech, as well as providing fast speech referenced control over the power of abrupt sounds while introducing only a very short delay.

Subjective evaluation of the SRL MKI and SRL MKII schemes conducted in the laboratory confirmed large reductions in noise loudness and preservation of speech quality. It was hypothesised that the SRL MKII scheme would provide the greatest reduction in the excess loudness of an audio signal compared with the loudness of the preceding speech conveyed by the audio signal for the least reduction in the speech loudness and quality. Using stimuli typical of those experienced in the three main intended applications (hearing aids, level-dependent hearing protectors and telephone headsets), this hypothesis held true and noise control was shown to be far superior to a conventional fixed-reference limiter while speech loudness and quality were maintained.

Acknowledgments

This is the fun part where I get the opportunity to thank the wonderful people in my life who put up with me every day, which in itself is something to thank them for.

Firstly, to my wonderful children, Jack and Thomas Fisher, thanks for putting up with the dining room being taken over with mountains of paper, for making me dinners and for enduring my singing and guitar playing between writing every paragraph – the music never ever stops.

To my wonderful friend, Jane McCredie, you are the best editor a boy could ever have. It's submitted, now let's go out!

To my thesis supervisor and boss, Adjunct Professor Harvey Dillon, thank you for giving me a job, for your amazing perception, your care, your fairness and your completely unflappable demeanour. You are an inspiration to me in countless ways. Thank you for all the times you've patiently explained things to me and for your never ending faith in me. The research goes on …

To my principal supervisor, Associate Professor Mridula Sharma, thank you for your positivity, your encouragement, for keeping the bureaucratic wolves at bay and your faith that eventually I will submit. I look forward to wearing that floppy hat with you.

Many thanks to you, Emeritus Professor Philip Newall, for helping to start me off on this journey, for encouraging me along the way, and, of course, for that rather dry sense of humour of yours. I still can't quite adopt your suggestion of naming a scale of acoustic discomfort as 'the Fisher' – but I'll always chuckle when recalling your dead pan suggestion of it.

To my wonderful collaborator, Dr Nicky Chong-White, for joining me for years in a project that sometimes appeared to be a mad and unachievable undertaking. For sticking with me through the hair-brained pursuit of SRL's perfection, for keeping the faith and returning after SRL MKI to make a fabulous MKII voice dominance detector – a huge thank you goes to you.

To Cong-van Nguyen, you are the smartest programmer I've ever met. Thank you for everything you've taught me about programming structure. Your ability to distil one day's complex problem into a solution that can be used again and again has impressed me no end. Thank you for your wonderful coding and for helping me to avoid programming the same thing twice.

To Mark Seeto, hawk-eye, thank you for your precision in scrutinising my work. You're a wonderful outlier amongst both statisticians and audiologists – few could claim to share that great intersection on a Venn diagram with a Masters in both disciplines. Thank you Mark, I think we need to shoot some pool, my shout.

A thank you to all my other colleagues at the National Acoustic Laboratories, past and present, you know who you are, and in particular, John Seymour for being a great coffee drinking partner and making my PC sing along with my brain (post coffee). Thank you also to Scott Brewer for sorting out my other PC issues (and that's not just my Political Correctness), to Greg Stewart for helping me with our mutual friends Mr HATS and Mr KEMAR, to Pam Jackson for always encouraging me and helping me to Savi-om and om, to Vivian Fabricatorian for helping me find the perfect piece each time when completing the NAL puzzle, to Dr Tim Gainsford for helping me with those ethical approvals, to Eunice Sarif for finding me those elusive publications, to Sam Youn for the wonderful music and friendship that kept me going, to Dr Fabrice Bardy, Dr Megan Gilliver, Dr Bram Van Dun, Dr Warwick Williams and Lyndal Carter (Dr to be) for endless discussions, wonderful friendship and distracting papers, to Dr Jorge Mejia for your statistical scepticism and your very cheeky laugh, and Katrina Freeston for your help with subject recruiting and a special thanks to all who participated in my experiments. A big thank you to you all.

Also a thank you goes to all my doctoral stable mates at Macquarie University, past and present, in particular, Dr Tobias Weller, Dr Adam Westermann, Dr Christos Oreinos, Baljeet Rana, Adam Weisser, Javier Badajoz and your inspirational supervisor, Dr Jorg Buchholz, and to Dr Nemat Rouhbakhsh, Dr Isabelle Boisvert, and Jens Cubick (ring-in from DTU, many thanks for your last minute help), and those of you who also gave up your time to participate in my experiments.

A big thank you to my extended family, Megan, Nicholas and Jeremy Jones, Andrew, Vesna, Deua and Jessie Stojanovski, Michelle, Steve, Rebecca and Alexandra Anderson for all your support. I am forever grateful to my parents, my father, the late Professor Bruce Harrigan Fisher and my mother, the late Mary Dawn Fisher for instilling in me the desire to learn and to use my skills and knowledge to help others. I am eternally grateful to my late wife, Sally Frances Nesta Fisher, for sharing with me her wonderfully inspiring and imaginative mind and for her faith in me.

To my eclectic, eccentric, best friend, Dr George Raicevich, you are the greatest information bowerbird I have ever had the pleasure of knowing. If Google tapped your brain it would explode. Thank you for inspiring me to think laterally. As Einstein said, 'The true sign of intelligence is not knowledge but imagination' well, you'd impress him on both counts.

Last but not least is you, Professor Robert Cowan, thank you Bob, for your faith in me and my work, for running the very successful HEARing CRC that has kept me in a job, and for all those cheeky comments. Well Michelangelo has something to say back at last, 'Hey Pope, the ceiling is complete.'

Michael Fisher, July 2016

Dedicated to

the memory of my late wife,

Sally Frances Nesta Fisher (nee Dawes)

1959 – 2007

and to Jack and Thomas Fisher

and Jane McCredie

Contents

List of publications

Publications

Fisher M, Chong-White N and Dillon H (2010). Speech referenced limiting: controlling the loudness of a signal with reference to its speech loudness. In *Audio Engineering Society Convention 129,* Audio Engineering Society.

Fisher M, Chong-White N and Dillon H (2011). Controlling the loudness of a communications signal with reference to its speech loudness. In *Forum Acusticum 2011,* pp. 79-82, European Acoustics Association.

Conference presentations

Fisher M, Chong-White N and Dillon H. Speech Referenced Limiting (SRL): a new approach to limiting the sound level from a hearing aid. Oral presentation at the International Hearing Aid Research Conference. Lake Tahoe, 11-15 August 2010.

Fisher M, Chong-White N and Dillon H. Speech referenced limiting of noise. Oral presentation at the 20th International Congress on Acoustics, ICA 2010. Sydney, 23-27 August 2010.

Fisher M, Chong-White N and Dillon H. Speech referenced limiting: controlling the loudness of a signal with reference to its speech loudness. Convention Paper 22 presented at the 129th Convention of the Audio Engineering Society, San Francisco, 4-7 November 2010.

Fisher M, Chong-White N and Dillon H. Controlling the loudness of a communications signal with reference to its speech loudness. Forum Acusticum 2011, 27 June – 01 July, Aalborg Denmark 2011.

Abbreviations

ALC	Automatic level control
ANOVA	Analysis of variance
ANSI	American National Standards Institute
ASI	Acoustic shock injury
BL	Band-limited
CB	Critical band
CN	Cochlear nucleus
CR	Compression ratio
CT	Compression threshold
dB	Decibel
dB SPL	Sound pressure level in decibels
dBA	A-weighted sound pressure level in decibels
dBr	Decibels relative to a given level e.g. digital saturation
DF	Diffuse field
DFT	Discrete Fourier transform
DRP	Drum reference point
DSP	Digital signal processing / processor
DFSR	Diffuse field saturated response
DTMF	Dual tone modulated frequency
ERB	Equivalent rectangular bandwidth
ERD	Equivalent rectangular duration
ERP	Ear reference point
FFT	Fast Fourier transform
FIR	Finite impulse response
FRL	Fixed-reference limiting / limiter
HATS	Head and torso simulator
HMM	Hidden Markov model
HRTF	Head related transfer function
HTL	Hearing threshold level
Hz	Hertz
I	Intensity
IFFT	Inverse fast Fourier transform
IIR	Infinite impulse response
Io	Reference intensity
ISO	International Standards Organisation
ITU	International Telecommunications Union

kHz	Kilo Hertz
LC	Loudness category
LPC	Linear predictive coding
LSP	Line spectra pair
LT	Limiting threshold
LTL	Long-term loudness
MCL	Most comfortable level
MEM	Middle ear muscle
MPO	Maximum power output
ms	Millisecond
NAL	National Acoustic Laboratories
NN	Neural network
p	Power, exponent, or significance level
PABX	Private automated branch exchange
RE	Real ear
RESR	Real-ear saturated response
RMS	Root mean squared
S	Sensation, specific loudness, or total loudness
s	Second
SDD	Speech dominance detector
SII	Speech intelligibility index
SMN	Stapedius motoneurons
SNR	Signal-to-noise ratio
So	Reference sensation
SPL	Sound pressure level
SpNR	Speech-to-noise ratio
SRC	Sample rate conversion / converter
SRL	Speech referenced limiting / limiter
SRL MKI	Speech referenced limiter mark 1
SRL MKII	Speech referenced limiter mark 2
SSPL	Saturated sound pressure level
STL	Short-term loudness
TC	Time constant
TTMN	Tensor tympani motoneurons
TTTP	Tonic tensor tympani phenomenon
VCN	Ventral cochlear nucleus
WB	Wideband
WDRC	Wide dynamic range compression
WOLA	Weighted overlap-add

Chapter 1

Introduction

1 Introduction

This thesis describes a method of improving listening comfort and acoustic safety for people listening primarily to speech produced by electronic devices such as headsets, telephones, headphones, hearing aids, cochlear implants, level-dependent hearing protectors and public address systems.

The thesis presents a novel method of controlling an audio signal that aims to preserve the quality and the intelligibility of speech conveyed by the audio signal while preventing other conveyed sounds from being louder than this speech.

The original motivation for this research arose from previous work by the author in the area of acoustic shock protection for telephone headset wearers.[1] Telephone headset wearers, particularly those with intensive headset usage (e.g. call centre employees), were occasionally exposed to sudden loud sounds from their headset and this sometimes resulted in acoustic shock injury (ASI).[2] Previous methods of controlling the level of the sound to prevent this occurring were found to be detrimental to speech quality and intelligibility in typical background noise environments.[3] A trade-off between providing good protection and providing good speech quality and intelligibility was required.[4] To improve this situation, novel methods of protection were developed such as shriek rejection, which suppresses a set of known offensive sounds, characterised by a narrow-frequency distribution, such as whistles.[5] However, there were many offensive sounds that a shriek rejecter could not identify and suppress and so conventional methods of sound level control were still required as a failsafe, bringing with them the inherent trade-off between providing comfort and good protection and providing good speech quality and intelligibility.

Another approach to this problem was needed: a solution that preserved the quality and intelligibility of the speech even at high sound levels, while suppressing all sounds that could lead to discomfort or injury, would be a major advance. This necessitated an approach that did not require the identification of potentially harmful sounds in order for them to be suppressed.

It was evident that listeners to electronically reproduced speech with access to a volume control typically adjusted the volume so the speech was at a comfortable loudness.[6] For speech to be at a comfortable loudness, it must by definition be comfortable at all frequencies. If any frequency causes discomfort, the user will most likely reduce the volume and/or alter any tone control available to them, unless doing so would prevent them from understanding the speech against whatever background noise was present.

This research proposed using the loudness of speech at specific frequencies as a control for ensuring that all sounds were comfortable. This method of control has been termed speech referenced limiting (SRL). A control such as this, in theory, would not affect speech, unless high-level noise was simultaneously present within the signal, as the speech would be its reference. It would control only sounds that exceeded this reference. These sounds might be noise alone or a mixture of noise and speech. Such a control system would reduce, at each frequency, the loudness of noise alone to the loudness of the speech. If the sound were a mixture of speech and noise at a given frequency, then both would be reduced, with the reduction increasing the more the noise exceeded the speech. Such a control scheme would need to identify when speech was dominant and hence when the speech reference should be updated. This conceptualisation of the problem and possible solutions led to the following hypotheses:

1. It was hypothesised that, if the sounds a person heard at a given frequency were no louder at that frequency than the speech they were comfortable listening to in that situation, it would be unlikely the listener would find these sounds any less comfortable in the short term than listening to the speech or that they would receive an acoustic shock.
2. It was hypothesised that if control was applied only to sounds with a loudness in excess of speech this could be achieved without affecting speech quality unless noise of a similar or greater loudness was simultaneously present within the signal from which the speech came.
3. It was hypothesised that if there was some temporal or spectral difference between the louder noise and the speech there would be an improvement in the perceived speech clarity and intelligibility from this control.

From a sound energy perspective, the ear would be unlikely to experience stimulation that exceeded the level it was acclimatised to with such a control applied. Therefore, it was unlikely that any somatic reaction to the sound energy would occur that would be any different to the somatic reaction to speech.

The desired outcomes of this project were:

- Listening comfort and safety
- Good speech quality and intelligibility

To achieve the above outcomes, regular estimates of the loudness of the audio signal were needed. Regular estimates of the loudness of the speech conveyed by the audio signal were also needed to form speech reference loudness estimates. Provided the loudness estimates of the audio signal did not exceed these speech reference loudness estimates, the audio signal would remain unchanged. This meant speech alone would be unaffected. If, however, the loudness estimates of the audio signal exceeded these speech reference loudness estimates, the audio signal would be reduced in level so ideally its loudness estimates would not exceed those of the speech.

1.1 Thesis structure

The thesis is structured into ten chapters. The first three chapters following this introduction (Chapters 2-4) review the relevant literature and establish the background research and data upon which the design of the SRL scheme is based. The fifth chapter describes the theory behind the SRL scheme and its general functioning. Of all chapters, this chapter provides the best description of SRL. This is followed by chapters on the SRL MKI (mark 1) scheme and its evaluation and the substantially more sophisticated SRL MKII (mark 2) scheme and its evaluation, concluding with a final chapter summarising the thesis. An overview of the thesis structure, chapter by chapter, is as follows:

Chapter 1 Introduction

Chapter 1 provides an introduction to the thesis, including the background to the undertaking of this research. It describes in general terms what SRL aims to achieve and provides an overview of the thesis, its novelty and its contribution to the field.

Chapter 2 Psychoacoustic factors

Chapter 2 reviews the literature on the psychological response to sound as it relates to the provision of acoustic comfort and safety for listeners to electronic sound reproduction devices that are primarily intended to convey speech. This review covers the major psychoacoustic data on loudness perception and, to a lesser extent, on timbre. It assesses the strengths and limitations of this data and the models derived from it. In particular, it looks at variability in the perception of loudness both within and between individuals. This review provides the psychoacoustic background that underpins the design of SRL.

Chapter 3 Acoustic shock and related factors

Chapter 3 reviews the literature on acoustic shock and related physiological and neurophysiological aspects in humans. In particular, the middle ear muscle responses and the mechanisms that trigger them are considered. The information contained in this chapter provided the original motivation for the creation of SRL, i.e. preventing acoustic shock. The literature on physiological and neurophysiological response to sound guided many aspects of the SRL design.

Chapter 4 Sound level control methods

Chapter 4 reviews the literature on various methods of automatically controlling the level of sound produced by electronic devices that are primarily intended to convey speech and examines how this fits within the greater field of sound reproduction. It identifies the specific applications in which automatic sound control is needed to provide comfort and safety for the listener. The review also considers the effect various methods of control have on the intelligibility of speech. This review establishes the range of applications for SRL and serves to provide a benchmark for its performance.

Chapter 5 Speech referenced limiting – the theory

Chapter 5 introduces the novel speech referenced limiting (SRL) concept, including the theoretical basis for it, and describes the SRL scheme in general terms. The concept described is the foundation on which the implemented and evaluated schemes, SRL MKI and SRL MKII, were based.

Chapter 6 SRL MKI scheme

Chapter 6 provides a detailed description of the first implemented SRL scheme, SRL MKI. This scheme was initially developed with a focus on telecommunications applications and was later extended to other applications, such as level-dependent hearing protectors and hearing aids. The chapter presents the detailed signal processing algorithm developed for the scheme. Many of the methods introduced here were subsequently used in the SRL MKII scheme, such as the very clean processing of the signal for reproduction.

Chapter 7 Subjective evaluation of SRL MKI

Chapter 7 reports on the subjective evaluation of the SRL MKI scheme. This involved three laboratory experiments to assess the SRL MKI scheme for application in hearing aids, hearing protectors and telecommunications headsets. The experiments provided evidence that the SRL MKI scheme had confirmed the hypotheses that it would control the loudness of the sound with reference to the speech loudness and that it would preserve the quality of the speech. The limitations of this first version of the scheme were, however, noted and used to guide the development of the significantly more advanced SRL MKII scheme.

Chapter 8 SRL MKII scheme

Chapter 8 provides a detailed description of the second implemented SRL scheme, SRL MKII, which was designed to address issues that arose during the implementation and evaluation of SRL MKI as well as to perform with greater reliability in a diversity of applications. The major advance with the SRL MKII scheme was its ability to estimate the speech loudness from frequency regions where speech was dominant while ignoring frequency regions where it was not. A second advance was its fast control over signal peaks using a shorter delay.

Chapter 9 Subjective evaluation of SRL MKII

Chapter 9 reports on the subjective evaluation of the SRL MKII scheme. This involved five laboratory experiments to assess the SRL MKII scheme for application in hearing aids, level-dependent hearing protectors and telecommunications headsets, and to assess its ability to respond to abruptly varying speech levels and noise simultaneously presented with speech. The experiments compared performance against a dual-speed, multi-band limiter almost identical to the SRL MKII scheme which employs fixed limiting levels set by the subjects. It also investigates the subjects' preferred limits and the variability in loudness perception both between and within subjects.

Chapter 10 Conclusion

Chapter 10 summarises the thesis, drawing on the collated information, the ideas developed and the results of the developmental and experimental work. It concludes on the major findings, providing recommendations on the application of SRL and suggests areas of further research.

1.2 Novelty and contribution to the field

To the author's knowledge, the concept of controlling the excess loudness, or simply the excess level, of an audio signal with reference to that of the speech conveyed by the audio signal is novel, both on a broadband and on a frequency-specific basis. The novelty of this research was accepted by patenting authorities in the five countries that have granted patents for the method and system: Australia, China, India, the United Kingdom and the United States of America.[7-11]

The main contribution this research makes to the field of sound control is:

- It shows that it is possible to automatically control an audio signal such that the excess in the loudness of sounds exceeding the loudness of the preceding speech conveyed by the signal is minimised without affecting the loudness or quality of the speech signal when louder noise is not concurrent with the speech.
- It shows that it is possible to create a system that provides arguably the greatest reduction in the excess loudness of an audio signal compared with the loudness of the preceding speech conveyed by the audio signal for the least reduction in the speech loudness and quality.
- It shows that noise concurrent with speech within an audio signal can be reduced in loudness by controlling the loudness of the noise with reference to the preceding speech and that the resulting speech quality is as good if not better than that obtained by applying a limiter with a user-selected limit.
- It argues that this method of reducing sound levels on a multiband basis, when the level or loudness in the band exceeds the level or loudness of speech in that band, is superior to conventional forms of compression and compression limiting in terms of speech intelligibility, as estimated by the speech intelligibility index.[12]
- It argues that the variability in loudness perception and somatic response to sound both within and between individuals is too great for a fixed-reference

limiting scheme to adequately provide loudness comfort and acoustic safety for all and at all times.

- It argues that the SRL method addresses both loudness comfort and acoustic safety issues through estimating and controlling for both loudness perception and somatic response to sound with reference to speech.

It is the author's hope that this novel method of sound control will be widely adopted and that future research will allow it to make an even greater contribution to preserving the comfort and safety of people listening to speech through electronic devices.

Chapter 2

Psychoacoustic factors

2 Psychoacoustic factors

2.1 Introduction

This chapter reviews the literature on the psychological response to sound as it relates to the provision of acoustic comfort and safety for listeners of electronic sound reproduction devices that are primarily intended to convey speech. In this review, I cover the major psychoacoustic data on loudness perception and, to a lesser extent, on timbre. I assess the strengths and limitations of this data and the models derived from it. In particular, I look at variability in the perception of loudness within and between individuals.

An understanding of loudness perception and its variability is crucial to the development and evaluation of any sound-limiting scheme, as will become clear in later chapters.

2.2 Loudness and stimulus intensity

The word 'loudness' is commonly applied to the entire perceptual range of sound intensity including the perception of 'softness'. Fletcher and Munson (1933) defined loudness as "a psychological term used to describe the magnitude of an auditory sensation".[13] Florentine et al. (2011) defined loudness as "the perceptual strength of a sound that ranges from very soft (or quiet) to very loud".[14] These and many other definitions describe loudness as the perceived strength of sound. The loudness of a sound has a complex non-linear relationship with its physical intensity. This section discusses literature on this relationship, an understanding of which is essential to any attempt to predict and control loudness based on the physical properties of sound.

A mathematical relationship between loudness and physical intensity was defined by Fechner in 1860.[15] This relationship was based on Weber's law, which states "that the change in a stimulus that will be just noticeable is a constant ratio of the original stimulus", as cited by his student, Fechner. Fechner's law (or Fechner's scale) states that "subjective sensation is proportional to the logarithm of the stimulus intensity". A century on, in 1960, Stevens, stated that the relationship between physical intensity and sensation was a power function not a log function.[16] Stevens had stated this in earlier papers with reference to calculating loudness.[17,18] He had been developing this view since defining the sone as a unit of loudness.[19] Sones double (or halve) in magnitude with a doubling (or halving) of the perceived loudness. The equation for Stevens' power law is:

$$S = S_0 \times (I/I_0)^P \qquad \textbf{(2-1)}$$

where: S is the sensation, i.e. loudness in sones

S_0 is the reference sensation, i.e. 1 sone

I is the intensity, i.e. sound intensity, W/m^2

I_0 is the reference intensity, i.e. 10 nW/m^2 (40 dB SPL)

P is the exponent

For loudness, P is typically 0.3. With P set to 0.3, a 10 decibel (dB) increase in intensity (10 dB = 10 · log10 (I/I_0) where $I/I_0 = 10$) results in a doubling of loudness, S = 2 for $S_0 = 1$. This relationship, although limited to certain conditions, justifies the use of the decibel as a unit for the estimation of loudness. Many loudness estimators have employed this relationship, using measures of averaged intensity in decibels to indicate loudness. One example is the LM100 loudness meter from DOLBY®.[20] The use of the decibel as a measure of loudness in broadcasting was standardised in the International Standard ITU-R BS1770-2, 'Algorithms to measure audio programme loudness and true-peak audio level'.[21]

Despite its prevalence in psychophysics, Stevens' power law has limitations when it comes to loudness estimation. It is an attempt to fit a simple mathematical relationship using a limited set of stimuli to a highly complex biological system's response. Some of these limitations are discussed by Stevens in his earlier papers.[17,18] The law is only applicable to narrow-band sounds and does not take into account the issues of masking or loudness summation. The law, without modification, does not hold true for low or high frequencies or for low or high sound intensities. This is evident if one attempts to fit Stevens' law to the equal loudness contour data produced by Fletcher and Munson decades earlier.[13] Different exponents would be required at different frequencies and for different levels to fit this data. More advanced fitting of a mathematical relationship of loudness perception to sound intensity has been performed by, among others, Zwicker and Scharf,[22] Zwicker and Fastl,[23] Moore and Glasberg,[30] Moore et al.,[24] Fastl and Zwicker,[25] and Florentine and Epstein.[26]

In summary, loudness, as a first approximation, has a relationship with sound intensity that may be represented by a power law with an exponent of 0.3. As a first approximation, a doubling (or halving) of loudness equates to a +10 dB (or –10 dB) change in sound intensity or pressure. In terms of providing acoustic comfort and safety through sound control, this simple relationship has great appeal but should be used carefully given its limitations.

2.3 Loudness and stimulus frequency

The acoustic comfort and safety of listeners to electronic devices reproducing sound is affected by the frequency of the sound components. An imbalance in the frequency content may result in excessive loudness in one frequency region and a loudness deficiency in another, making listening a potentially unpleasant experience. Fletcher and Munson (1933) investigated the balance of loudness perception across frequency.[13] They produced the first set of equal loudness contours. These contours show the sound level as a function of frequency for which an equal loudness sensation occurs. They use the concept of loudness level which they define as 'The loudness level of any sound shall be the intensity level of the equally loud reference tone at the position where the head is to be placed'. They used a reference tone with a frequency of 1 kHz and produced equal loudness contours for every 10 dB step in its sound level. Independently, a unit of loudness level called the phon was introduced by Barkhausen in the 1920s.[25] It was defined to be equal to the sound pressure level (SPL) of a 1 kHz tone in dB, i.e. X phons is the loudness level of a 1 kHz tone at X dB SPL. Sometime after Fletcher and Munson's experiment the phon came into use as the unit of loudness level for equal loudness contours. Fletcher and Munson's classic experiment has been improved upon, by a number of researchers, using more modern equipment and more subjects. The results of this are standardised in the International Standard, 'Normal Equal-Loudness Level Contours', ISO 226.[27]

The contours show a pronounced elevation in hearing threshold with decreasing stimulus frequency below 500 Hz and a strong expansion in the loudness growth with sound level at low frequencies, which leads to a flatter sound level for equal loudness at high sound levels. At other frequencies the contours show some non-linearity in the loudness growth with sound level, exhibiting steeper growth at low and high levels and shallower growth at mid-levels. Of particular note is the frequency region from 2 to 5 kHz (around and above the ear-canal resonance frequency) where the curves display an increased sensitivity at all sound levels and a marked increase in sensitivity at high sound levels.

In the context of providing loudness comfort to listeners of electronic sound reproduction devices, the equal loudness contours provide a guide to the frequency balance for equal loudness comfort across frequency, at least for tones. However, as discussed later in this chapter, natural sounds have an intensity that decreases with increasing frequency and therefore produce less loudness with increasing frequency, i.e. less than equal loudness. The listener's familiarity with this spectral balance creates their normality or expectation. A departure from this, which can be introduced through the use of electronic sound reproduction devices, such as the telephone system or other low-fidelity sound devices, results in a distinct tonal quality being apparent to the listener, e.g. the sound of a voice transmitted by a telephone is easily recognised. Perceptual adjustment to an altered spectral balance, however, occurs during listening, such as when listening to a telephone, and this leads to an adjustment of the individual's expectation of how loudness should vary with frequency.

In summary, the equal loudness contours provide a guide to the frequency balance for equal loudness comfort, at least for tones. For naturally occurring stimuli the expectation of the loudness decreases with increasing frequency but an individual's expectation may be temporarily altered through acclimatisation to an altered spectral balance. As will be discussed in subsequent chapters, an abrupt change in the expected spectral balance that a listener is acclimatised to, such as an increase in sound energy around the ear-canal resonance frequency, may cause discomfort or even injury to a listener. Therefore, the methods to deliver acoustic comfort and protection need to adapt to the listener's preference at a given time.

2.4 Loudness and stimulus duration

To provide acoustic comfort, the effect of the duration of a sound on loudness needs to be considered. The early work on the relationship between duration and loudness was summarised by Zwislocki in 1960.[28] He concluded from his work and that of others that loudness level increases by approximately 10 dB for a 10-fold increase in duration, e.g. from 10 ms to 100 ms. He concluded that the integration was neural summation at a high level, which could be approximated using a first-order, low-pass filter with a time constant of 200 ms, at least at threshold. He revised this to 100 ms for mid to high sound levels in 1969.[29] As will be discussed later in this chapter, others have considered shorter and longer integration times.

The effect of pulse density on temporal integration was investigated by Zwicker and Fastl.[23] Short (5 ms) pulses of a 2 kHz tone were presented at 57 dB SPL with repetition rates varied from 1 Hz to 200 Hz. The results show approximately a 12 phon change in loudness levels although there was little change in loudness level up to a repetition rate of 5 Hz, and only a 2 phon increase at 10 Hz. This data supports the concept that, for pulse densities exceeding 10 ms per 200 ms, loudness is proportional to average power of the stimuli, at least at this sound level, and generally supports the 10 dB / 10-fold increase in density or duration relationship. It also supports an integration time constant in the order of 100 ms.

In the 1990s temporal integration rates were found to vary non-monotonically with level by Buus et al. (1997)[30] and Florentine et al. (1996).[31] Florentine's experiment involved determining the relative sound level required to achieve equal loudness for stimuli with various durations and sound levels. The stimuli were 1 kHz tones and broadband noises. The general pattern was that loudness integration for both the 1 kHz tone and the broadband noise was greatest at mid sound levels. For 5 ms and 200 ms stimuli durations, Florentine et al. found that equal loudness perception occurred when the 5 ms long stimuli were about 10 to 12 dB above the 200 ms stimuli which were set at near threshold. This rose to about 18-19 dB when the 5 ms tone was at 56 dB SPL and the 5 ms noise was at 76 dB SPL. At 100 dB SPL the difference fell back to about 10 dB for the tones and 13 dB for the noises. Their data indicates a slightly higher loudness integration rate than 10 dB for a 10-fold increase in duration at mid sound levels and less than this at low and high sound levels.

The relationship between the loudness growth data and the loudness integration data as a function of sound level appears to have an opposite but complementary trend. When loudness growth is steepest (i.e. at low and high sound levels) temporal integration is least and when loudness growth is shallowest (i.e. at mid sound levels) temporal integration is greatest. This observation led to the equal-loudness-ratio hypothesis proposed by Florentine et al.

The rate of loudness growth for sounds of brief duration, however, does not follow the 10 dB / 10-fold increase in duration relationship because sounds become broader in frequency when their duration is very short. This leads to an increase in loudness due to the summation of components at different frequencies.

In summary, the laboratory data indicates that a 10 dB increase in loudness level per 10-fold increase in duration and an integration time in the order of 100 ms applies at comfortable loudness levels provided the stimulus duration is not very short. To provide loudness comfort a device needs to be able to estimate and control sound as quickly as the auditory system integrates and reacts to it.

2.5 Loudness and stimulus direction / presentation method

When considering providing acoustic comfort and safety to listeners of electronically reproduced sound, the effect of the method of presentation needs to be considered. The presentation method affects the vibration reaching each cochlea and the bilateral composition of the presentation affects the neural processing and the resulting control within the auditory processing system. The control of outer hair cells in the individual cochlea and the control of the muscles in each middle ear are influenced by neural signals from the contralateral as well as the ipsilateral cochlea.[32]

Assuming normally functioning outer and middle ears, the dominant pathway for sound to reach the cochlea is via air conduction of sound, which is converted to mechanical vibration by the tympanic membranes and transmitted via the middle ear ossicles to the oval window of each cochlea. The loudness of air-conducted sound depends on how it reaches the tympanic membrane. The anatomy of an individual's outer ear, combined with that of their head and torso, influences the sound pressure at the tympanic membrane and hence the perceived loudness of a sound in a frequency-dependent manner. This frequency dependence varies strongly with the direction of the sound source and has been documented in detail by Shaw.[33-35] Shaw's data shows that, at specific angles, the transformation from the uninterrupted field to the tympanic membrane, across frequency can vary by up to 20 dB and, at specific frequencies, the sound level across angles can also vary by up to 20 dB.[33] Both variations can individually produce, in theory, loudness differences of up to four times. Added to this is a substantial variation in data across subjects as illustrated in the responses of individuals presented by Shaw.[36] The combined effect is that the individuals' auditory system is tuned to their ears providing substantial individualised variations in sound level with variation in sound source direction.

The specific binaural processing provided by an individual's outer ears in the field is in most cases substantially modified by the use of electronic sound reproduction devices. This is the case for many devices that present sound from the listener's environment to the listener (e.g. hearing prostheses without microphones in the ear canal and level-dependent hearing protectors / headsets). This is also the case for devices that present sound not received by microphones located on the individual (e.g. telephone handsets, communications headsets and headphones). Externalisation of sound is reduced, if not removed, through the loss of binaural head-related transfer functions (HRTFs).[37] The apparent proximity of a sound source in the near field is also affected by changes in the HRTFs.[38] [39] Externalisation and proximity are important to a listener's sense of personal space, the perception of potential danger and somatic reaction to it.[40] This is particularly true when the listener cannot see the sound source, such as when listening to a telephone or other audio-only signal. Loudness perception is altered by the absence of other stimuli, this is discussed later in this chapter.

Monaural presentation by electronic sound reproduction devices, such as a telephone handset or communications headset further changes the perception of sound including loudness. Monaural presentation to an ear alters the control of its outer hair cells and middle ear muscles by the contralateral ear. Monaural presentation also introduces altered and possibly inappropriate control over the contralateral ear's outer hair cells and middle ear muscles. Binaural function, including the binaural summation of loudness, is removed. Many researchers, such as Moore et al. (1997), consider the effect of binaural loudness summation to produce a doubling of the monaural loudness for an equal sound in each ear and hence a halving of the binaural loudness with monaural presentation.[24] Others consider it to be level-dependent. Fastl and Zwicker (2006) consider it to be a doubling at low sensation levels, i.e. about 10 dB, and reducing to a factor of about 1.4 at high sensation levels, i.e. about 5 dB.[25] Epstein and Florentine (2014) found lower degrees of binaural loudness summation, which not only varied with presentation level, but were also reduced when sounds were presented in the field (with an earplug used to create the monaural condition) compared to their headphone presentation condition.[41] They found that the binaural loudness summation factor reduced to approximately 1.06 when the sounds were presented at 55 dB SPL in the field and an image of the sound source was provided to the subjects on a video. As well as altering loudness, monaural presentation affects binaural noise suppression[23] and internal protection mechanisms for the ear.[42] These are important considerations when providing acoustic comfort and safety.

The response of the electronic sound reproduction device's transducer(s), including variation in its response, due to manufacturing variability and aging, and in its coupling to the ear[43] all lead to an altered sometimes unpredictable presentation of sound at the tympanic membrane. For example, the sound levels and frequencies produced by an ear bud and its effect on the amplitude and phase of external sounds varies with its position in the ear and this may vary in an unpredictable manner as the person moves their head.

In summary, the alteration of the sound 'expected' by an individual's auditory system as a result of electronic sound presentation, the variability in presentation technology and the variability in its performance in an individual together results in alteration and variation of the individual's perception of loudness and response to sound. This evidence supports the view that methods to deliver acoustic comfort and protection need to be adaptive to the listener's preference at a given time.

2.6 Loudness and stimulus bandwidth

The acoustic comfort and safety of a listener to an electronic sound reproducing device is affected by the distribution of frequencies that it simultaneously produces. It was well understood as early as the 1930s that the loudness of wideband sounds was greater than narrow-band sounds of the same level.[13] The concept of a minimum or critical bandwidth (CB) at which this difference occurred developed over the next two decades.

In the 1950s, Zwicker et al. determined a minimum or critical bandwidth at which a difference in loudness perception occurred as a function of frequency.[44] Zwicker formally defined a relationship between critical bandwidth and frequency using a scale which he called the Bark.[45] In 1983, Moore and Glasberg derived an alternative critical bandwidth (and scale), which they called the Equivalent Rectangular Bandwidth, or ERB.[46] This was based on a number of experiments involving the detection of a tone in noise where the noise was notch filtered around the tone frequency. In comparison to the bandwidths in the Bark scale, the ERB continues to decrease in width with decreasing frequency below 500 Hz and is narrower overall. Both scales, however, have a bandwidth that is approximately proportional to their centre frequency for frequencies above 500 Hz. The one-third octave bandwidth scale, although wider than both the Bark scale CB and the ERB, except at low frequencies, is also used in loudness estimation, such as in the International Standard, ISO 532, 'Acoustics - Method for calculating loudness level'.[47]

Methods of loudness estimation that employ the auditory band concept involve transforming frequency-specific band energy to frequency-specific loudness, using bandwidths such as one-third octave, Bark scale CBs or ERBs and combining the frequency-specific loudness estimates to create a total loudness estimate. This combination process is called spectral loudness summation. The transfer function from power, or basilar membrane excitation, to frequency-specific loudness is typically based on Stevens' power law but with specific modifications. Just as there are a variety of bandwidths in use, there are a variety of methods for how the frequency-specific loudness estimates are generated and how they are combined to produce a total loudness estimate. Method A of the International Standard, ISO 532, 'Acoustics - Method for calculating loudness level',[47] uses the following equation to calculate the total loudness.

$$S_t = S_m + F \times (\sum S - S_m) \qquad \textbf{(2-2)}$$

where: S_t *is the total loudness, in sones*

S_m *is the maximum specific loudness, in sones*

F *is the band weighting factor, 0.15 for one-third octave*

S *is the specific loudness in each band, in sones*

Method B of ISO 532 uses a graphical method to sum specific loudness in third-octave bands to calculate the total loudness. The method includes a graphical spreading of loudness into the higher bands which is included within the sum. More recent summation methods also diverge from the ISO 532 method A. Moore et al. and Glasberg and Moore produced a total loudness estimate by equally adding all the frequency-specific loudness estimates[24,48] as did Chalupper and Fastl.[49] More recently, Moore's 1997 method, which is standardised in the American National Standard S3.4 'Procedure for the Computation of Loudness of Steady Sounds',[50] and, to a lesser extent, the method of Zwicker, which is standardised in ISO 532 Method B,[47] have been criticised for overestimating spectral loudness summation.[51]

From a loudness control perspective, loudness summation over frequency needs to be taken into account. Data on the degree of spectral loudness summation varies between studies and has been shown to depend on both sound level and duration by Anweiler and Verhey (2006) and Verhey and Kollmeier (2002).[52,53] Röhl et al., using a loudness categorical scaling, found that, for 4,000 Hz centred pink noise with bandwidths of 50, 500, 1,500, 3,000, 6,000 and 8,000 Hz, loudness summation on average reached a maximum when the stimuli were around 65 dB SPL.[54] Loudness summation became slightly negative when the stimuli were near 0 dB SPL and approached zero at 90 dB SPL. Zwicker and Fastl show a reduction in loudness summation at higher sound levels, from a maximum of approximately 15 dB at 60 dB SPL to 10 dB at 80 dB SPL.[23]

Verhey and Kollmeier found that a 1 second noise centred on 3,200 Hz with a bandwidth of 6,400 Hz compared to 200 Hz produced loudness summation effects of 17 dB at 45 dB SPL, reducing to 15 dB at 65 dB SPL.[53] They found about a 4 dB increase in the effect of loudness summation when comparing data for the 1 second duration stimuli with that of 100 ms duration stimuli and found about a 7 dB increase in the effect when comparing it with that of 10 ms duration data. However, a change in loudness with duration was only present when the bandwidth was less than 3,200 Hz.

Anweiler and Verhey (2006), using loudness scaling, found that loudness summation peaked around 45 dB SPL and diminished at high and low sound levels. However, for a duration of 1 second, the maximum loudness summation (3,200Hz versus 200 Hz) was only 7 dB. This increased to 10 dB for a duration of 10 ms at 65 dB SPL. Using loudness matching, they found similar amounts of spectral loudness summation to Verhey and Kollmeier at durations of 1 second, and sound levels of 45 and 65 dB SPL, but only a very small effect of stimuli duration on loudness summation.

The interaction of stimulus level, bandwidth, duration, and measurement method is clearly complex. To add to this complexity, the level of sub-critical-band noises has been found to be higher than that of tones in a loudness matching experiment. This was found to be up to 8 dB when the noise was almost a critical bandwidth wide by Hots et al. (2014).[55] Furthermore, Röhl et al. (2011) found higher loudness ratings for 50 Hz and 500 Hz wide pink-noise bands centred on 4 kHz at 70 dB SPL compared to the same stimuli with a 1,500 Hz bandwidth.[54] They also investigated the effect of stimulus bandwidth using functional magnetic resonance imaging (fMRI) and the same stimuli. The voxel (volume of activity in the fMRI scan) for the 50 Hz wide pink-noise band was typically 2.7 times that for the 1,500 Hz wide pink-noise band of the same SPL. The voxel was also larger for a 500 Hz wide pink-noise band than for the 1,500 Hz pink-noise band of the same SPL. This psychoacoustic and brain imaging data is contrary to what one would expect for narrower band sounds compared with broader band sounds. The reverse occurs when the bandwidth is increased beyond 1.5 kHz. Röhl et al. consider that the effect at low bandwidths may be explained by peak listening to large amplitude fluctuations resulting from a narrower bandwidth. These recent papers indicate that there are more effects of bandwidth worthy of investigation.

In summary, in relation to providing loudness comfort and acoustic safety, one can ignore the effects of loudness summation at low to mid sound levels and concentrate on mid to high sound levels, i.e. 60 dB SPL or more. Given the variability in the data it is difficult to put a precise figure on the degree of spectral loudness summation that occurs at these levels: for full bandwidth stimuli, it appears to range up to about 15 dB at mid levels, and reduces at high levels for long duration stimuli. For short duration stimuli the spectral loudness summation effect appears to be several dB higher than for long-duration stimuli at mid levels but it too reduces at high levels.

2.7 Loudness estimators

The above discussion has looked at various aspects of loudness, its relationship to intensity, and the effect of frequency, direction/presentation method, duration and bandwidth. These are all important to understanding the perceptual effects of sound presented to the listener and help to provide guidance on how sound needs to be controlled in order to provide listening comfort and safety. The data is typically composed of averages produced under specific conditions which are highly dependent on the method used in the measurement. For example, Anweiler and Verhey (2006) found a clear dependence of loudness summation on stimulus duration when using loudness scaling but only a small dependence when using loudness matching.[52] Limitations in the data need to be taken into account in forming a view on loudness estimation in relation to comfort and safety.

While I draw on many of the aspects of loudness estimation described above, I do not use an existing loudness estimator in this thesis. A familiarity with some of the existing estimators' features does, however, assist in understanding the methods of loudness estimation and control I have developed. There are several prominent dynamic loudness estimators. These include the loudness estimators of Glasberg and

Moore, (G&M),[48] and of Chalupper and Fastl, (G&F).[49] In addition to these sophisticated estimators, there are simple loudness estimators such as the A-weighted long-term average sound level, Leq(A),[56] the LM100 loudness meter from DOLBY®,[20] and the estimators complying with the International Standard ITU-R BS1770-2, 'Algorithms to measure audio programme loudness and true-peak audio level'.[21] The LM100 tracks the Leq(A) using a 10 second sliding window,[57] while the ITU-B1770-2 uses 400 ms gated blocks with a 75% overlap.[21] Like the Leq(A) measure, they are not responsive to short-term changes in loudness and therefore are unsuitable for the estimation of loudness comfort and safety.

Both the G&M and C&F estimators are based generally on the model proposed by Zwicker,[23,58] but they differ in several ways. Briefly, the G&M estimator includes a static model for the outer and middle ear described by Moore et al.[24] It creates a 'running spectrum' using six time-aligned fast Fourier transforms (FFTs) in parallel with lengths of 2, 4, 8, 16, 32 and 64 ms. This enables high frequency resolution but low time resolution (i.e. more temporal smearing) at low frequencies and the opposite at high frequencies. Every 1 ms, the FFT outputs in the following frequency ranges are selected from the longest to the shortest FFTs respectively: 20 to 80Hz, 80 to 500 Hz, 500 to 1250 Hz, 1250 to 2450 Hz, 2450 to 4050 Hz and 4050 to 15000 Hz. From these frequency-specific FFT outputs, an excitation pattern is generated every 1 ms at frequencies spaced every 0.25 ERBs using the method described in Moore et al.[24] The excitation pattern is transformed to a frequency-specific loudness pattern and is summed to give the 'instantaneous' loudness. A short-term 1st order integrator with different attack and release times, 22 ms and 49 ms respectively, is applied to the instantaneous loudness to create an estimate of the short-term loudness, STL. The attack is designed to increase the loudness level estimate at a rate of roughly 10 phons per 10-fold increase in stimuli duration up to 100 ms. The release was designed to "give reasonable predictions of the overall loudness of amplitude modulated sounds".[48] A longer term 1st order integrator is applied to the STL to produce an estimate of the long-term loudness (LTL) using an attack time of 99 ms and a release time of 2000 ms.

The C&F estimator first high-pass filters the signal to take into account the low-frequency roll-off within the lowest critical band. A filter bank, implemented using the Fourier-t transform, separates the signal into critical bands. Envelope signals are formed from these signals with equivalent rectangular durations of 4 ms sampled every 2 ms. The bands are weighted to approximate the static effect of the outer and middle ear. The weighted output, called the excitation, is transformed to specific loudness. The specific loudness is decayed over time using a non-linear low pass filter to simulate the effect of post masking, based on the concept described in an earlier model of Zwicker's.[58] The specific loudness is spread from lower to higher bands to simulate the effect of upwards spread of masking. The resulting specific loudness pattern is summed to give the instantaneous loudness every 2 ms. A short-term 1st order integrator with a time constant of 125 ms is applied to the instantaneous loudness to create an estimate of the short-term loudness (STL).

The common elements of these models are: frequency weighting to approximate static effects of the outer and middle ear; the separation of the signal into auditory bands, i.e. excitation; non-linear transformation of the band signals to specific loudness; summation of specific loudness to produce total instantaneous loudness; low-pass filtering to simulate temporal integration. The models differ in: their auditory bandwidths; the fine details of their transformation from excitation to specific loudness; their simulation of temporal and frequency spreading and their temporal integration time constants. In particular, the C&F loudness estimator has more temporal integration prior to loudness summation.

The G&M and the C&F estimators have been jointly assessed in their ability to match subjective assessment of loudness for numerous sounds. Rennies et al. (2010) found that the estimators generally predicted the main trends in the data,[59] but that the short-term time constants of both models were slightly too small for some sounds. The temporal integration prior to loudness summation in the C&F loudness estimator appeared to enable it to better predict the loudness resulting from dynamic changes in the spectrum. With regard to speech-like signals, Rennies et al. (2013) found that the G&M LTL estimator better matched subjective loudness data.[60] With regard to 'technical' sounds, i.e. machine sounds, Rennies et al. (2015) found that the accuracy of both models was highly dependent on the stimuli, and they were least accurate when the sounds had large temporal variations.[61] They found that the G&M LTL estimator generally gave a better match with subjective loudness data.

In summary, the simple loudness estimators considered are inadequate in estimating short-term loudness comfort for highly fluctuating sounds. The G&M and the C&F loudness estimators are similar in concept. They are both based on psychoacoustic data from laboratory experiments and perform well on this type of test material. Their performance is less accurate estimating the loudness of more complex sounds containing greater temporal variation. None of the loudness estimators provide frequency-specific measures of the loudness of sound as outputs. However, the M&G and C&F models could be modified to provide them. Frequency-specific measures of the loudness are of interest if one wants to estimate the pleasantness of sound and listening comfort as is discussed in the next section.

Loudness perception can vary greatly in ways that existing mathematical models relating sensation to physical parameters do not consider. The section on variability in loudness perception discusses this in more detail.

2.8 Timbre, sharpness and pleasantness

The spectral content or timbre of a sound affects listening comfort. It is not simply the total loudness of a sound but also the frequency-specific loudness that affects listening comfort.

A well-known sound which causes most people to cringe is that of finger nails scraped across a blackboard. Halpern et al. (1986) investigated the perceptual effect of scraping a sharp object over a slate floor which they say 'mimics the sound of finger nails scraping across a blackboard'.[62] This sound was found to be the most unpleasant sound out of 16 sounds of similar amplitude by 24 adult subjects. Spectral analysis revealed that it had several strong harmonics with frequencies around 2.8 kHz, 4.2 kHz, 5.6 kHz and 7 kHz. However, there did not appear to be a component at the difference frequency of 1.4 kHz, i.e. the fundamental was missing. They presented high-pass and low-pass filtered versions of the stimulus, all at equal RMS amplitude, to 12 subjects. Decreasing the low-pass filter's cut-off frequency from 8 kHz to 3 kHz had no effect on unpleasantness of the stimulus, nor did high-pass filtering the stimulus with a cut-off frequency of 2 kHz. However, as the cut-off frequency was increased to 3 kHz and then to 4 kHz the unpleasantness reduced significantly, a further increase in the cut-off frequency past 4 kHz had no effect on the unpleasantness. It is clear that the 2.8 kHz component was primarily responsible for the unpleasantness. Furthermore, they found only a small difference in the loudness perception of the filtered sounds. This is contradictory to most data on loudness summation. They, however, did confirm that the subjects halved their loudness ratings for a 10 dB decrease in stimulus level indicating their method and subject responses were good. They also attempted to remove the temporal fine structure and found this had no effect on unpleasantness. It would appear from this data that the frequency-specific sensation (i.e. the frequency-specific loudness) possibly combined with some residual temporal variation rather than the overall loudness determined the perception of unpleasantness and hence listening comfort. This heightened sense of unpleasantness for a frequency component around the ear-canal resonance frequency corroborates with data on acoustic shock incidents as discussed in Chapter 3.

The pleasantness of auditory sensation has also been investigated by Fastl and Zwicker.[25] They describe a measure of sharpness, the 'acum', which is monotonically related to stimulus frequency. The greater the high-frequency content of the stimulus the greater the perceived sharpness. Pleasantness decreases with increasing sharpness of the stimuli.

High-level, high-frequency sound produces a perception of sharpness and unpleasantness. Loud sounds, and particularly those perceived as high in pitch, are largely absent in the natural world. The more intense natural sounds are perceived as having a lower pitch, for example, the roar of the sea, the howl of the wind and the pitch of the human voice. Natural sounds perceived as having a high pitch, such as the hiss of a snake, the whistle of a bird, and the speech phoneme /s/ are of significantly lower intensity compared with lower-pitch sounds. Speech, for example, has an average energy distribution that is greatest at the lower middle frequencies around 400 Hz. The average energy distribution of speech decreases with increasing frequency, as illustrated in Figure 2-1 from the data of Byrne et al. (1994).[63]

As discussed previously, the human auditory system is more sensitive to sounds of higher frequency than those of lower frequency. Equal loudness contours show an increased sensitivity to high-frequency sound.[27] This increased sensitivity to high frequencies compensates to some extent for the decrease in intensity of natural sounds at high frequencies. However, it makes the unnaturally occurring high-level, high-frequency sounds sharp and unpleasant. One may speculate that this higher sensitivity has evolved to compensate for the lower intensity of high-frequency natural sounds. In particular, the auditory system is very sensitive to high-pitch sounds with frequencies in the range of 2,000 Hz to 5,000 Hz primarily due to the combined resonance of the concha and the ear canal. The relationship between the sound pressure at the eardrum compared to the undisturbed free field for sound sources directly ahead of the average listener given by Shaw[33] is shown in Figure 2-2.

The effect of this relationship on the falling spectrum of speech is shown in Figure 2-3. The spectrum of speech at the eardrum is flatter than in the free field.

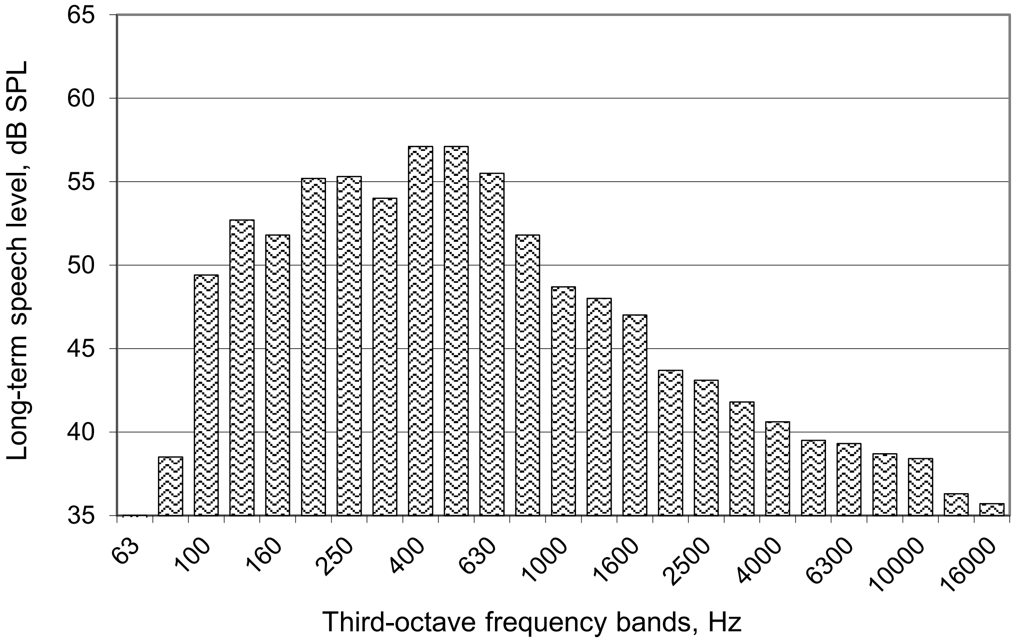


Figure 2-1. *Long-term average speech spectrum in the free field (total level: 65 dB SPL).*

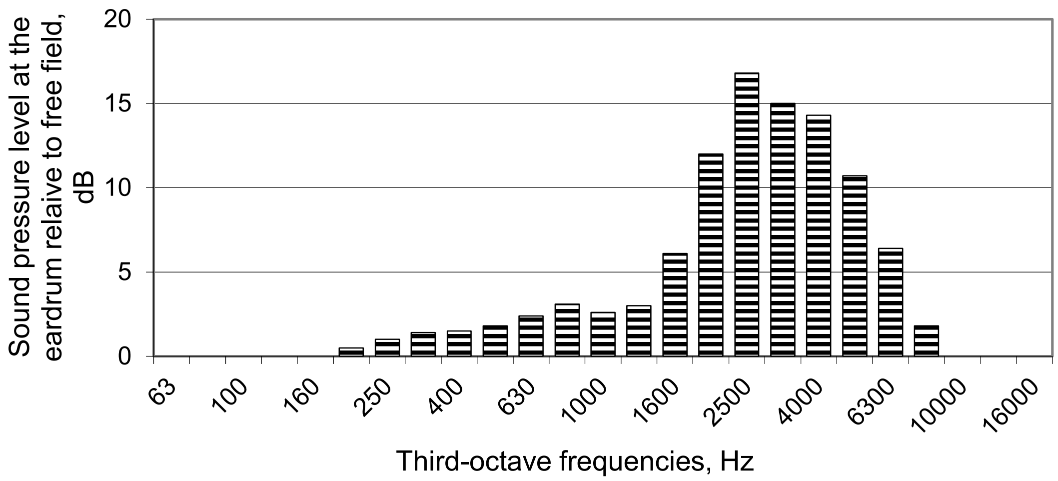


Figure 2-2. *Sound pressure at the eardrum relative to the free field.*

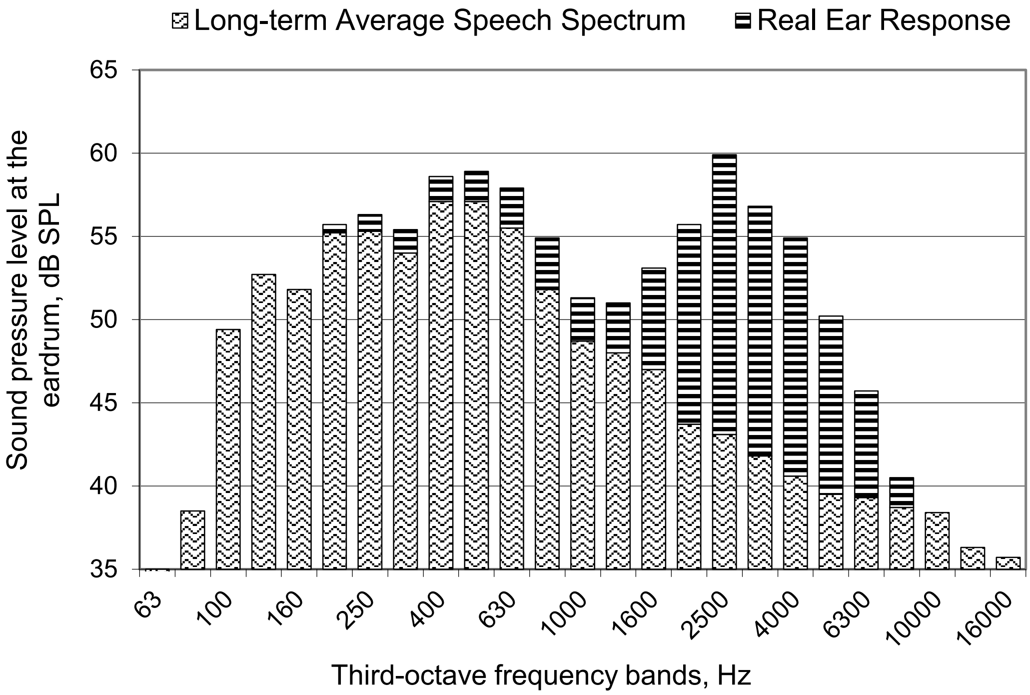


Figure 2-3. *Long-term average speech spectrum at the eardrum.*

Unlike natural sources of sound, man-made devices are capable of producing high-frequency sounds at high intensities. Warning devices such as smoke alarms and umpires' whistles are both high in frequency and intensity and typically resonate at the frequencies at which the auditory system is most sensitive. This is designed to have an alarming and influential effect on the human psyche. A comparison of the spectra of a whistle (at a considerable distance, e.g. 20 metres) to average speech (in very close proximity, i.e. 0.2 metres) at the eardrum is depicted in Figure 2-4. The frequency-specific loudness of the whistle at 2.5 kHz is about four times that of the speech. Although the whistle illustrated in Figure 2-4 has significantly more energy at its resonance frequency, and hence specific loudness, compared with the speech, the two sounds have the same broadband long-term level in the free field of 65 dB SPL. Electronic sound reproduction devices are capable of distorting the spectra of natural sound, such as exaggerating the energy around the ear-canal resonance frequency to produce an un-natural listening experience.

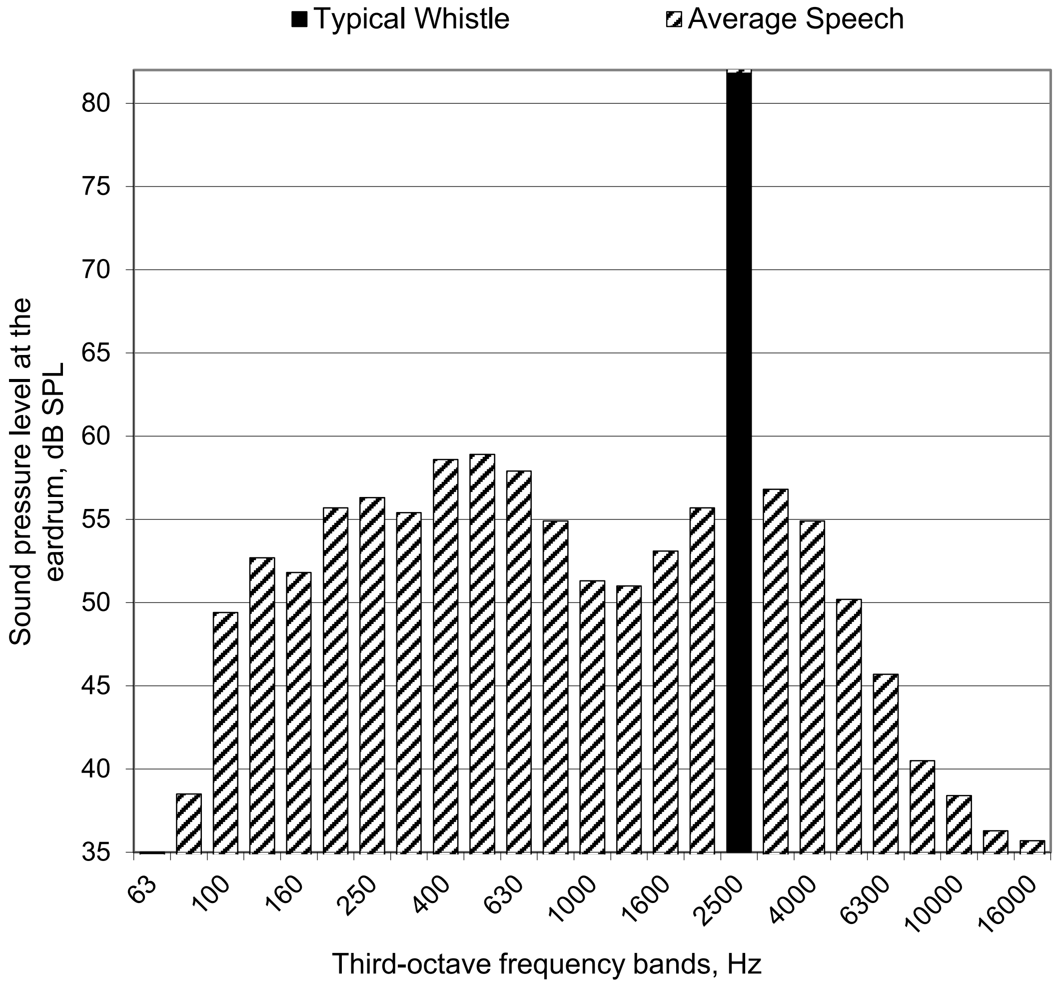


Figure 2-4. *The spectrum of a whistle and the average speech spectrum at the eardrum. Both sounds have a level of 65 dB SPL in the free field.*

In summary, the intensity of natural sound decreases with increasing frequency and the expectation of the natural timbre of sound follows this distribution. Sharpness increases and pleasantness decreases with an increase in high-frequency energy in relation to low-frequency energy. This appears to only occur with man-made sounds, which may include natural sounds unnaturally reproduced by electronic sound reproduction devices. With regard to providing listening comfort, a device needs to control any increase in the frequency-specific loudness above the listener's comfortable levels. It therefore follows that methods to deliver acoustic comfort need to adapt to the specific loudness that a listener finds comfortable at a given time.

Loudness perception can vary greatly in ways that existing mathematical models relating sensation to physical parameters do not consider. The next section discusses this in more detail.

2.9 Variability in loudness perception

Existing models of loudness perception, such as those by Zwicker,[22] [58] [47] Moore,[24] Glasberg and Moore,[48] and Chalupper and Fastl,[49] are based on averaged perceptual data. They do not account for variations in the individual's perception of loudness, apart from, in some cases, consideration of the average effects of hearing loss. Loudness perception varies between individuals, and within an individual as a result of psychological state and context. Successful loudness control therefore needs to be adaptive and referenced to the listener's individual preference at any given time.

Loudness perception has been found to vary with age, gender, individual characteristics, involvement, context, comprehension and with other sensory input, as set out below.

2.9.1 Age

In a study of 799 subjects aged from 17 to 92, Coren determined the most comfortable level (MCL) for discourse in quiet.[64] Before the age of 40, the MCL rose by about a third of a dB per year, or 1.8 dB every 5 years. After the age of 65, it rose by over a half a dB per year, or 2.8 dB every 5 years. Based on this data, the MCL of a typical middle-aged person is 10 dB higher than it was when they were in their late teens. Accordingly, if the speech level is set at the MCL for a middle-aged person, an average person in their late teens will find it to be double their preferred loudness or MCL. Clearly, any electronic device reproducing speech that is used by people of different ages needs to provide the user with a means of adjusting the level, e.g. a volume control, and any loudness control system needs to take this into account.

2.9.2 Gender

Rogers et al. studied 50 young adults, half female and half male, and found that the MCL for listening to male discourse in quiet was 6 dB lower for females than males.[65] The MCL for the discourse was 56.2 dB SPL for females and 62.1 dB SPL for males. When background noise (babble) was added, they found that the acceptable noise level (ANL) or acceptable signal to noise ratio was very similar for both genders, 10.4 dB for females and 11.4 dB for males. The acceptable background noise levels were 45.8 dB SPL for females and 50.7 dB SPL for males. Similarly, Kellaris and Rice found a preference for music at lower sound levels by females compared with males.[66]

Thomas and Jones studied the uncomfortable loudness level (ULL) of 26 individuals, half who had shown a high annoyance to noise, and half who had shown a low annoyance to noise in an initial survey of 122 subjects.[67] They found that the female participants had 13 dB lower average ULL than the male participants.

Therefore, gender, like age, also introduces variation in preferred loudness levels. This gives further weight to the need for speech reproduction devices to provide users with level adjustment, and for loudness control systems to be responsive to variation in loudness preference.

2.9.3 Individual characteristics and involvement

Why is it that one person may use a jack hammer without ear protection while another person 20 metres away finds the loudness of the jack hammer intolerable? Why is it that some people are excited by and enjoy the excessive sound levels at car races while others find them offensive? Kardous and Morata report that sound levels at car races are very high and note this to be part of the allure of the sport.[68] Could it be that greater involvement with the sound source changes the perception of its loudness, making high-intensity sounds appear softer? Fastl investigated an aspect of involvement using a kind of virtual reality.[69] He recorded moving images using a video camera mounted on the dummy head used for the audio recording. Subjects rated the loudness of both sound alone and sound accompanied by a moving image. Fastl found an average 8% reduction in loudness ratings when the sound was accompanied by the moving image. For some individuals, the reduction in loudness was more than 50%. Although this does not show participation, one would assume there is greater involvement with the sound, through an improved virtual reality presentation, which has led to this reduction in loudness. As noted in the next section the decrease in loudness was not as pronounced with the addition of visuals without the virtual reality component. I speculate that the characteristics of the individual and the degree of their involvement with the sound source influence their judgement of its loudness. Further research in this area would help to develop a better picture of these relationships.

As with age and gender effects, individual characteristics and involvement can also affect preferred loudness levels. Any electronic device reproducing speech that is used by people with different characteristics, in different situations, and with different degrees of involvement, needs to be level adjustable and any loudness control system needs to take this into account.

2.9.4 Context and visuals

Context also affects the individual's perception of sound. High-level traffic noise may be tolerable when walking down a busy street, but would be too loud if an audio/video recording of the traffic was played at the same level when at home. Fastl investigated aspects of this by adding still pictures and moving pictures to a sound recording of a train.[69] He found that still pictures could reduce the perceived loudness by an average of 2.5% while moving pictures could reduce it by 5%. When he changed the colour of the train, he found that subjects perceived the red train to be louder than the green train. However, the impact of colour alone on loudness perception was found not to be statistically significant by Parizet and Koehl.[70]

The fact that not being able to see the source of a sound influences loudness perception has ramifications for all sounds that are disconnected from visual perception of the source, such as sounds heard through a telephone, headset or headphone. Fastl's findings that these sounds are perceived as louder means any loudness control system needs to take this perception into account.

2.9.5 Comprehension

The comprehension of sound would appear to influence its perceived loudness. Segregation and grouping of sounds has been investigated in detail by Bregman.[71] Just as musical ability varies between individuals, so does the ability to segregate sounds. Zendal and Alain investigated the differences in the ability of musicians and non-musicians to segregate simultaneously occurring sounds.[72] They found that musicians were better at this task and their synchronous cortical responses to the stimuli were different from those of non-musicians.

One individual will segregate sounds in a different way to another and therefore have different opinions on the loudness of individual sounds within a complex sound. For example, one individual may say: 'The trombone player was so loud I could hardly hear the trumpet'. The person they are talking to, however, may hear the music as a whole and hence make no distinction between the sound of the various instruments in the brass section or their loudness. As this example illustrates, people with a different comprehension of a sound have a different perception of the loudness of its components. It may be that a difference in the loudness perception of a sound's components results in a different loudness perception of the entire sound. It may be that the more one understands a sound the less 'noisy', i.e. ill-defined, it appears and therefore the less loud it appears. If seeing the source of a sound reduces its loudness, it would appear plausible that understanding sound or understanding the source of a sound may also reduce its loudness. Although this is speculative, anecdotal evidence suggests this area could benefit from further research. The implication for this potential variability in loudness perception is that any protection system needs to adapt to the user's preferred loudness.

2.10 Summary

This chapter has reviewed psychoacoustic data related to providing acoustic comfort and safety for listeners of electronic sound reproduction devices that are primarily intended to convey speech. Loudness perception has been considered in detail. The intensity, frequency, duration, bandwidth and direction/presentation of sound have been discussed in terms of their effect on the psyche of the average normal-hearing person. The data on these relationships and the fitting of mathematical formula to them has been reviewed and the relevance to acoustic comfort discussed. Data on the variability in the perception of loudness, within and between individuals, has been reviewed and implications for providing acoustic comfort have been considered. The main conclusions drawn from this review are:

- Loudness growth can be approximated using a simple power law under limited conditions, and a 10 dB intensity increase (decrease) per doubling (halving) of loudness relationship holds under limited conditions. The limitations, however, need to be taken into consideration when using this relationship in controlling loudness.
- The frequency balance for equal loudness comfort may be guided by the equal loudness contours. The expectation of loudness decreases with increasing frequency but an individual's expectation may be temporarily altered through acclimatisation to an altered spectral balance. To provide listening comfort a device needs to consider the expected spectral balance the listener is acclimatised to at the time.
- Loudness growth with duration/density of a sound can be approximated using the 10 dB /10-fold change in duration/density and a time constant in the order of 100 ms provided the stimulus duration is not very short. To provide

loudness comfort a device needs to be able to estimate and control sound as quickly as the auditory system integrates and reacts to it.

- As a result of electronic sound presentation there is an alteration of the sound 'expected' by an individual's auditory system. The variability in presentation technology, its performance in an individual, and in the individual's perception of loudness and response to sound is complex. Taking into consideration the listener's preference at a given time would appear to bypass many of these factors, and provide the best approach to providing listening comfort for unexpected sounds.
- The relationship of loudness to a sound's bandwidth / spectral content, particularly with dynamic changes, is complex and could benefit from further research. Loudness summation at mid sound levels (i.e. 60 dB) may be up to 15 dB, decreasing at high sound levels.
- Only the psychoacoustic-based loudness estimators provide short-term loudness estimation suitable for guiding control of loudness. The short-term loudness estimations, while accurate when using specific laboratory-generated stimuli, are less accurate in determining the loudness of temporally fluctuating non-laboratory-generated stimuli.
- Timbre of a sound influences the listening comfort. Sharpness increases and pleasantness decreases (which is presumed to also decrease listening comfort) with an increase in high-frequency energy relative to low-frequency energy. A narrow-band energy component with a frequency around that of the ear-canal resonance frequency can be particularly unpleasant. With regard to providing listening comfort a device needs to control any increase in the frequency-specific loudness over that which is comfortable.
- Loudness perception varies greatly between and within individuals. There are many variables involved in the perception of loudness, and this means caution is necessary when applying laboratory data and models in the field. As we don't fully understand all the mechanisms that influence the perception of sound we need to design protection methods that deal with the unknowns as much as possible. Because people's perception of loudness is not fixed an adaptive protection scheme that uses their preferred loudness at any given time to control the loudness of out-of-character succeeding sounds has advantages over a fixed protection scheme.

Chapter 3

Acoustic shock and related factors

3 Acoustic shock and related factors

3.1 Introduction

This chapter reviews the subject of acoustic shock and related factors. It looks at the neurophysiological mechanisms involved in acoustic startle and acoustic shock, the reports on acoustic shock incidents, the symptoms that arise, and the sounds involved. In particular, this chapter considers the middle ear function, the mechanisms involved in controlling the transfer of energy to the cochlea, the middle ear defence mechanisms and the injuries that can occur. The effect of fear and anxiety on the response to sound is considered and, in particular, the tonic tensor tympani phenomenon is reviewed.

The desire to prevent acoustic shock, combined with the need to provide good speech intelligibility for telephone headset users, was the original motivation for the research described in this thesis and, ultimately, the development of the SRL system.

3.2 Physiology of the middle ear

3.2.1 Overview

The middle ear contains a mechanism for transmitting the pressure variations that sound waves present to the tympanic membrane on to the cochlea via its oval window membrane.[73,74] The two membranes are connected via a chain of small bones (the middle ear ossicles), which are supported by a number of ligaments and two middle ear muscles. These muscles control how much of the force at the tympanic membrane reaches the oval window in a frequency-dependent manner, changing the force, for example, when the person is chewing,[75] when intense low-frequency sound is present,[76] or when an abrupt or high-level sound is received.[74] The air-filled middle ear cavity containing these membranes, ossicles, muscles, ligaments and the cochlea's equalising membrane (the round window) is vented through to the nasopharynx via the Eustachian tube. The Eustachian tube opens periodically to adjust for changes in barometric pressure so that the tympanic membrane can operate at its optimum. All of these elements combine to create a highly complex control and defence system within the middle ear.

3.2.2 The middle ear

The middle ear contains three ossicles: the malleus (or hammer) attached to the tympanic membrane's manubrium (or ridge); the incus (or anvil) connecting the malleus to the stapes; and the stapes (or stirrup), which is attached to the oval window of the cochlea via the annular ligament of the stapes footplate.

The ossicles are supported by ligaments attached to the surrounding tissues and bone. Two muscles attach to the ossicles:

1. The tensor tympani muscle is attached at one end to the manubrium of the malleus (i.e. where it connects to the tympanic membrane) and at the other end to the wall of the Eustachian tube. The tensor tympani muscle is innervated by the motor branch of the trigeminal (V^{th} cranial) nerve.

2. The stapedius muscle is attached at one end to the neck of the stapes (i.e. close to its point of connection with the incus) and at the other to a channel within the wall of the middle ear cavity. The stapedius muscle is innervated by the facial (VII^{th} cranial) nerve.

The primary function of the middle ear is to match the low impedance of the air outside the tympanic membrane to the high impedance of the fluids in the inner ear. The oscillating air particles may have large displacements but they only have a small force, whereas the highly incompressible fluids in the inner ear require strong forces to produce displacement. The tympanic membrane moves with the air particle motion in the ear canal and hence transforms the acoustic energy into mechanical energy. The motion of the tympanic membrane is transmitted to the oval window membrane of the cochlea by the middle ear ossicles. The levering action of the malleus, incus and stapes results in a lever ratio of 2. This lever ratio, combined with a 15:1 area ratio between the tympanic membrane and the footprint of the stapes, results in a good impedance match between the air and the inner ear fluids at around 1 kHz. The total pressure gain provided by the middle ear including the ear drum appears to be 0 dB at low frequencies, increasing to about 20 dB around 800 Hz and decreasing thereafter. This estimate has been obtained by Puria et al. from measurements on four human cadaver ears.[77] However, the transfer function for live humans varies significantly.[74]

3.2.3 Middle ear muscle responses

The tensor tympani and stapedius muscles alter the transfer of energy from the tympanic membrane to the oval window. The stapedius muscle is activated by high-level sound, particularly at low frequencies. It stiffens the movement of the stapes and pulls the foot of the stapes away from the oval window to reduce the energy transferred to the oval window. This action is called the stapedius reflex or acoustic reflex.[73] This reflex is more sensitive to low frequencies.

The tensor tympani muscle is activated by a startle and it stiffens the malleus to reduce the energy transferred to the incus. This action is called the acoustic startle reflex. It can be activated by an unexpected sound or event such as a puff of air to the eye.[78]

Both muscles are activated when a person shouts, protecting the cochlea from the intensity of a person's own voice.[78] They can also be activated by a person's own voice when talking, as well as by chewing or yawning. In a human, the tensor tympani muscle is not believed to normally respond to sound energy unless the sound is of a high intensity and/or the sound is startling. By contrast, in most other mammals the tensor tympani muscles actively respond to sounds at lower levels as well as high levels and to startling sounds.

3.2.4 The innervation and sensitivity of the tympanic membrane

The tympanic membrane is innervated with three nerves:

1. The auriculotemporal branch of the trigeminal (V^{th} cranial) nerve innervates the anterior half of the tympanic membrane.

2. The auricular branch of the vagus (X^{th} cranial) nerve innervates the posterior half of the tympanic membrane.

3. The tympanic branch of the glossopharyngeal (IX^{th} cranial) nerve innervates the inner surface of the tympanic membrane.

The tympanic membrane is extremely sensitive. It alerts the person, through producing a sensation of pain, when there is too much pressure on it, such as:

- when diving to depths of more than approximately two metres
- when directly touched
- when there is an infection in the middle ear
- when it is overexerted by the tensor tympani muscle

3.3 Neurological pathways that stimulate the middle ear muscles

The pathway for the middle ear muscle (MEM) reflex as put forth by Mukerji et al. (2010) is described below.[79]

The action potentials generate nerve firings in the spiral ganglion cells within the cochlea. These propagate along the auditory nerve to interneurons in the ventral cochlear nucleus (VCN), which is part of the cochlear nucleus (CN). The interneurons project on to the MEM reflex motor neurons, located near the motor nuclei of the facial nerve or the trigeminal nerve. However, the exact pathways are currently not well understood. The stapedius motoneurons (SMNs) then project along the facial nerve to innervate the stapedius muscle. The tensor tympani motoneurons (TTMNs) then project along the trigeminal nerve to the tensor tympani muscle.

In addition to the path from the ipsilateral CN to the SMN (which may be indirect), there is a path from the contralateral CN. There are also speculated to be paths from the cortex, the locus coeruleus and superior ovilary complex. These provide non-auditory stimulation of the stapedius muscle and alter its sensitivity to auditory stimulation.

In addition to the path from the ipsilateral CN to the TTMN (which may be indirect), there is a path from the contralateral CN. There are additional paths that provide non-auditory stimulation of the tensor tympani muscle and alter its sensitivity to auditory stimulation, but these paths are yet to be well described in humans.

3.4 Acoustic startle integration time

It has been shown in rats that the temporal integration time constant for the acoustic startle response is 3 ms.[80] There appears to be no reason to believe that the integration time constant would be different in other mammals given their similarities in other aspects of the acoustic startle reflex. Acoustic startle time constants from other mammals have been applied to humans.[81] There is also a propagation delay, defined as the time delay or latency from the sound of a click being presented to the human ear to a response appearing at the brainstem. In the case of humans, this is approximately 5 ms (wave V).[82] It may therefore be the case that, for sustained sounds, the full effect of startle will occur within about 11 ms (the delay plus two time constants). However, for startle prevention, it is the integration time that is of consequence, as the delay is largely inconsequential. It is therefore necessary to prevent abrupt increases in sound (relative to the sound level being experienced) using a time constant that is faster than the 3 ms integration time of the startle response.

3.5 Acoustic reflex integration time

Unlike the extremely fast acoustic startle reflex integration time, the integration time associated with the stapedius muscle reflex appears to have a time constant similar to loudness integration. The time constant for this somatic response has been reported to be in the order of 200 ms.[83] The relationship between the stapedius reflex threshold and duration is steeper (25 dB per 10-fold change in duration) for durations shorter than 50 to 80 ms.[74]

3.6 Fear, anxiety, muscular tension, acoustic startle, the tonic tensor tympani phenomenon, and the resulting symptoms

Fear, anxiety and other psychological factors may cause various muscles in the body to contract. This applies to the stapedius and the tensor tympani muscles and has been confirmed using electromyographic measurements during middle ear surgery.[84]

Anticipation of a loud sound leads to middle ear muscle contraction, an effect measured by Djupesland.[74] A toy pistol was shown to a normal-hearing subject and the person was told that it could produce a loud bang. No further information was given. Later, the toy pistol, 'which had been kept concealed after it had been shown, was brought to view and lifted into a shooting position', Djupesland wrote. Of the 75 subjects tested, 70 subjects elicited rapid contraction of many muscles in the head and neck at the sight of the pistol. All the subjects shut their eyes and acoustic impedance of their ears simultaneously changed, indicating contraction of one if not both the middle ear muscles. On a subsequent presentation of the toy pistol, the subjects, having realised the loud bang would not occur, neither produced visible contraction of their muscles or a change in the acoustic impedance of the ear.

Djupesland also reported on the results of performing electromyography on the orbicularis oculi, stapedius and tensor tympani muscles in 30 patients with otosclerosis during a stapedectomy operation. In 24 of 30 patients, the presentation of the toy pistol led to contraction of several muscles in the head and neck and increased activity of the orbicularis oculi muscle and the stapedius and tensor tympani muscles.

These findings demonstrate that anticipation of a loud sound leads to a middle ear muscle response. One may speculate that people having been shown a particular electronic sound device and told that it can produce a loud sound are likely to be fearful of it and tense their middle ear muscles in anticipation of the noise. One may also speculate that, having been shown that the electronic sound device will not produce a loud sound (e.g. when an acoustic shock protection device is connected to a headset), people will release this tension.

Tension of the tensor tympani muscle can be tonic (i.e. continuous), an effect measured by Klockhoff and Westerberg.[85,86] According to Djupesland,[74] these researchers initially observed that, 'in some individuals the acoustic impedance of the ear fluctuated appreciably, irregularly and rather slowly'. They noted that for a small change in the ear canal pressure there was an extremely large change in the middle ear's impedance. According to Djupesland, they concluded that: '…the impedance changes were caused by fluctuating tonic contractions of the tensor tympani muscle, "tonic tensor phenomenon." The spontaneous impedance fluctuations were completely abolished by both succinylcholine and sodium pentothal, which shows that the phenomenon is of a muscular origin.' There were several reasons for the conclusion that the contracting muscle was the tensor tympani and not the stapedius muscle. First, the impedance changes were larger than the maximum acoustic reflex response that can be obtained from the stapedius muscle. Second, the phenomenon had been observed in many ears in which the stapedius was unable to influence the impedance due to either otosclerosis, paralysis of the stapedius muscle due to facial palsy or discontinuity of the ossicular chain. Third, where there was a unilateral section of the tensor tympani muscle in a bilateral tonic tensor tympani case, the fluctuations ceased in the operative ear but continued in the other ear.

Djupesland lists the following symptoms observed in patients exhibiting the tonic tensor tympani phenomenon (TTTP):

a. A sense of pressure and fullness in the ear, sometimes otalgia.
b. Tinnitus and/or other transient acoustic sensations.
c. Abnormalities of sound perception, such as slight episodic waxing and waning of sound, and speech discrimination problems. Pure tone audiometry shows normal hearing.
d. Dizziness of a nonspecific type, a subjective feeling of disequilibrium. Conventional routine testing of vestibular dysfunction is usually negative. Faint spontaneous nystagmus may, however, occur and "directional preponderance' is a relatively common finding in the electronystagmogram.
e. Tension headache is present in the majority of cases. Simultaneous electromyographic recording from the temporal muscle and impedance recording of the tensor tympani contractions show that the fluctuation of muscle tone was of the same character but not synchronous (Klockhoff and Westerberg, 1972).
f. Elevated psychic tension seems to be the essential etiologic factor in almost all cases.

Section 3.8 describes a report of acoustic shocks from 1957, 14 years before Klockhoff and Westerberg discovered TTTP. In this 1957 paper, Palva states, 'A sensation of dullness and blocking of the ear was felt and the ringing in the ear continued for some time after the event'.[87] This would appear to correlate with symptoms a and b in Djupesland's list above.

3.7 Ear pain resulting from sound exposure

It appears highly plausible that the reaction of the middle ear muscles in response to a startle and/or sustained high levels of sound, with or without the tonic tensor tympani phenomenon, results in a sensation of pain being produced by at least one if not all three nerves that innervate the tympanic membrane.

3.8 Acoustic shock from electronic sound reproduction devices

In 1957, following complaints of harmful effects from telephone signals, a study by Palva found that 14% of 139 telephone exchange workers in Turku, Finland, were suffering occupational deafness.[87] Palva wrote: 'The aural symptoms were stated to be aggravated especially, if an alarm signal was delivered directly into an operator's ear from another exchange. A sensation of dullness and blocking of the ear was felt and the ringing in the ear continued for some time after the event.' Some alarm signals were of the order of 130 dB, the author wrote. The hearing loss of the employees was characterised by a mid-to-high frequency (1500-4000 Hz) dip in sensitivity and was described as a result of acoustic trauma.

With the steady growth in the number of people using headsets in call centres from the 1980s onwards the number of incidents related to acoustic exposure from a headset increased. In 2001, a report by Milhinch on 103 cases of acoustic shock in call centres belonging to Australia's largest telecommunications company, Telstra, was released.[88 2] A pattern to the sounds causing the incidents and the resulting symptoms emerged.

All 103 cases investigated by the author of the papers were reported between 1995 and 1999 and were considered significant as they involved loss of time from work. The gender breakdown was 89% female and 11% male, a significantly higher proportion (binomial distribution; $p<0.001$) of females than would be expected from average employment figures for call centre workers of 74% female and 26% male. The age range was from 21 to 63 years, with a mean age of 34 years. Repeated acoustic shock incidents had occurred in 19% of cases, and 124 incidents were documented in total. Figure 3-1 shows the number of symptoms reported for the 124 reported acoustic shock incidents.

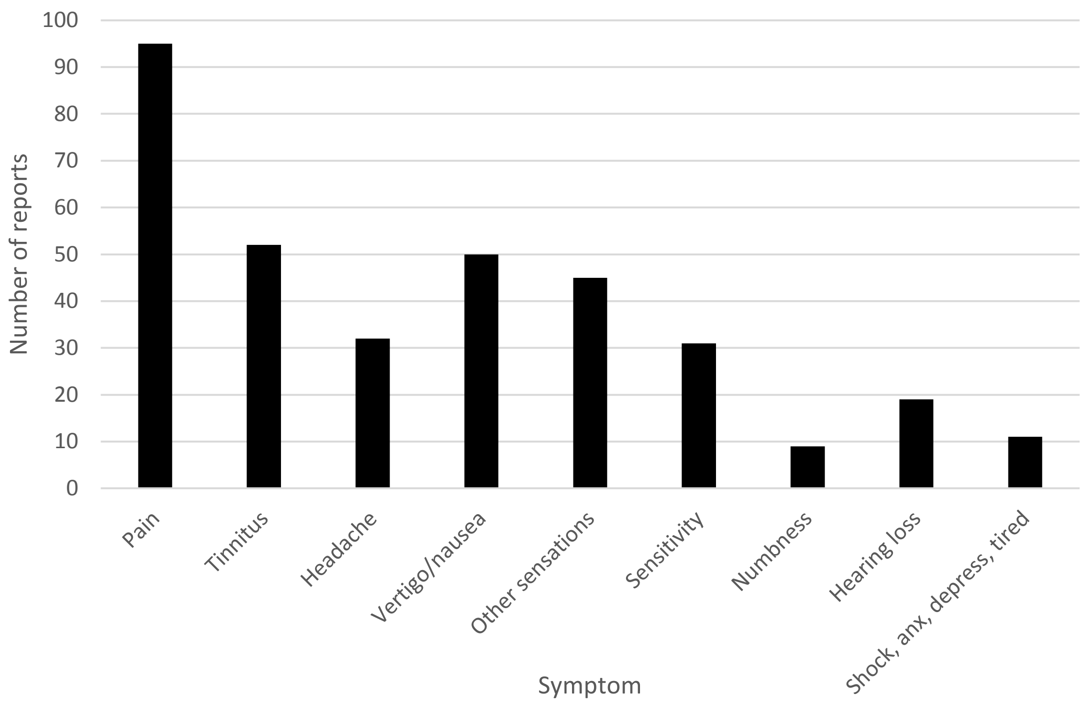


Figure 3-1. *Reported symptoms in 124 incidents of acoustic shock (adapted from Milhinch[88]).*

The data suggested that the stimulus was frequently a tone with a frequency of 2.3-3.4 kHz, at intensities ranging from 82 dB SPL to 120 dB SPL at the real ear with a rise time between 0 and 20 ms. When hearing loss, defined as a rise in threshold, occurred, it resulted in slight dips in the region from 4 to 6 kHz.

A similar pattern of symptoms in 18 call centre workers claiming acoustic shock injury in the United Kingdom was reported by Lawton.[89] The hearing loss found in these two studies was not as great as that found by Palva, where the exposure was typically at a higher sound level, nor that reported by Beastall[90] or Guyot, which involved cordless telephones.[91]

Another study in the United Kingdom reported on 30 patients who had presented with acoustic shock.[92] The otological symptoms reported were very similar to those reported by Milhinch and Lawton, with tinnitus and pain being the two most frequently occurring symptoms. The number of otological symptoms reported in the 30 patients is shown in Figure 3-2.

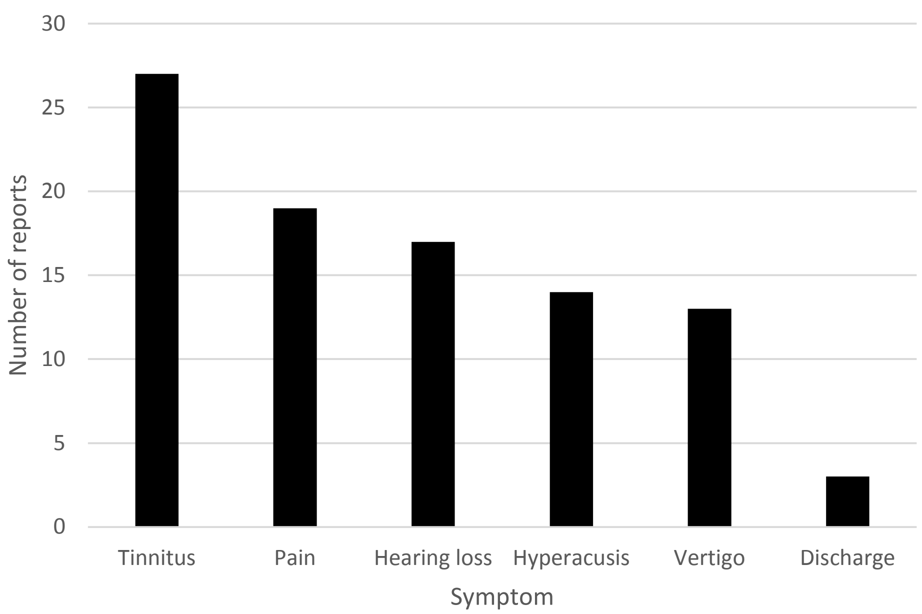


Figure 3-2. *Otological symptoms reported in 30 patients presenting with acoustic shock (adapted from Parker et al.[92]).*

This study found a greater occurrence of hearing loss compared to the findings of Milhinch and Lawton. In 12 cases, the hearing loss was determined to be sensorineural and, in six of these cases, a 4/6 kHz notch was evident. Slightly over half the cases were single incidents and the exposure was bilateral in 11 cases.

The sounds and their tally are shown in Figure 3-3. The most frequently reported sound was a squeal. The second most common was electrical clicking and the third was an alarm. It would appear that the commonly occurring offending sounds were either narrow-band harmonic series or impulsive in nature.

The main sources of the sounds were telephone headsets (22 patients, 73%), personal radio communication equipment, (four patients, 13%), and alarms, (four patients, 13%), with one patient wearing a hearing aid in the affected ear at the time of the incident.

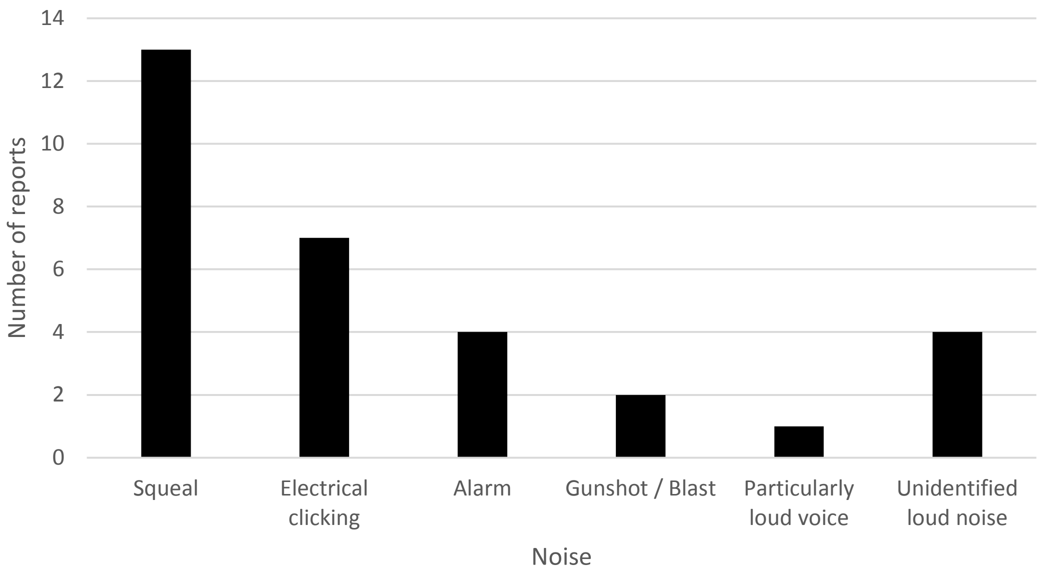


Figure 3-3. *The occurrence of sounds reported to cause acoustic shock in 30 patients (adapted from Parker et al.[92]).*

A multi-clinic study of the prevalence of the tonic tensor tympani syndrome in patients suffering from tinnitus and hyperacusis was conducted by Westcott et al.[93] Of the 345 subjects suffering from either tinnitus or hyperacusis, 68 (19.7%) were identified as suffering from acoustic shock. Hyperacusis was significantly greater in the acoustic shock group by a factor of approximately two ($p<0.001$). Acoustic shock was approximately twice as frequent in Australia/New Zealand compared to the other two countries participating in this study, Spain and Brazil.

3.9 Abrupt power increase, acoustic startle and shock

In the large majority of cases of acoustic shock, the stimulus has been reported to be either a mid-to-high frequency sound (e.g. the tones reported by Milhinch and the squeals and alarms reported by Parker et al.) or an impulsive sound (e.g. the electrical clicking reported by Parker et al.). Unlike an abrupt low-frequency sound, abrupt mid-to-high frequency sounds have a shorter rise time due to having a shorter period. This makes their initial power far higher than low-frequency sounds and more like impulsive sounds. As discussed, the integration time of the acoustic startle appears very short, i.e. 3 ms. It is therefore not surprising that the sounds with the most abrupt rise in power are associated with reports of acoustic shock. This correlation of data suggests, if not confirms, that methods of acoustic shock prevention need to prevent abrupt increases in sound (relative to the sound level being experienced) using a time constant that is faster than the integration time of the startle response.

3.10 Some industry and regulatory responses to acoustic shock risk

Most authorities dealing with telecommunications recognise acoustic shock as an issue that needs to be addressed in their standards. The European Telecommunications Standards Institute (ETSI) conducted a study, 'Acoustic safety of Terminal Equipment (TE): An investigation on standards and approval documents, ETSI TR 101 800 V1.1.1' to look at this issue.[94] It mainly reported on the specified maximum sound pressure limits in a large number of countries around the world. These were reported to be very high in all countries. The lower ones were in the order of 118 dBA at the ear reference point (ERP)[95] and many were considerably higher.

In Australia, the largest telecommunications company, Telstra found that reducing the sound level to approximately 85 dB SPL at the real ear, using a customised version of the Plantronics M10 limiting amplifier, could significantly reduce the occurrence and intensity of acoustic shock injury.[3] It could not, however, be completely eliminated, and those who had previously experienced an acoustic shock appeared to remain more susceptible to injury than others. The company also found that operators were having difficulty understanding telephone speech at this limited level in the typical background noise levels of the call centres (which were intended not to exceed 55 dBA but were often higher in practice). The call centre workers also complained about the poor speech quality resulting from the limiting adding distortion to the speech. As a result, the company developed an equipment specification focused on both the reduction of sounds reported to cause acoustic shock and provision of good speech intelligibility. As will be discussed in Chapter 4, alternative methods of processing were subsequently developed to improve speech quality and protection.

In Australia, the committee that sets the Australian Standard (*AS/CA S004:2013 Voice performance requirements for Customer Equipment*)[96] responded to the problem of acoustic shock through modifications to this standard and the production of guidelines designed to improve acoustic safety. The resulting document is *G616:2013 Acoustic safety for telephone equipment.*[97] It recommends that the lowest possible maximum limiting level be employed for the situation and that this maximum level should not exceed 102 dB SPL at the drum reference point (DRP).[95]

3.11 Summary

This chapter has presented the neurophysiological mechanisms believed to be involved in acoustic startle and shock, the symptoms that arise from it, the occurrences, and action taken by a large telecommunications company and standards body in Australia. It is hoped that, through the application of the SRL scheme that is the focus of this thesis, acoustic shock arising from listening to speech reproduction devices will be significantly reduced, if not completely eliminated.

Chapter 4

Sound level control methods

4 Sound level control methods

This chapter reviews the literature on methods of automatically controlling the level of sound produced by electronic devices that are primarily intended to convey speech, with an emphasis on their provision of acoustic comfort and safety. The review also considers the effect that such control has on the intelligibility of speech.

The control methods considered include:

- Peak clipping
- Limiting
- Compression

The applications considered are:

- Headphones
- Headsets
- Level-dependent hearing protectors
- Hearing-aids (and cochlear implants)

In relation to the above applications, research conducted into effects of automatic level control (ALC) methods is most comprehensive in the hearing-aid field. This review therefore draws principally on research into that field in examining some of the deeper technical issues. It establishes the range of applications where a scheme such as SRL can assist in providing improved acoustic comfort and safety for the listener.

4.1 Sound level control in electronic devices – the big picture

The majority of electronically reproduced sound that people hear has had its sound level automatically controlled. Members of the public are largely unaware that most sound recordings, film and television soundtracks and radio programs have been processed in this manner. They may be surprised when listening back to their personal sound recordings, such as when using a Dictaphone, by the lack of control of extraneous noises or by a similar lack of sound control when listening to a hands-free telephone call accompanied by the noise of the caller washing dishes.

There are essentially two different conditions in which ALC is used and these cause very different effects. They are:

1. ALC applied to the desired signal only (and not to the background noise)
2. ALC applied to all signals (both the desired signal and the less desirable noise)

These are important differentiations which largely indicate whether the speech-to-noise ratio and hence speech intelligibility will be enhanced or degraded through the application of ALC. The use of ALC in these two conditions is discussed in the following two sections.

4.1.1 ALC applied to the desired signal only

ALC is applied to most sound recordings, film and television soundtracks and radio and television transmissions. In production, ALC is applied to individual voices, instruments and sound effects, prior to mixing these sounds together, and to the final mix for overall level control and psychoacoustic loudness effect.[98] Due to the selective application of ALC, only the desired signals are increased in level. When reproduced by the consumer, the sound level is therefore more consistent, which helps to keep it above whatever background noise is present in the consumer's environment. This consistency increases the average speech-to-noise ratio resulting in an improvement in speech intelligibility provided the speech is not distorted by the ALC. The application of ALC to other sounds such as music have been found in one study to be well tolerated,[99] high levels of compression in another study have been found to be detrimental,[100] although awareness of the processing and preference varies between individuals, genre and the processing applied.[101]

ALC is applied by radio stations so the sound is at a consistent level within the listener's environment, such as when listened to against the background noise of a car, and to ensure it has a strong level when compared with other radio stations, in addition to technical requirements.[102,103] ALC is also applied by television stations for consistent level within the typical viewing environment but to a lesser extent than radio stations as there is an expectation of variation in the sound level (e.g. with drama productions).[104] There is, however, no doubt that advertisements are generally more compressed and amplified, creating a greater loudness than other program material.[105]

These examples show that, when carefully applied, ALC enhances the ratio of the speech to the noise in the listener's environment and as a result improves listening comfort, safety and speech intelligibility. While they illustrate how perception of speech can be improved through providing a consistently good signal-to-noise ratio, there is no evidence that ALC improves speech quality, nor any reason why it should.

A similar scenario applies to speech presented via a (tele)communications headset or handset or a public address (PA) system with the microphone in close proximity to the voice (e.g. a PA system in a passenger aircraft).[106] In each of these cases, the ALC is applied only to the desired speech (assuming there is no noise or other signal contaminating it, as will be discussed below) and improves the ratio of the speech to the noise in the listener's environment resulting in an improvement in speech intelligibility but not in the speech quality.

There are several additional issues related to (tele)communications headsets. Unlike the selected signals in the production cases (sound recordings, film and television soundtracks and radio programs), the signal in these cases comes from an uncontrolled source, a microphone in the other party's environment, and may therefore contain noise or other signals, as well as potentially uncontrolled signals from the (tele)communications network itself. These issues were discussed in Chapter 3. An additional concern in the headset case is the level of the background noise in the listener's own environment. It may be so high (such as in a call centre, a factory, or on an airport tarmac) that the user needs an excessive headset level to hear the speech, which despite the use of ALC, may increase the potential for injury from noise emanating from the headset.

4.1.2 ALC applied to all signals

In the following applications, the ALC is applied to all signals (the desired speech and the typically less desirable (background) noise):

1. Hearing protectors with level-dependent amplification
2. Hearing aids
3. Cochlear implants

In normal operation, the source of the signal for these devices is one or more microphones located on the device. As a result, speech and noise both come from the listener's environment in a combined form, although possibly with an improvement in their ratio due to directional processing. ALC is therefore applied to the noise as well as to the speech. Application of ALC results in co-modulation of the speech and noise and produces a detrimental effect.[107-109] In the absence of selective amplification (i.e. an enhancement of the spectro-temporal differences between the speech and noise), the level of lower-level noise will be increased as much as the speech by the application of ALC. Other than ensuring that all components of the speech are above the effective audibility thresholds of the individual (which may be raised due to noise reaching the ear via leakage or bone conduction in the hearing protector case or by hearing loss in the hearing aid case), there is no speech intelligibility advantage in further raising the level of the signal in this situation as the speech-to-noise ratio remains unchanged. In fact, there is potentially a disadvantage in raising it if the rise in level creates a greater masking of the speech or if the speech reaches such a level that the speech level distortion factor comes into play and reduces its intelligibility. Both these effects are well established and documented in the American National Standard, 'Methods for Calculation of the Speech Intelligibility Index' (the SII).[12]

There has been a considerable amount of research into the application of ALC to hearing aids, see Dillon (2012) and Souza (2002) for general overview and Moore (2008) for an overview of ALC speeds.[110-112]. A plethora of strategies have been developed and evaluated within the hearing-aid research field. Improvements in audibility and listening comfort have often been reported. Intelligibility of low-level speech has been shown to be improved by bringing the lower levels of speech above the audibility thresholds of the individual. Comfort has been shown to be improved by limiting high-level sounds. Many performance differences between the strategies have been demonstrated, however, in the vast majority of experiments, ALC strategies when compared with linear amplification with the volume manually adjusted by the user do not yield an improvement in speech intelligibility.

4.2 Methods of automatic sound level control

4.2.1 Device saturation

The default means of protecting a listener from excessive exposure to sound from an electronic sound reproduction device is the saturation of its electronics and in particular the saturation of the transducers (i.e. speakers and earphones) that convert the electrical signal to sound. Speakers and earphones have limits on the maximum physical displacement of their diaphragms and this restricts the maximum sound pressure they can produce. The displacement of the diaphragm generally follows the electrical driving signal level up to the point of maximum displacement, beyond which it cannot follow the signal and distortion of the sound occurs. The increase in the sound level resulting from a further increase in the signal level is typically no more than 3 dB for a sine wave. New frequency components, which did not exist in the signal, are produced and are normally considered unpleasant by listeners, as discussed in the next section.

The maximum sound level produced in the real ear by a saturating headphone can be very high, in the order of 125 dB SPL or more for some devices. Fortunately, many driving amplifiers will reach their maximum output before saturation of the headphone is reached.

4.2.2 Peak clipping

Peak clipping is the simplest, though also generally considered the poorest sounding, method of reducing the saturated sound level produced by an electronic sound reproduction device.[110,111,113,114] Because a peak clipper does not require active electronics, it can be inexpensively and easily applied. In telephone applications, it typically consists of diodes wired across the electrical connections of the reproduction transducer's coil or across the output of the circuit driving the transducer. The operating voltage of the diodes can be adjusted to clip the signal at a given voltage and therefore can control the maximum sound level the transducer produces. Distortion of the waveform results from clipping, and like the device saturation method previously discussed, it results in new frequency components being produced that did not previously exist in the signal and these are normally considered unpleasant by listeners.

Because peak clipping is a very harsh form of sound control, the level at which it acts is normally set to prevent only very high sound levels being produced as otherwise it would adversely affect speech and other desirable sounds.[113]

Peak clipping is also used in amplifiers for some headphones, headsets and hearing aids. It can be applied in the digital domain as well as in the analogue domain. However, as discussed, it is generally considered the worst sounding method of control and should therefore only be considered as a last resort for acoustic protection and should not be applied at levels that could affect speech. These limitations mean that, with peak clipping, high-level noise signals will still be able to significantly exceed normal listening levels (i.e. comfortable speech levels), meaning there is a risk of acoustic shock injury. This is the case with telephone headsets when not used with additional protection as discussed in Chapter 3.

4.2.3 Limiting and compression

Limiting, or compression limiting as it is often termed, is a method of automatically reducing the level of the signal that will (eventually) drive a sound reproduction transducer. It is a form of compressing (i.e. squashing) the changes in the sound level which is characterised by a high compression ratio and typically has a fast reaction to an increase in the signal level above a reference level, known as the limiting threshold (LT) or compression threshold (CT).

The compression ratio (CR) is the ratio of the change in the input level to the change in the output level. A limiter is typically set to have a CR greater than 8:1 (i.e. an 8 dB change in the input results in a 1 dB change in the output).[115]

A single-band limiter is the simplest form of limiting. The level of the signal is estimated, often without weighting the frequency components, although in some situations it may be weighted.[102] A simple estimator typically has two response times: the speed with which its estimate increases with an increase in the signal level (the attack time) and the speed with which its estimate decreases with a decrease in the signal level (the release time).

The estimate of the signal level is compared with the CT/LT. The excess level of the signal, determined by the amount by which the estimated level exceeds the CT/LT, is then reduced by the amount specified by the CR to produce the level controlled signal. With a limiter, the CR is often set to ∞:1. The effective compression ratio, however, will be different to this, due to the response times of the estimator and the method used to measure the controlled signal. The effective attack time (not the envelope attack time but the time to reduce an increase in the signal level) and the effective release time (not the envelope release time but the time to increase the level following a decrease in the signal level) also depend on the signal levels, the CR and the CT. For hearing aids, there are standards for determining these compression effects, including the reaction times, such as those specified by the *American National Standards Institute (ANSI), S3.22-2003, Specification of hearing aid characteristics.*[115]

4.2.3.1 Single-band limiting plus wide dynamic range compression (WDRC)

This is the same method as described above for a single-band limiter combined with a WDRC (a compressor having a lower CT, a lower CR and slower time constants than the limiter). The WDRC may be multi-band, where the signal is first split into a number of bands and separate compressors are applied in each band and then recombined prior to limiting. For example, an experimental hearing aid with a two-band WDRC developed by Moore and Glasberg (1988) uses this approach.[116]

4.2.3.2 Multiband limiting

The signal is split into a number of frequency bands and a limiter is applied to each band. After the application of limiting the bands are combined to create the limited signal. The limiting of the band signals is linked in some devices to constrain excessive changes to the spectral balance.[102] A similar but alternative approached to this uses a single adaptive filter to apply different degrees of limiting at different frequencies.[117,118]

4.2.3.3 Multiband limiting plus wide dynamic range compression (WDRC)

This is the same method as described immediately above but is combined with a single or multiband compressor with lower CT(s), lower CR(s) and slower time constants than the limiter. See Dillon (2012) and Souza (2002) for a general coverage of multiband schemes in hearing aids.[110,111]

Some multiband schemes use percentile estimates of the sound levels in order to adjust the gain in multiple bands. One such multiband scheme using percentile estimates of the input signal was developed for hearing aids by Ludvigsen (1997).[119] Another using the percentile estimates of the output signal in a scheme called Adaptive Dynamic Range Compression (ADRO) was developed by Blamey et al. (2005).[120,121]

4.3 Applications of ALC to acoustic comfort and safety

4.3.1 Headphones

In most cases, the material people listen to through headphones is already compressed and limited and is therefore unlikely to contain abrupt changes in sound levels that would cause discomfort or injury when used with most consumer devices. There is, however, a concern about long-term exposure causing hearing damage. One product that addresses this in consumer devices is a broadband level-adjustment device called the Limitear developed by Glover (2012).[122] The device can draw its power from the headphone audio signal to reduce the signal applied to the headphones over time, thereby reducing the exposure.

Unlike the consumer, professionals working with audio (such as radio station panel operators, disc jockeys, recording studio engineers, stenographers and transcriptionists) are likely to be exposed to uncontrolled sound, from equipment malfunction, operator error and extraneous noises in recordings made in the field. To deal with this, a number of professional headphone limiters have come on the market, such as the headphone limiter developed by the British Broadcasting Corporation, manufactured by Canford (UK). This has an adjustable limiting threshold but contains no compensation for specific transducers. Limiters have also been installed on headphone circuits for court stenographers and medical transcriptionists. Unfortunately, there appear to be no published studies on the effectiveness of these devices. There is, in general, a lack of scientific studies on the effects of commercially available protection equipment on people. This comment doesn't just relate to headphones, this applies to most equipment in this field. There are, of course, many carefully worded marketing claims showing no data.

4.3.2 Headsets

As discussed in Chapter 3, telephone headset wearers are at significant risk of acoustic shock injury. The telephone system is largely uncontrolled in terms of the signals it conveys through its audio channels.

4.3.2.1 Built-in peak clippers for headsets

Because of the uncontrolled nature of the telephone audio signal, most telecommunications headsets contain diodes to peak clip the signal so that a specified maximum sound level is not exceeded. The maximum level specified is typically 118 dBA SPL at the ear reference point (defined in *International Telecommunications Union Standard ITU-T Recommendation P.57 Artificial ears)*,[95] although standards specifying maximum levels and the measurement of them vary between countries.[94] This is a very high sound level to be exposed to and is likely to be very uncomfortable for a listener with normal loudness sensitivity. As discussed in Chapter 3, this level of protection has been found not to prevent acoustic shock injury.

4.3.2.2 Limiting amplifiers

Some telecommunications headsets require an amplifier between them and the telephone, or telephone console as it is often termed. Most headset amplifiers also contain some type of ALC. Both single-band and multiband limiter strategies have been used and some devices also incorporate WDRC or a slow-acting automatic volume control as well as expansion at low levels to reduce telephone line noise.

The Plantronics M10 and M12 limiting amplifiers are examples of fast-acting compression limiting and downward expansion using analogue electronics.[123] These devices do not have compensation for the sensitivity or frequency response of the headsets they are used with, meaning the limited sound level varies depending on the sensitivity and frequency response of the headset. This is also the case with the single-band analogue device from GN Netcom, the MPAII.

The first acoustic shock protection device was a digital device that used a Texas Instruments' floating-point digital signal processor. It contained a number of novel processing strategies. To deal with the variation in sensitivity between headset types, it employed a method, called headset-tailored limiting, which corrects for the response of headset type to provide more controlled limiting levels. This method was developed by the author of this thesis and was licensed for use in the SoundShield™ device as part of a package of strategies including shriek rejection (which is discussed in the following section).

The Plantronics M15D and the GN Netcom 8210 are examples of digital devices containing advanced ALC strategies designed to address the issue of acoustic shock. The M15D uses the ADRO[121] multiband processing strategy implemented in the digital signal processing chip developed by the DSP Factory.[124] The GN Netcom device uses a multiband compression / limiting strategy implemented in a Texas Instruments' fixed-point digital signal processing chip. The advantage of these multiband limiting strategies over single-band strategies is that the power of the speech is spread over a number of bands and therefore has a lower level in each of the bands. This enables the limiting thresholds to be lower in the bands for a given broadband speech level. As a result of having lower thresholds in the bands, narrow-band sounds are limited to a lower level and therefore greater protection is provided against narrow-band sounds. Neither, of these devices, however, contain correction for the headset type.

4.3.2.3 Shriek rejection

The Telstra experience, as discussed in Chapter 3, showed that headset limiters were not able to simultaneously provide protection from acoustic shock and adequate speech quality and intelligibility in the context of call centre background noise. As discussed in Chapter 3, the sounds that were reported to cause acoustic shock were often narrow-band high-frequency sounds, such as tones produced by faults within the telephone network, whistles from malicious callers, feedback from cordless telephones, misdirected fax machines and so on. The high pitch of these sounds resulted in them being called 'shrieks'.[2] To address this problem, a solution other than fixed-reference limiting was required.[4]

It was noted by the author of this thesis that many of the sounds reported to cause acoustic shock were spectrally different enough from the speech that an algorithm could be developed to identify and reject them without reducing the level of the sound in other frequency regions. Speech could therefore be almost entirely preserved while one of these offensive sounds was simultaneously rejected. An algorithm was developed that analysed the spectrum of the signal on an ongoing basis looking for high-level narrow-band energy that persisted over a minimum period. The algorithm applied frequency-agile notch filters to the signal to remove these sounds while preserving the sound at other frequencies. The author called the method *shriek rejection*. It was patented[5] by the author and licenced to the company Polaris for manufacture and sale under the name SoundShield™.

4.3.3 Level-dependent hearing-protectors

Level-dependent hearing-protectors are often used instead of devices without amplification in situations where the wearer wishes to hear softer sounds in between loud sounds. The softer desirable sound is often speech. These devices can, for example, assist workers on a building site who converse between using loud tools.

There are a number of level-dependent hearing protectors on the market. They are designed so that the long-term exposure from the amplified sound is below a given limit, such as 85 dBA. One device (Peltor Protac II), investigated by the author of this thesis, limits the sound exposure to about 82 dBA $_{\text{diffuse field equivalent}}$ when measured on a HATS acoustic mannequin.[125] It uses a single-band fast-acting compressor limiter with a soft knee making its performance what could almost be termed WDRC. It limits speech and noise alike. As expected the speech quality is affected by the protection processing due to the fast-acting compression limiting. Co-modulation of desired speech and competing sounds is clearly present. More advanced strategies have been developed using multiband approaches to reduce noise using active noise cancellation. Multiband strategies have also been used to limit exposure from both the field and a communications channel.[126,127] Concerns about degradation in the wearers localisation performance from using independent ALCs in each ear processor have arisen, which may be addressed by linking the ALC of each ear processor.[128,129]

4.3.4 Hearing aids

As discussed in Section 4.1.2 and Section 4.2, there are a great number of ALC strategies used in hearing aids. Methods of limiting were discussed but not the criteria for setting the limiting level. The most evaluated method for setting the maximum power output (MPO) is the National Acoustic Laboratories' Saturated Sound Pressure Level prescription, NAL-SSPL.[113] It is based on data from both normal hearing and hearing impaired listeners. The chosen MPO lies midway between the maximum output level to avoid discomfort and the minimum output level needed to avoid excessive saturation of speech with an average level of 75 dB SPL.

4.4 Summary

There are essentially two different conditions in which ALC is applied. In the first condition, ALC is applied to the desired signal only, and in the second condition, ALC is applied to the desired signal and the less desirable noise together. In the first condition, the ALC can improve the speech-to-noise ratio and hence improve speech intelligibility, but generally this cannot occur in the second condition, except through reductions in off-frequency or temporal masking of speech, or reductions in the speech level distortion factor.

There are many first condition situations where the audio signals have been very carefully controlled by ALCs (and/or manually controlled by professionals) so when provided to the listener, e.g. the consumer of professionally produced audio/sound, they are unlikely to cause discomfort or injury.

There are other first condition situations in which the signals are uncontrolled and require automatic processing in order to be comfortable and safe. These include: headphone listening, such as performed by sound engineers, broadcast operators, stenographers and transcriptionists, and headset listening, such as performed by call centre operators, two-way radio operators, airport ground crew and so forth. In these situations, an increased speech level through volume control adjustment or ALC can increase the received speech-to-background-noise ratio and hence improve speech intelligibility. The increased level, however, leaves the listener more open to experiencing discomfort and injury from noise within the signal and hence an intelligent ALC is required.

There is the second condition in which less desirable noise is contained within the signal carrying the desirable speech, such as when using level-dependent hearing protectors or hearing aids. The sounds presented in these applications are uncontrolled and require ALC to be comfortable and safe. Unlike the first condition in which the speech-to-noise ratio could be improved through ALC incorporating an increase in level, the ALC, in this second condition, is required to use the spectro-temporal difference between the speech and the noise to reduce masking of speech and control the speech level distortion factor in order to improve speech intelligibility. At the same time, the ALC should not degrade the speech intelligibility through co-modulation of the desired speech and the less desirable noise.

Device saturation and peak clipping are very poor forms of acoustic protection, ALC strategies that dynamically adjust the amplification gain are preferable. In general, multiband ALC strategies offer better control than single-band ALC strategies in terms of comfort and safety. This is particularly true in terms of controlling narrow-band noise. Additional methods of potentially harmful noise reduction such as shriek rejection offer an improvement over this in particular applications, such as headset protection. Shriek rejection directly addresses the need for noise suppression with minimal sacrificing of the speech but it only works for a specific set of sounds. This Chapter has shown a number of weaknesses in the current methods available for providing loudness comfort, acoustic safety and good speech intelligibility. These and the issues presented in Chapters 2 and 3 are addressed with the novel solution presented in the next chapter.

Chapter 5

Speech referenced limiting – in theory

5 Speech referenced limiting – in theory

This chapter introduces the novel speech referenced limiting (SRL) concept and describes the SRL scheme in general terms. The concept it describes is the foundation on which the implemented and evaluated schemes, SRL MKI and SRL MKII, were based.

In Chapter 2, variability in loudness perception both within and between individuals was discussed. This variability in loudness perception was shown to be substantial, particularly, in relation to the sound level at which loudness discomfort occurred. In Chapters 3, the issues of somatic response to sound and acoustic shock were discussed. The variability in individuals' physiological and psychological response to abrupt changes in sound and the sounds that trigger adverse reactions were considered. In Chapter 4, established methods to prevent loudness discomfort and acoustic shock were considered and the limitations of these methods discussed. These chapters provide the background to the novel approach presented in this chapter: that is, the use of the characteristics of speech, to which the listener is acclimatised, as a reference for controlling other sounds.

Speech is arguably the sound most listened to by humans. There would appear to be no better reference to use for loudness perception than the loudness perception of speech. This chapter introduces the novel concept of controlling sounds with reference to speech, which I call speech referenced limiting. This approach aims to provide loudness comfort and prevent adverse somatic responses, i.e. acoustic shock, for people primarily listening to speech reproduced by electronic devices. The concept is described in the following section.

5.1 The speech referenced limiting concept

Conventional limiters limit signals that exceed a fixed level and I therefore define them as 'fixed-reference limiters' (FRL). Figure 5-1 shows an input/output plot of amplification that includes a fixed limiting level.

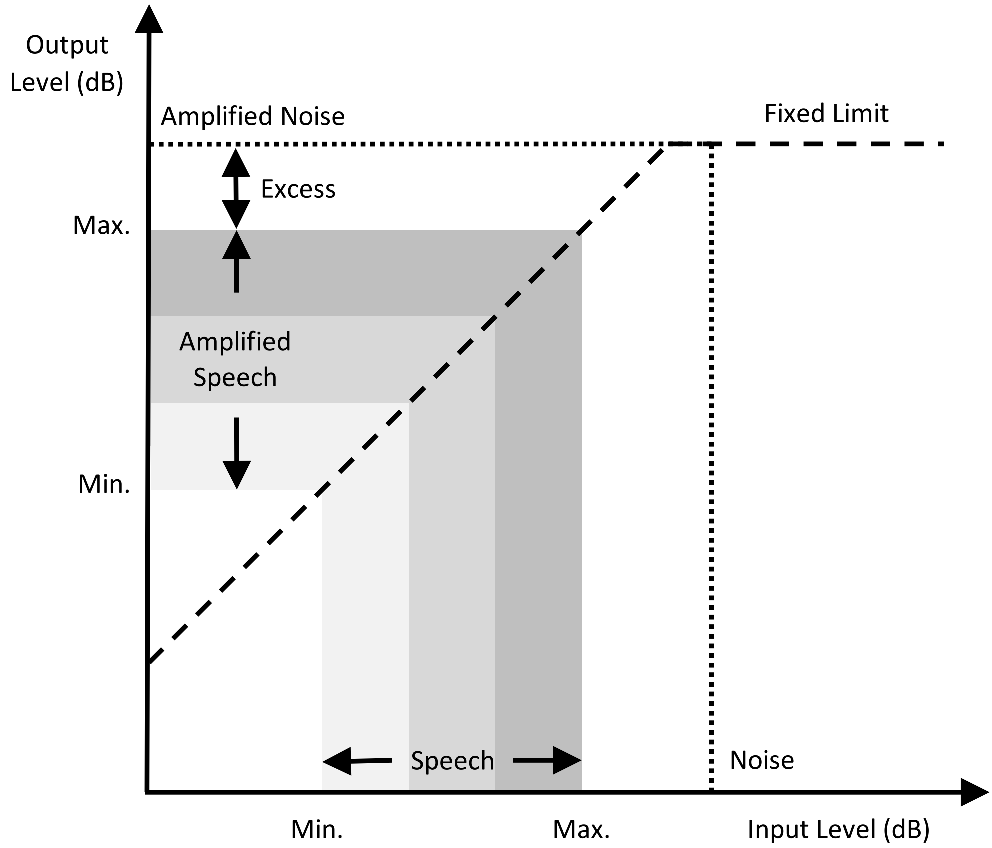


Figure 5-1. *Input/output plot of amplification with a fixed limiting level in excess of the maximum speech level.*

The dynamic range of the input speech is shown and this is mapped through amplification to produce the amplified output speech dynamic range. In the case depicted in this figure, the limiting level is greater than the maximum level of speech and therefore no limiting of speech occurs. A noise at an input level greater than the maximum level of speech is shown. While the noise receives amplification, it does not receive as much amplification as the speech due to limiting. However, the amplified noise level exceeds the maximum level of speech by the excess shown. If the amplified speech is at a comfortable level, then the amplified noise will be at a less comfortable level.

Figure 5-2 shows a similar input/output plot to Figure 5-1. The amplification, however, differs in that the fixed limiting level has been reduced to below the unlimited maximum level of amplified speech. The input noise, although having greater amplitude than the maximum level of the input speech, has undergone greater limiting, resulting in the amplified noise level not exceeding the maximum level of amplified speech.

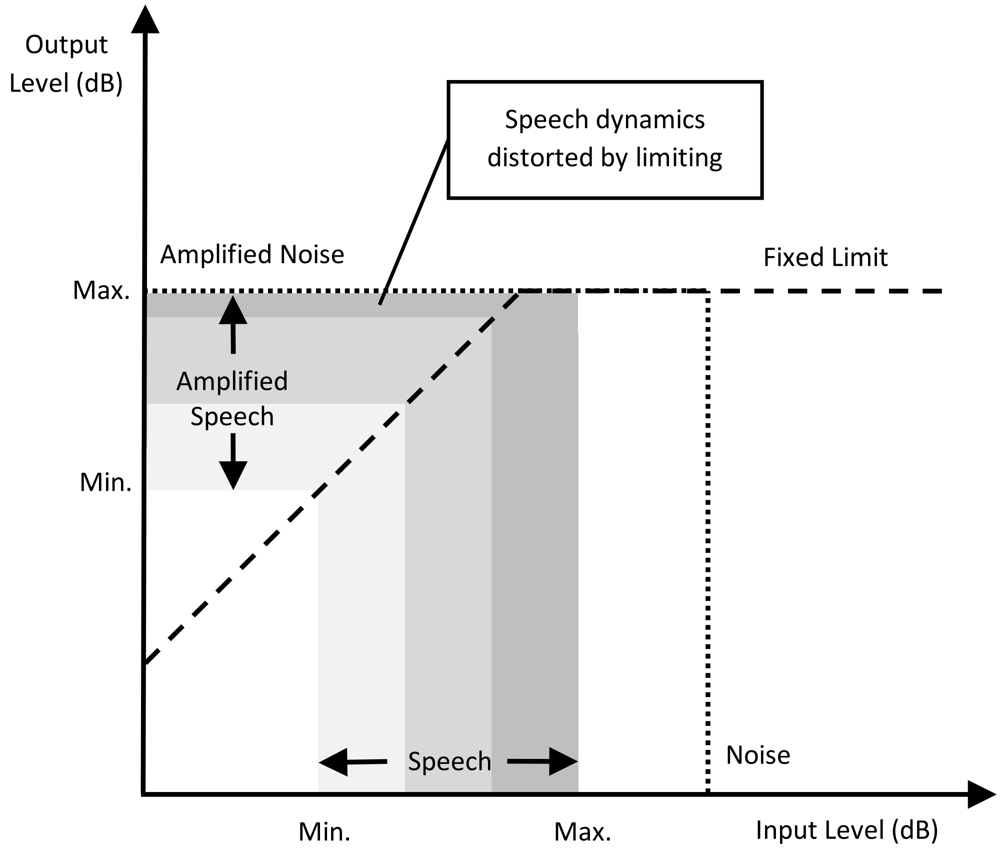


Figure 5-2. *Input/output plot of amplification with a fixed limiting level at a lower level than the maximum speech level.*

This control, however, has been at the expense of distorting the dynamics of the amplified speech, through compressing its dynamics and limiting its maximum level; this is reported to result in degradation of the perceived speech quality and intelligibility,[130,131] as was discussed in Chapter 4.

Figure 5-3 shows a similar input/output plot to Figure 5-1. The amplification, however, differs in that the limiting level is variable and is dependent on the maximum level of the speech. As the reference for this variable limiting level is the maximum level of the speech, I call this 'speech referenced limiting' in contrast to 'fixed-reference limiting'. In theory, the speech is not limited because the limiting level is equal to the maximum level of speech. Non-speech signals of a level greater than the speech, such as the noise level shown, will be reduced to the speech level. The excess noise level shown previously in Figure 5-1 has been removed. The variable limiting level tracks the maximum level of the speech as depicted.

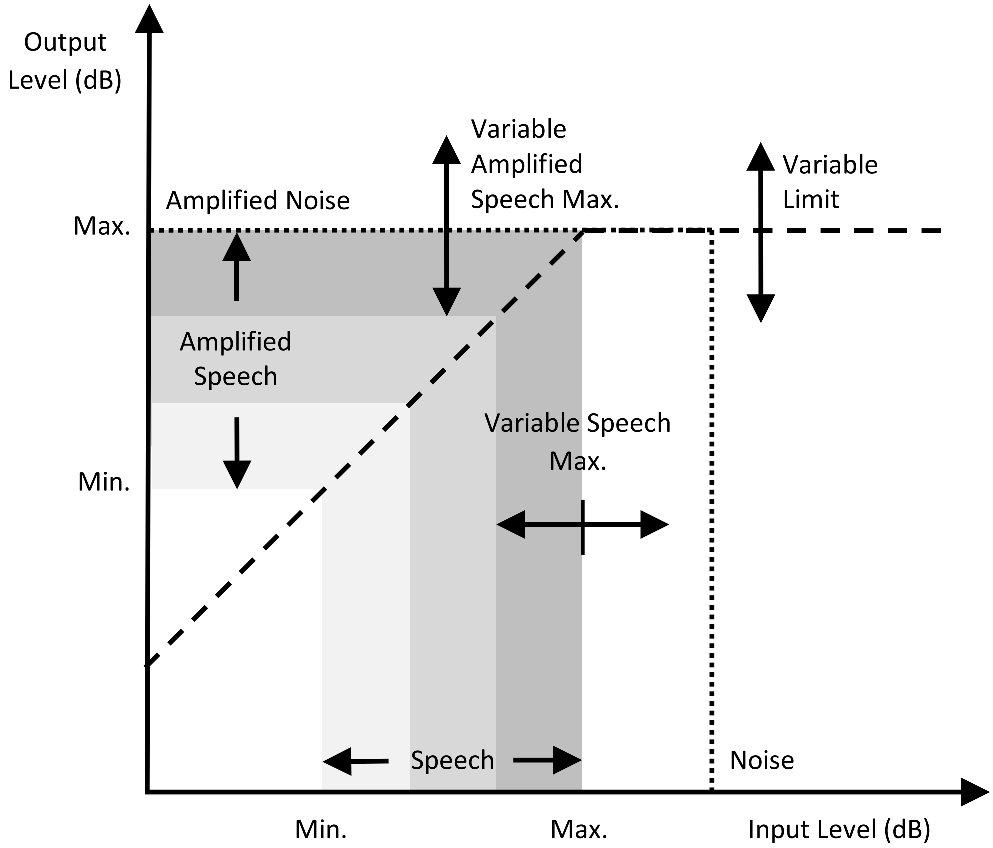


Figure 5-3. *Input/output plot of amplification with a variable limiting-level dependent on the maximum speech level.*

The single-band control mechanism illustrated in Figure 5-3 may be applied to a multi-band system in which the signal is controlled in a number of frequency bands. This means the frequency-specific loudness of a signal may be controlled with reference to the frequency-specific loudness of speech measured in an array of narrow-band filters using the appropriate bandwidths, such as those given in the *International Standard ISO 532B*,[47] based on earlier versions of the *International Standard 61260:1995, Electroacoustics — Octave-band and fractional-octave-band filters*,[132] or the critical band filters of Zwicker,[45] or the equivalent rectangular bandwidth filters derived by Moore and Glasberg.[46] An illustration of the multi-band control of a non-speech signal is shown in Figure 5-4. In frequency bands 10 to 20, the levels of the non-speech signal exceed the speech reference levels and are reduced as indicated by the arrows; the amount of gain reduction, as indicated by the length of the arrows, equals the amount of excess.

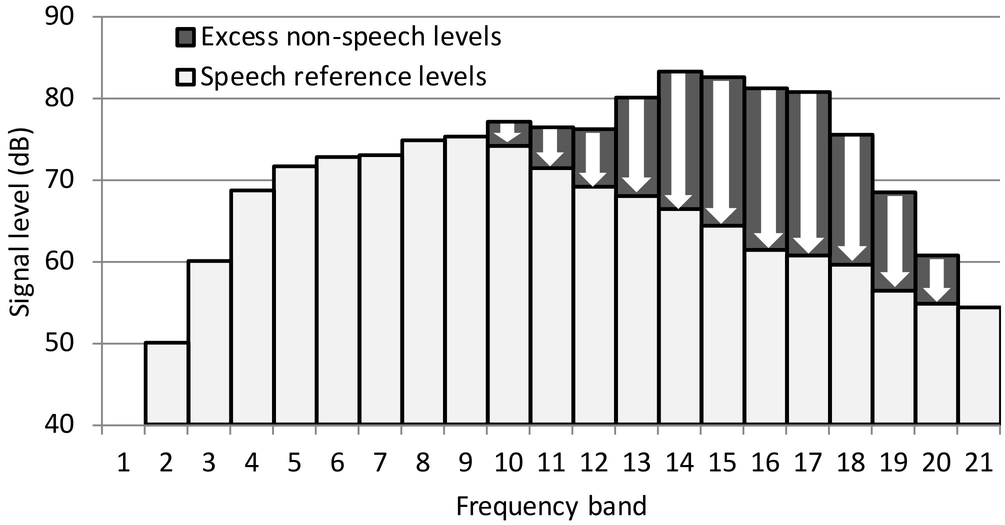


Figure 5-4. *Reduction in the levels of a non-speech sound with reference to speech levels in multiple bands.*

In theory, the use of the maximum levels of speech in bands as reference levels for limiting means there is no limiting of speech in the absence of noise. Limiting of speech only occurs when the in-band noise level is significant in relation to the in-band speech level, i.e. greater than approximately -3 dB relative to the maximum speech level. With the noise at this level the maximum level of the combined speech and noise signal within the band exceeds the speech reference level by approximately 2 dB and the limiting starts to become significant. If the noise that causes the limiting has an in-band maximum-to-mean power ratio of 5 dB, typical of many machines and multi-talker babble and the speech has a typical in-band maximum-to-mean power ratio of 15 dB,[63] then the ratio of the mean speech power level to the mean noise level will be around –7 dB. At this speech-to-noise ratio, the contribution to intelligibility of speech in this band will be less than 27% of its potential when there is no noise using the American National Standards Institute's Methods for calculation of the speech intelligibility index (SII).[12] The calculation is given by Equation 5-1.

$$K_i = \frac{(E'_i - D_i + 15)}{30} \qquad \textbf{(5-1)}$$

where: E'_i *is the equivalent speech spectral level*

D_i *is the equivalent disturbance level*

i *is the band number*

In this case, the equivalent speech spectral level E'_i is 7 dB below the equivalent disturbance level D_i and hence the band audibility, excluding the speech level distortion factor, is 8/30 or 0.27.

Reduction of the band level as the in-band noise increases has little to no effect on its in-band speech intelligibility, provided the speech remains above the absolute hearing threshold and above any off-frequency masking threshold in this frequency region. This is because the gain reduction does not change the in-band signal to noise ratio. However, the reduction in the band level reduces the masking of speech in other bands and therefore, in theory, improves the intelligibility of the speech. Because the other bands in which there is significant noise are also being reduced in level, any off-frequency masking within this band is also reduced. Although the scheme does not directly try to optimise the SII, unlike a scheme developed by Kates,[133] its reduction of bands with significant noise partially achieves this.

By using the maximum levels of speech as reference levels for limiting, the scheme, in principle, provides *the greatest limiting of noise for the least limiting of speech* making it arguably the optimal method for limiting high-level noise in speech systems. To my knowledge this approach has never been attempted or considered before.

5.2 The SRL scheme

A schematic diagram of the concept for the SRL scheme is shown in Figure 5-5. The input signal is modified adaptively by the adaptive modifier to produce the output signal. The adaptive modifier controls the level of the signal on a frequency-specific basis. The input signal is also passed to a pair of loudness estimators, the current loudness estimator and the speech loudness estimator. These estimate the frequency-specific, total and the 'peak' loudness of the input signal. The current loudness estimator estimates the loudness of the input signal continuously, while the speech loudness estimator estimates the loudness of the input signal only when it is dominated by speech. The input signal is also passed to a speech dominance detector which analyses the signal to determine if it is dominated by speech. When the speech dominance detector detects a dominant speech signal, it provides an instruction to

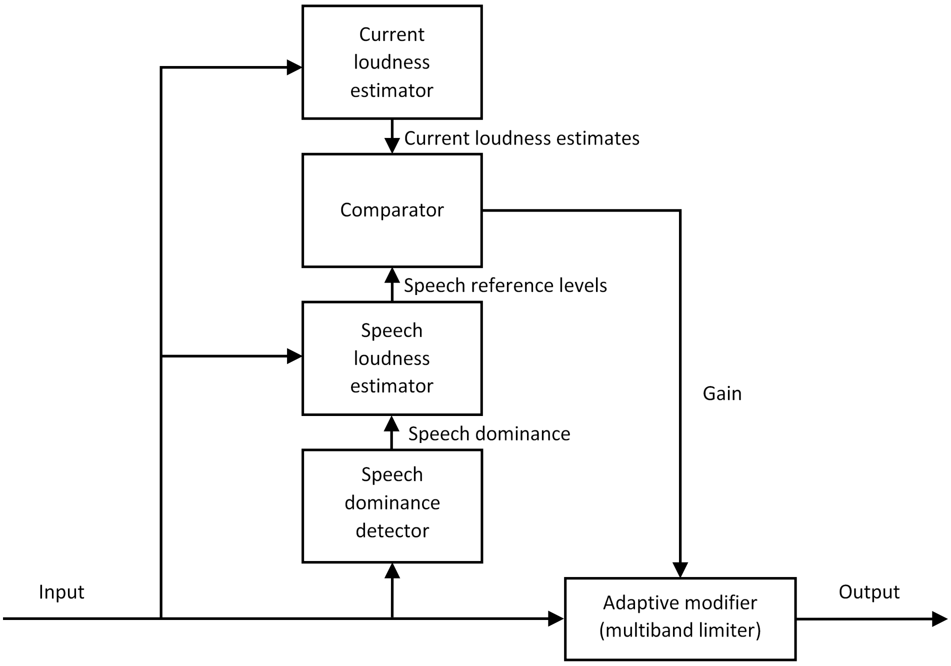


Figure 5-5. *The simplified SRL scheme.*

the speech loudness estimator to update its loudness estimates with recent estimates of the frequency-specific, total and 'peak' loudness from which an updated set of speech reference levels is produced. The current loudness estimates of the signal are compared with the speech reference levels by the comparator. If they exceed the speech reference levels, then reduced gain values are produced. These gain values are passed to the adaptive modifier, which reduces the amplification applied to the signal by the amount specified and hence reduces the excess.

Some of the analysis processes performed to produce the current loudness estimates and the speech loudness estimates are common to both. These include frequency analysis and power estimation. These can be performed by an analyser that precedes the two estimators. This modification to the conceptual SRL scheme is shown in Figure 5-6. The processes contained within the grey box (current loudness estimation, speech loudness estimation and comparator) constitute the complete gain calculation process.

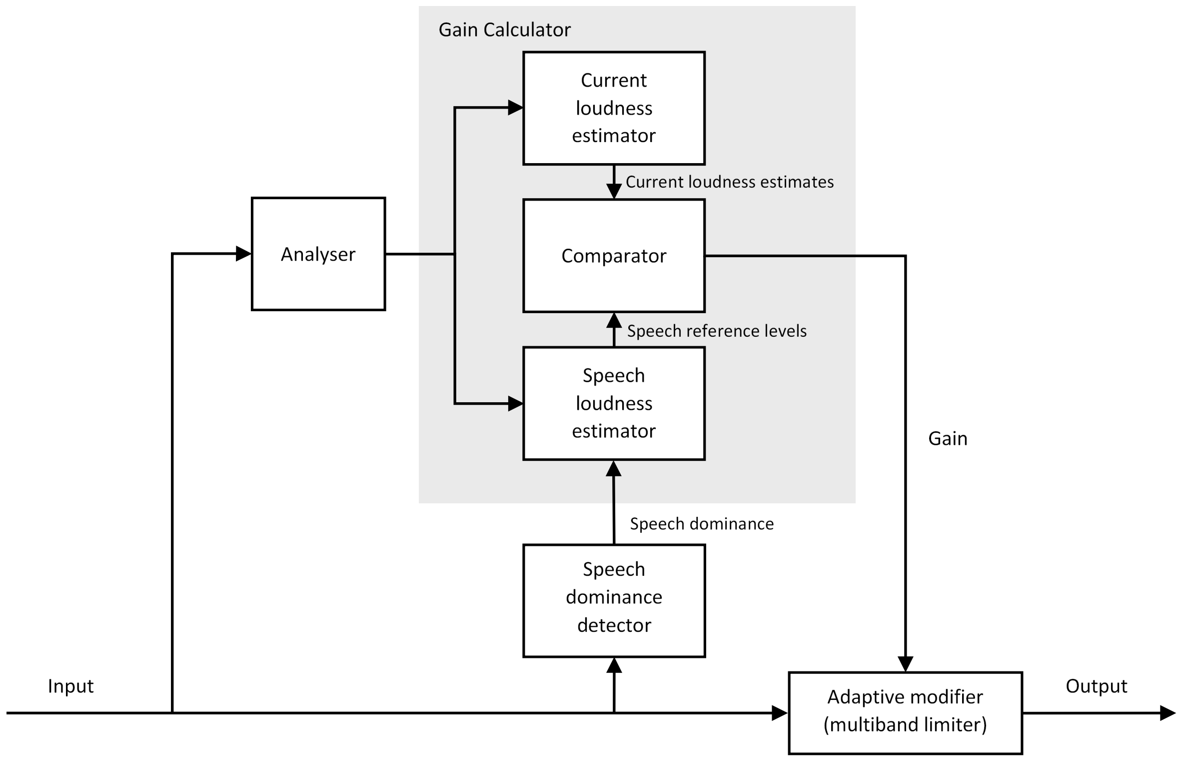


Figure 5-6. *The simplified SRL scheme with a separate analyser.*

5.2.1 Loudness estimation

The estimation of the loudness of time-varying signals is an area in which there has been considerable investigation by many researchers; this has been discussed in Chapter 2. Many methods of loudness estimation are available, from very simple estimators to sophisticated models. In designing this scheme, I have been mindful that different loudness estimation approaches might be necessary across the many potential applications for this scheme. For some applications, the signal processing resources are limited and only simple loudness estimators can be employed, while for others powerful processors are available and more sophisticated models can be utilised. For this reason, I have developed a scheme that can accommodate both.

At its most basic, an approximation of short-term loudness is simply the short-term integration of the frequency-weighted sound power, e.g. the A-weighted sound level meter defined in the *International Standard IEC 61672-1:2013. Sound level meters - Part 1: Specifications.*[56] Similarly, an approximation to the frequency-specific, short-term loudness is simply the short-term integration of the sound power in auditory-based frequency-bands. The bandwidth of these bands may simply be based on a logarithmic scale approximation such as one-third of an octave as given in the International Standard 61260:1995, Electroacoustics — Octave-band and fractional-octave-band filters,[132] or the critical band filters of Zwicker[45] or the equivalent rectangular bandwidth filters derived by Moore and Glasberg.[46]

As the SRL scheme uses the estimated relative loudness between the current signal and its past speech, it is unnecessary to estimate the absolute loudness of the signal in order to perform speech referenced limiting. This is one of the great advantages of SRL over conventional limiting schemes. It enables protection to be provided without knowing the characteristics of the reproduction system. It is, however, important that the method of estimating the loudness of the current signal and estimating the loudness of the speech are the same, since otherwise they cannot be meaningfully compared.

While it is generally desirable to control the relative loudness within bandwidths comparable to those of the human auditory filter, it is not strictly necessary to do so. There can be some advantages in using other frequency resolutions. For example, in the telecommunications field, high-frequency, narrow-band signals have been reported to be a common cause of acoustic shock by Milhinch[2] and others, as discussed in Chapter 3. By employing frequency resolutions finer than the auditory bandwidth at high frequencies, the specific loudness of the speech in each band will be lower due to it being distributed over more bands. However, the specific loudness of a narrow-band signal within the finer bandwidth of the band will remain unchanged, assuming it remains within one band. It therefore will be further suppressed and will appear softer than the specific loudness of speech in this frequency region; and will be less likely to cause a somatic response. This may be of benefit for people who are suffering from an acoustic shock injury and, as a result, have developed a hyper-sensitivity to high-frequency, narrow-band sounds, as discussed in Chapter 3. Another possible advantage is that greater computational efficiency may be obtained using non-auditory frequency resolution, such as linear spaced frequency bands.

A simple form of loudness estimation is shown in Figure 5-7. The input signal is applied to a filter bank which splits the signal into K frequency bands. There are many filtering techniques that can be used to separate the signal into a number of frequency bands, including infinite impulse response (IIR) filter banks, finite impulse response (FIR) filter banks, wavelets and discrete Fourier analysis. Crochiere and

Rabiner describe many of these techniques in their book of multi-rate DSP.[134] Each filtering method has its advantages and disadvantages in terms of computational efficiency, precision, stability, delay and phase. In particular, the Gammatone filter bank[135] which is based on the ERB has a performance which is well matched to masking data.[136] The SRL scheme may be incorporated within other systems that provide their own filter bank; accommodation for this arrangement is shown in the second conceptual SRL schematic, Figure 5-6, where there is a separate analyser block.

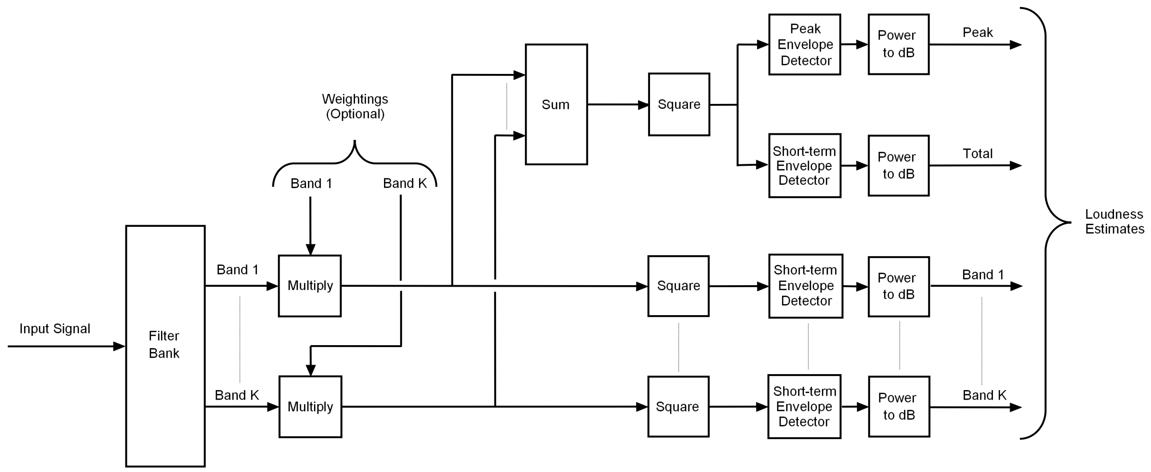


Figure 5-7. *Current loudness estimator.*

The band signals produced by the filter bank are optionally multiplied by weightings. Weighting is not essential for the system to perform the process of speech referenced limiting as it does not change the difference between the specific loudness estimates of the current signal and the speech reference levels. However, these weights can help to make the difference between the total loudness estimate of the current signal and its speech reference level, and the 'peak' loudness estimate of the current signal and its speech reference level, more likely to be indicative of those differences experienced by the listener. To achieve this, the weighting at the band centre frequencies should include the transfer function of the reproduction system including its transducer response at some defined listening point. For head/ear-worn devices this may be the ear reference point (ERP) or the drum reference point (DRP)[95] measured in the appropriate coupler[137] or simulator[138] respectively.

The weights should also include any unaccounted for outer and middle ear average static responses such as those responses provided by Moore et al.[24] It is, however, not necessary to know the absolute transfer function, i.e. in sound pressure level, SPL to obtain reasonable estimates of the weighting needed for the total and 'peak' estimates; the system merely needs to know the frequency response. If, however, fixed-reference limits are used in addition to the speech reference limits then the weights should account for the absolute transfer function from the digital level to the acoustic level. The weights may need to be estimated in some cases where there is great variability in the measured data or such data is unavailable. For example, if the SRL scheme is applied to a system that produces sound in the field, and the position of a listener varies, an average estimate will be more appropriate. The weighted band signals are combined by the summation block as shown to produce a total signal. The weighted band signals and the total signal are squared to produce the band power signals and the total power signal respectively.

The band power signals are applied to short-term envelope detectors to produce envelope signals for the bands. These detectors should provide an approximation to the frequency-specific short-term perception of loudness produced by the auditory system. There may, however, be an advantage in the rate that they integrate loudness being slightly faster than that of the human, in order to control the sound in a slightly pre-emptive manner. Integrating the signal over the short-term provides short-term envelope detection and uses an exponential averaging technique commonly referred to as 'leaky integration'. As discussed in Chapter 2, loudness integration rates are level-dependent and also depend on the bandwidth of the stimuli. A number of different integration rates and time constants are in use. One such short-term time constant is that defined in the *International Standard IEC 61672-1:2013, Sound level meters - Part 1: Specifications*.[56] The standard specifies a time constant of 125 ms which approximates loudness integration rates found in some earlier psychoacoustic experiments as discussed in Chapter 2. Some psychoacoustic experiments have suggested that the integration times should be faster with increasing frequency. However, more recent experimental data shows no significant dependence on frequency.[48] Typically, loudness integration time constants are in the order of 100 ms as discussed in Chapter 2.

One form of short-term envelope detection is the 1st order IIR low pass filter or 'leaky integrator' with switchable coefficients. These coefficients determine the time constants and are switched depending on whether the current sample of the input power applied to the envelope detector is greater or smaller than the previous calculated envelope sample. If the input sample is greater than or equal to the previous calculated envelope sample, then an attack coefficient and its corresponding input scaling factor are selected to be the a1 and b0 coefficients of a standard first-order filter respectively, where b0 = 1 – a1. Otherwise, a release coefficient and its corresponding input scaling factor are selected to be the a1 and b0 coefficients of the filter respectively. The envelope signal resulting from the leaky integrator increases exponentially at a rate determined by the attack coefficient when the input sample is greater than or equal to the previous calculated envelope sample. Otherwise the envelope decreases exponentially at a rate determined by the release coefficient.

The total power signal is applied to two envelope detectors - a peak envelope detector and a short-term envelope detector - to produce a peak envelope and a total envelope respectively. The envelope detectors are identical to the short-term envelope detector described above except that the peak envelope detector employs faster time constants. The attack time constant of the peak envelope detector needs to be as fast if not faster than the integration time constant of the startle reflex. There is of course no perception of 'peak' loudness, however, the startle reflex responds to the same neural firing provided to the ventral cochlear nucleus (VCN) by the cochlea via the auditory nerve, i.e. the 'instantaneous' frequency-specific loudness. This is presumably integrated over frequency (to some degree) and time, possibly by the VCN, using a very fast time constant, to create a neural signal that excites the tensor tympani motoneurons to activate the tensor tympani muscle, as discussed in Chapter 3. The peak envelope, the total envelope and the band envelopes are converted to decibels by the power-to-dB converters to produce the final loudness estimates in decibels.

Figure 5-8 shows how a sophisticated loudness model can be incorporated into the loudness estimation, this is similar to that proposed by Glasberg and Moore.[48] Like the simple loudness estimator, most models also accept optional weightings to account for the digital-to-acoustic transfer function. The loudness model produces 'instantaneous' estimates of the total and specific loudness of the input signal in sones.

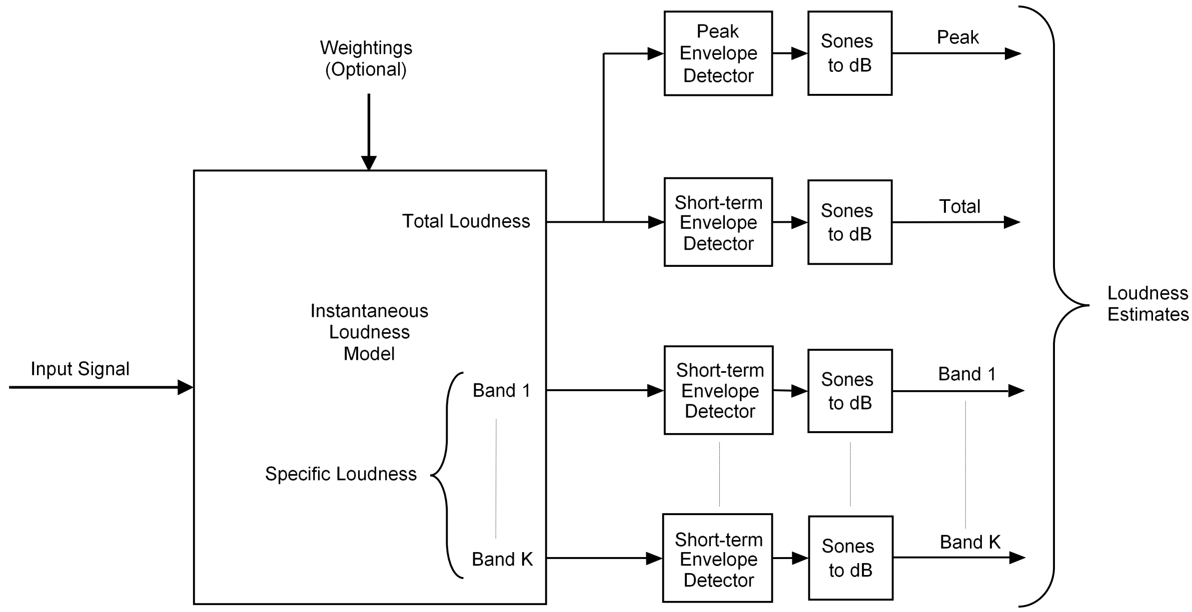


Figure 5-8. *Loudness estimator incorporating a loudness model.*

The instantaneous estimates produced by the model are applied to envelope detectors with similar time constants to those previously described for the simple loudness estimator. The smoothed loudness estimates are converted from sones to decibels. One of the main advantages of using a sophisticated loudness model is that its total loudness is based on summation of the specific-loudness estimates rather than power summation as is the case with the simple loudness estimator described above. Also, unlike the simple power-based model, the spread of excitation between bands and the complexity of loudness growth are modelled. However, because the SRL scheme uses the relative loudness and seeks to minimise this relative difference, the modelling of the complexity of the loudness growth function is less important. For the SRL system to perform the speech reference limiting method, the speech loudness estimator must be largely the same as the current loudness estimator, including the choice of time constants. This ensures that the loudness estimated by the speech estimator is the same as the loudness estimated by the current estimator, which minimizes the changes made by SRL to speech. A simple form of the speech loudness estimator, excluding those elements that are common with the simple form of the current loudness estimator, is shown in Figure 5-9.

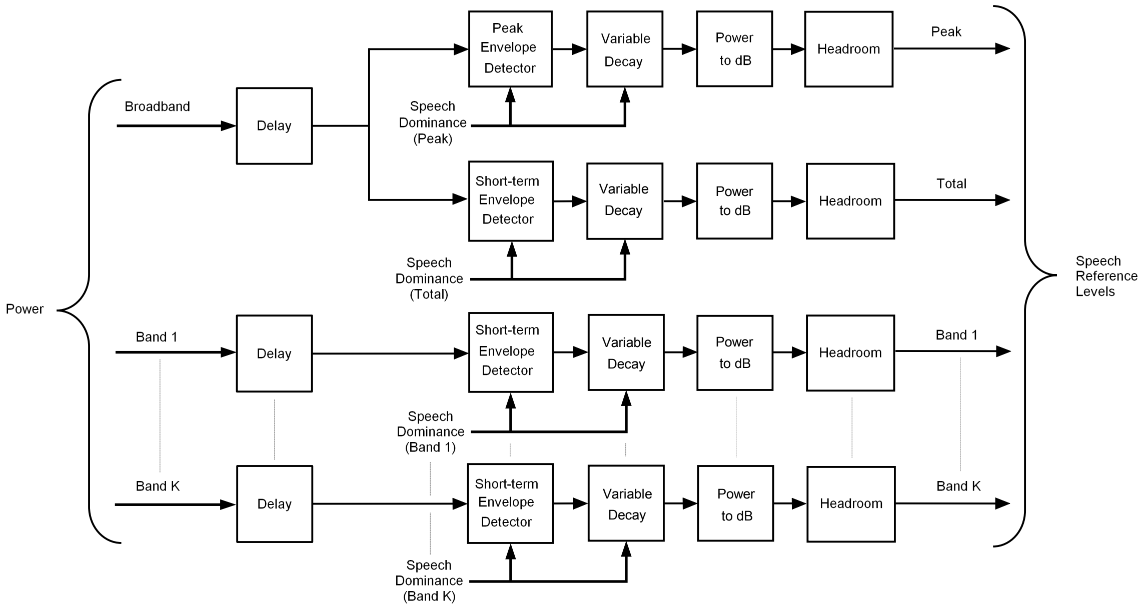


Figure 5-9. *Simplified speech loudness estimator.*

Unlike the current loudness estimation, the speech loudness estimation is updated only when the signal is dominated by speech. This is achieved by adapting the short-term envelope detectors to their input signals only when the speech dominance control provided to them indicates that updating their estimate is required, as shown in Figure 5-9.

The speech loudness estimator should also maintain a memory of the speech loudness when the speech is absent, in a manner similar to a human listener. The memory of loudness is not easy to reliably approximate because it is quite variable and changes with time.[139] In this simple implementation of the speech loudness estimator, the memory of the speech loudness is modelled as a variable decay applied to the speech loudness estimate, as shown in Figure 5-9. The decay rate of the memory of the speech loudness is faster when speech is dominant. If the new speech is louder than the memory of past speech loudness, it immediately adopts the new loudness. If, however, the new speech is softer, it more quickly forgets its memory than it would if there was no dominant speech present and adapts to the new, softer level. A person's expectation of the minimum loudness of speech is modelled by a minimum level below which the memory of speech loudness does not fall. These minimums are based on the data of Pearson et al.[140] combined with that of Byrne et al.[63] When there is no absolute digital-to-acoustic transfer function available to set this minimum, it may be determined statistically from a previously determined history of speech levels.

In the ideal system, the loudness estimators for the current signal and speech are identical and in sync. This means that, when only speech is present in the signal, the outputs of both estimators are identical and therefore no gain reduction is applied to the signal. However, in a real-time system, the speech dominance detector takes time to assess the signal before it makes a decision as to whether the signal is dominated by speech. In order for the loudness estimation of the speech to be in sync with the instructions from the speech dominance detection, the input to the speech loudness estimator needs to be delayed by an equal amount. This means that the signal's current loudness estimate will be a little ahead of the speech loudness estimate by the amount of the delay. The estimated current loudness will therefore briefly exceed the estimated speech loudness when there is a rise in the estimated current loudness level and this will result in gain reduction being briefly applied to the speech. This can be, to some extent, addressed by adding a small margin (i.e. headroom) to the estimated speech loudness so that speech slightly louder than the estimated speech levels can be free of gain reduction. Any headroom added to the estimated speech loudness, however, also enables noise to be louder than the speech by the amount of specified headroom. The value that the headroom is set to is therefore a compromise between providing strict control of the noise loudness and enabling speech free of any limiting when it has increased in loudness prior to its loudness estimate increasing. Headroom values of a few dB have been shown in informal listening trials of the SRL system to be an appropriate compromise. In summary, the resulting speech reference levels are headroom-adjusted estimates of the memory of dominant speech loudness features.

5.2.2 Speech dominance detection

The aim of the speech dominance detector is to determine when speech is the dominant signal so that the speech loudness estimator knows when to update its estimate of the speech loudness. It is preferable to have the speech dominance detector also determine the frequency region over which the speech is dominant and instruct the speech loudness estimator to update its estimate only over this frequency region. This speech dominance frequency range information was not created or used in the simple SRL MKI scheme as this scheme was originally only intended to be used with (tele)communications signals. It was, however, extensively used in the SRL MKII scheme.

It is desirable that the speech loudness estimates not be significantly corrupted by noise that is present along with the speech. An error of up to +2 dB in the short-term level will not lead to a significant error in the speech loudness estimation. The error corresponds to a loudness ratio of about 1.14. A +2 dB error in the short-term level corresponds to added noise of a level up to approximately -3 dB relative to the speech level. This short-term speech-to-noise ratio (SpNR) of 3 dB is therefore a good threshold above which it is reasonable to declare that speech is dominant within a band.

Because the system needs the speech detected only when it is dominant, the task of dominance detection is more easily achieved than in speech detection generally. This is because speech needs to be detected only when the short-term SpNR in narrow frequency regions is 3 dB or more. Furthermore, it is not necessary for the detector to always detect dominant speech. It merely needs to detect dominant speech often enough that the estimates of speech loudness, and hence the speech reference levels, are appropriate for the current speech. Therefore, a large number of false negatives in the speech dominance detection is acceptable. The speech dominance detector is tuned to minimise false positives in its detection as these may lead to the inclusion of high-level noise in the speech loudness estimates and hence the speech reference levels.

In developing this novel process of speech dominance detection, a large number of established approaches to speech detection were considered, and various algorithms were created and tested. These included using traditional speech coding approaches such as linear predictive coding (LPC) and line spectra pair (LSP) coding and speech identification methods such as hidden Markov models (HMM) and neural networks (NN) applied to spectral and cepstral representations of speech. Of the various speech detection methods developed and assessed during the course of this research, the most reliable method of speech dominance detection in the low frequencies was a rule-based scheme that analysed the time-varying harmonic content of the voiced spectra using both spectral and cepstral analysis. The most reliable method of speech dominance detection in the high frequencies was a rule-based sibilance detector that analysed the time-varying spectrum. As the speech dominance detection methods that my colleague at the National Acoustic Laboratories, Dr Nicky Chong-White, and

I developed are currently not covered by intellectual property protection, the detail of the methods cannot be disclosed within this thesis except at a general level as described further in Chapters 6 and 8. Even if the detail could be disclosed it is, in any case, beyond the scope of this thesis.

5.2.3 Comparator and gain calculator

The purpose of the comparator is to determine if there are excess loudness features present in the current signal compared with the speech reference levels that need to be reduced and the amount of reduction required. The comparator also restricts the range of the speech reference levels so that they are within the bounds of predefined maximums and minimums before performing a comparison. This protects against the use of extreme speech reference levels that may have been inadvertently generated as a result of incorrectly classifying high-level noise as speech or detecting very low-level speech, e.g. speech at a large distance from the listener. A speech reference level that was too high would potentially expose a listener to high-level noise and one that was too low would result in excessive limiting and hence signal distortion. The maximums and minimums may be referenced to:

- absolute sound levels where the digital-to-acoustic transfer function is known,
- digital-to-voltage transfer functions where the sensitivity of the typical transducers is known, or
- a range of expected digital levels based on the application.

A schematic diagram of the comparator and gain calculation illustrating the application of these bounds is shown in Figure 5-10.

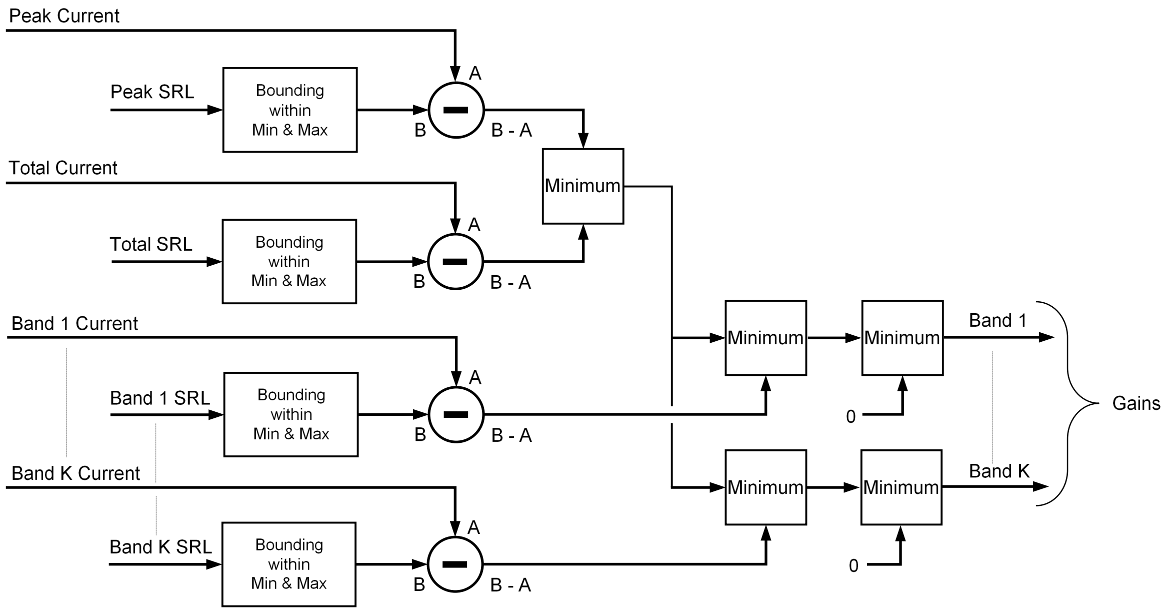


Figure 5-10. *Comparator and gain calculator.*

Deciding on the best method for combining the 'peak', total and frequency-specific loudness controls is not black and white. The simplest approach is for the gain reduction in each band to be the minimum gain prescribed by the 'peak' loudness calculation, the total loudness calculation and the frequency-specific loudness calculation for that band. This arrangement and the final generation of the gain is shown in Figure 5-10. This is, however, an oversimplified approach. Limitations, for example, are that high-level, narrow-band sounds result in a high total loudness which lead to all bands being limited. This broad limiting may help preserve the overall spectral balance but it is not the best approach for maintaining good speech intelligibility when this kind of narrow-band noise is present or, for that matter, any high-level noise is present that has a spectral shape different to that of the speech. Different approaches to this were therefore used in the implemented schemes, SRL MKI and SRL MKII.

5.2.4 Adaptive modifier

The adaptive modifier modifies the input signal to produce the output signal in accordance with the gains provided by the comparator. It is essentially an adaptive filter that may be controlled to provide broad band attenuation as well as frequency-selective attenuation of the signal. There are many methods that may be employed to adaptively filter a signal including adaptive IIR filters, adaptive FIR filters, IIR or

FIR filter bank analysis followed by adaptive modification of the amplitude within the bands and reconstructive synthesis, discrete Fourier analysis followed by adaptive modification of the complex spectrum and inverse discrete Fourier analysis with reconstruction using techniques such as overlap add or overlap save.[117,134,141-143]

The SRL method is not restricted to using any particular method of adaptive filtering. It has been designed so that it can work with any established method of adaptive filtering such as the Weighted Overlap Add (WOLA) scheme in the WOLA audio processor from On Semiconductor.[124,144] Many of the methods mentioned above to some degree either suffer from frequency-domain aliasing, time-domain aliasing, long delay, or phase distortion and are susceptible to quantisation noise and distortion. Of all the filtering topologies, the adaptive linear-phase FIR filter suffers from these issues the least. Unlike the standard filter bank / Fourier transform based approaches, it does not introduce frequency aliasing or time domain aliasing into its reconstructed signal.[145] Nor does it introduce the phase distortion of IIR based approaches or non-linear-phase FIR filter approaches such as Gammatone filter banks.[135] Furthermore it is the filter type most resistant of all to noise and distortion produced by quantisation. It can produce very clean results using low-precision devices. A disadvantage is the trade-off between the delay it introduces (in the linear phase case) and its frequency resolution. For both the SRL MKI and MKII schemes the adaptive linear-phase FIR filter approach was adopted. This approach is discussed in detail in the next chapter on the SRL MKI scheme. However, in other evaluations, not reported in this thesis, the SRL MKII has been successfully included and evaluated in a number of different processing schemes that use filter-bank based adaptive filtering.

5.3 Summary

The SRL scheme is a novel alternative method to the conventional forms of sound-level control. The method measures dynamically varying levels of speech, from which it creates a set of time-varying speech referenced levels, which it uses to control the loudness of sound with respect to. By using the maximum levels of speech as reference levels for limiting, the scheme, in principle, provides *the greatest limiting of noise for the least limiting of speech* making it arguably the optimal method for limiting high-level noise in speech systems. This chapter has described the scheme in general terms, including the processes of:

- Frequency analysis
- Speech dominance detection
- Gain calculation comprising:
    - Current loudness estimation
    - Speech loudness estimation
    - Comparisons of the estimates and gain calculation
- Adaptive modification of the signal

The unique core functions of speech dominance detection and gain calculation may be used with a variety of analysis / modification schemes. In this thesis, two schemes, SRL MKI and SRL MKII are described and evaluated. Both of these schemes are based on the general concepts presented in this chapter.

Chapter 6

SRL MKI scheme

6 SRL MKI scheme

6.1 Introduction

The SRL MKI (mark 1) scheme was initially developed with a focus on telecommunications applications. It was envisaged that it would reside within the digital telephone network and/or the customer's equipment, e.g., private automated branch exchange (PABX), telephone consoles and headset amplifiers. It was therefore initially designed to be used with digital telephone signals with a sampling rate of 8 kHz.

The speech dominance detector was optimised to deal with telephone filtered speech with a bandwidth of less than 4 kHz which generally has a good speech-to-noise ratio. The scheme was later enhanced to deal with wideband telephony and other potential applications such as level-dependent hearing protectors and hearing aids. New processing sampling rates of 16 kHz, 24 kHz, 32 kHz and 48 kHz were introduced. The speech dominance detector was enhanced to deal with speech with a bandwidth of up to 8 kHz using a sampling rate of 16 kHz in addition to its existing mode of operation for speech sampled at 8 kHz. Signals with sampling rates exceeding 16 kHz were converted to 16 kHz for speech dominance detection. However, all other processing was done at the native sampling rate of the signal.

The development work was undertaken in MATLAB[146] and the scheme was coded in the C programming language.[147] Two versions were created:

- A real-time version running on a TMS320C6727 DSP Processor[148]
- A sound file processing application running under Windows

6.2 Overall architecture

An overall schematic of the SRL MKI scheme is shown in Figure 6-1. The scheme is essentially the same as that depicted in Figure 5-6 in Chapter 5 but with the broadband analysis and control performed separately from the multi-band analysis and control.

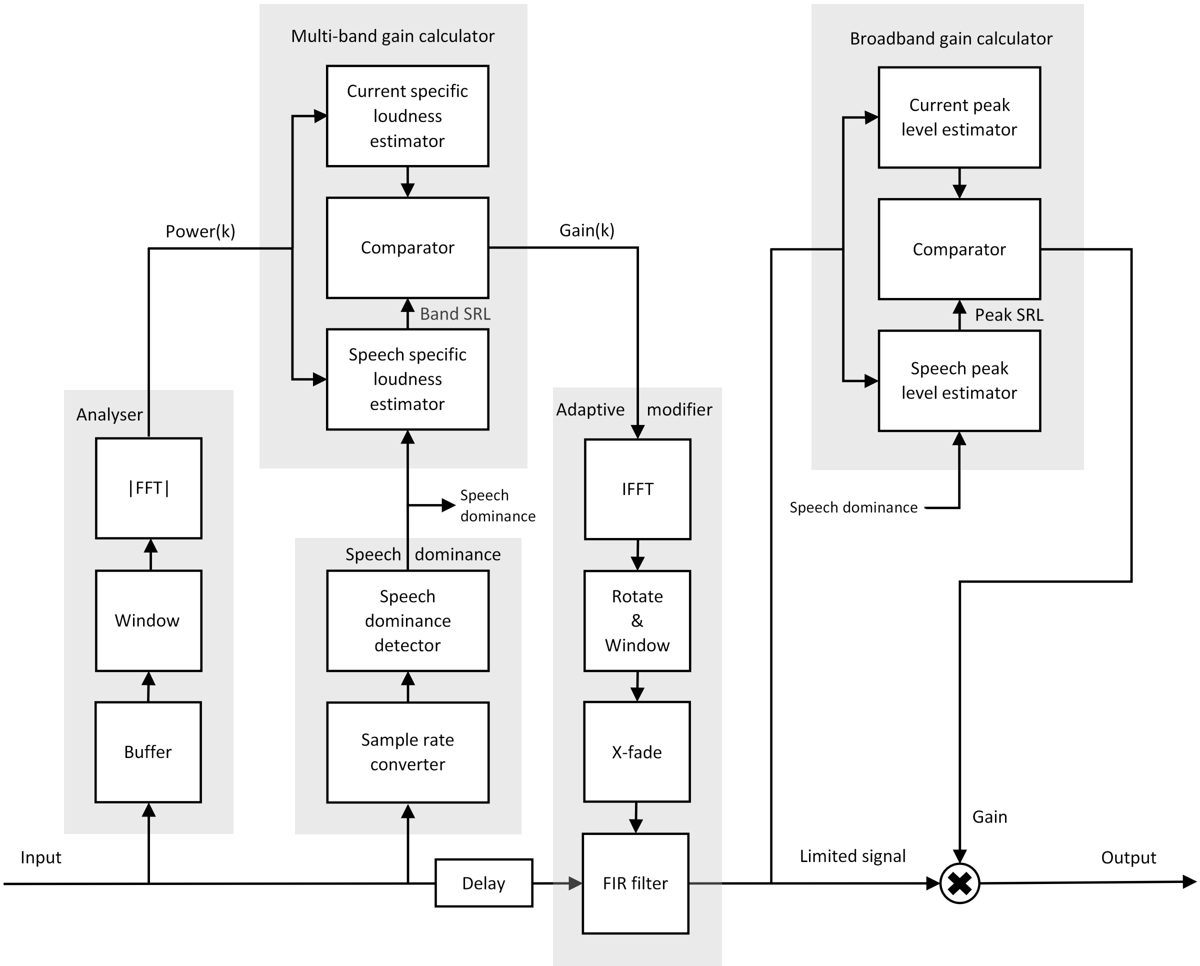


Figure 6-1 *The SRL MKI scheme.*

The four main processes depicted in Figure 5-6 in Chapter 5 are present, these being:

- analyser,
- speech dominance detector,
- multi-band gain calculator (current loudness estimator, speech loudness estimator, comparator) and
- adaptive modifier.

In addition to these there is a delay prior to the adaptive modifier within the processed signal path and a multiplier following it, that provide additional compensatory delay and broadband attenuation of the signal respectively. There is also a broadband gain calculator which takes its input from the limited signal at the adaptive modifier's output and controls the multiplier to provide the broadband attenuation.

6.3 The processed signal path

In keeping with the desire to have very clean processing of the signal, even when using low-precision signal processors, the input signal only passes through a delay, a finite impulse response (FIR) filter and a multiplier on its way to the output as shown in Figure 6-1. This typically will result in the signal only being quantised twice, once for the sum of products in the FIR filter and once for the multiplier. This structure therefore provides a very clean processed signal path. The accepted input signal sampling rates are 8, 16, 24, 32 and 48 kHz. The FIR filter has: a linear phase response and an 8 ms impulse response, which introduces a constant delay of 4 ms into the signal.

In the absence of any limiting, the input signal is multiplied by unity within both the FIR filter and the multiplier resulting in an output signal that is an exact replica of the input signal regardless of the processing precision. The replica is, however, delayed by two delays: one introduced by the linear-phase FIR filter of 4 ms and a second delay introduced by the delay block with an adjustable delay period of zero to 16 ms. The amount of delay is discussed later in this chapter.

In practice, the implemented versions of the scheme employ floating point arithmetic and storage. As a result, the cumulative quantisation noise/distortion introduced by the processing is well below the noise of the input signal.

6.4 The analyser

As discussed in Chapter 5, there are advantages in having narrow-frequency bands in some applications of SRL. The resolution of the frequency analysis should ideally be as fine as the human auditory system. In telecommunications applications, however, there are benefits to it being even finer in the high frequencies, so that the speech level per band in these high frequencies is low, resulting in low speech reference levels. These low speech reference levels in the high-frequency region assist in the control of high-frequency narrow-band sounds that may cause an acoustic shock in individuals who are hyper-sensitive to high-frequency, narrow-band sounds, often as a result of a previous acoustic shock. If the bandwidth of the psychoacoustic-based Bark scale in the low frequencies is considered then a frequency resolution of around 100 Hz is necessary.[45] Such a resolution used with a linear frequency scale gives very low speech reference levels in the high-frequency bands resulting in significant suppression of narrow-band sounds.

A linear frequency scale was therefore identified as preferable for this application making the fast Fourier transform (FFT) the preferred method of multi-band frequency analysis over other filter bank methods. In addition to providing a linear frequency scale, the FFT has the advantage of being a computationally efficient method of frequency analysis. Using an FFT with a time window of around 8 ms would approximate the desired frequency resolution. From an auditory temporal processing perspective, this is similar, assuming the use of a rectangular window, to the 8 ms equivalent rectangular duration (ERD) of the auditory system 'window' found by Moore et al.[149] The FFT lengths employed and the resulting band (bin) frequency resolution for the various sampling rates supported by the SRL MKI scheme are shown in Table 6-1.

Sampling rate	FFT length	Frequency resolution
(kHz)	**(points)**	**(Hz)**
8	64	125
16	128	125
24	256	93.75
32	256	125
48	512	93.75

Table 6-1. *FFT length and frequency resolution for the supported sampling rates.*

Although the FFT lengths given in Table 6-1 for sampling rates of 24 kHz and 48 kHz indicate a temporal duration of 10.666... ms the actual window is 8 ms which has been padded out with zero value samples to reach a temporal duration of 10.666... ms.

Ideally, a system should analyse and control the signal without a delay being necessary, i.e. on an instantaneous basis. However, due to computational limitations as well as the delay introduced by real-time analysis filtering some delay is inevitable. The analysis needs to track the dynamic changes of the signal in order to control them and hence control the perceptual and somatic response to them. Provided a delay is applied to the controlled signal that is at least as long as the analysis update period, and the analysis window is as fine as the ERD of the auditory system, then an analysis period up to an ERD is acceptable for perceptual control. For the SRL MKI scheme, a multi-band analysis and processing update rate of 250 Hz was selected. This corresponds to an update period of 4 ms, which is half the period of the unpadded FFT window length and half the ERD of the auditory system found by Moore et al. The buffer therefore extracts 8 ms long blocks of the input signal (i.e. the analysis window length) every 4 ms (the analysis update period) with each block overlapping the previous block by 4 ms - a 50% overlap. Table 6-2 shows the block length and block overlap for each of the scheme's supported sampling rates.

Sampling rate	Block length	Block overlap
(kHz)	(samples)	(samples)
8	64	32
16	128	64
24	192	96
32	256	128
48	384	192

Table 6-2. *Block and overlap lengths for the supported sampling rates.*

The 8 ms blocks were weighted with a windowing function. There are numerous analysis window functions that could have been selected for the purpose. Many of these have been extensively analysed by Harris.[150] The window shape ideally should have a temporal integration that is no slower than the auditory system's and a frequency selectivity as fine as the auditory system's. The chosen window, known as a Hann function, applies to the input blocks a half-period cosine response that has been raised to the power of 2.

The Hann window was selected because of its smooth time response which has -6 dB duration of 4 ms, which is finer than the ERD resolution and has good frequency selectivity although it is not as fine as the auditory system in the low frequencies. It is suited to being used with a 50% window overlap as, at this low degree of overlap, the overlapping windows add to unity. Therefore, there is no signal power loss as noted by Harris. The time and frequency response of the Hann window is shown in Figure 6-2 and Figure 6-3 respectively.

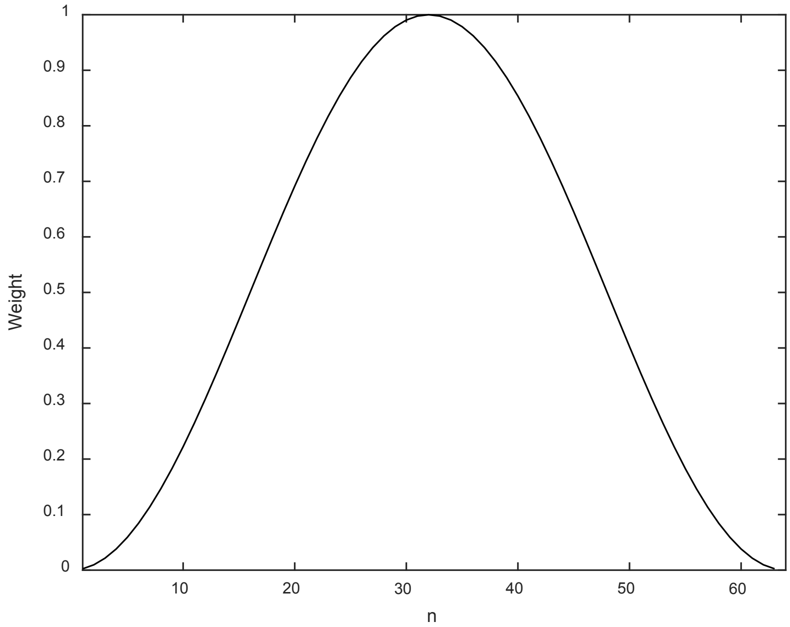


Figure 6-2. *Time response of a Hann window.*

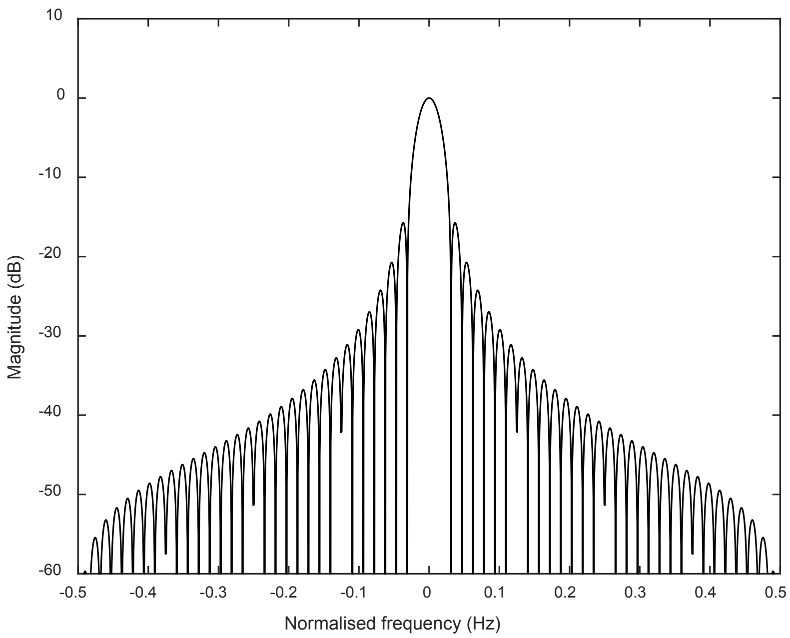


Figure 6-3. *Frequency response of a Hann window.*

6.5 Speech dominance detection

The speech dominance detector has two different modes of operation, only one is selected for a signal of a given sampling rate, the modes are:

- narrow-band mode, using a sampling rate of 8 kHz
- wideband mode, using a sampling rate of 16 kHz.

For signals with a sampling rate of 8 kHz the narrow-band mode is activated. For signals with a sampling rate of 16 kHz or higher the wideband mode is activated.

To detect speech dominance in signals with sampling rates in excess of 16 kHz, sample rate conversion (SRC) is applied to the signal to produce an input signal for the speech dominance detector. A bandwidth of around 7.7 kHz captures all the significant features of speech even though the input signal may contain acoustic signal energy up to 24 kHz. According to the speech intelligibility index (SII),[12] the contribution to speech intelligibility of speech components with a frequency above 7.7 kHz is around 1.5%. The amount of energy in the region above 7.7 kHz diminishes with increasing frequency and, similarly to its minimal contribution to speech intelligibility, it contributes almost nothing to detection of speech dominance. Because the benefit gained from attempting to capture information from these higher frequencies is very small and the processing load to do so is so large, a speech dominance detector with a bandwidth of around 7.7 kHz and a 16 kHz sampling rate is preferable. Table 6-3 shows the sampling conversions for the various sampling rates supported.

In each case the relationship between the input sampling rate and the output sampling rate is a rational factor and therefore may be accurately performed. For an input sampling rate of 32 kHz, a 128 tap linear-phase FIR filter was used to remove the frequency components above approximately 7.7 kHz before down-sampling it by a factor of 2. For an input sampling rate of 48 kHz, a 192 tap linear-phase FIR filter was used to remove the frequency components above approximately 7.7 kHz before down-sampling it by a factor of 3. For an input sampling rate of 24 kHz, the input signal was up-sampled by a factor of 2 and a 192 tap linear-phase FIR filter was used to remove the frequency components above approximately 7.7 kHz before down-sampling it by a factor of 3. In all three cases of sample rate conversion, a signal delay of 2 ms was introduced.

Input sampling rate	Output sampling rate	SRC Ratio
(kHz)	**(kHz)**	
8	8	1
16	16	1
24	16	2/3
32	16	1/2
48	16	1/3

Table 6-3. *Speech detector sample rate conversion (SRC).*

As outlined in Chapter 5, most of the detailed processing performed by the speech dominance detector cannot be revealed due to the current lack of intellectual property protection applied to its novel methods. However, some performance aspects, related to its interface, can be revealed. First, it takes typically 40 ms to detect speech dominance plus a further 2 ms when sample rate conversion is used. Second, the speech dominance detector provides a single binary signal that indicates when speech is dominant, which it updates every 20 ms.

6.6 The adaptive modifier

The adaptive modifier applies a linear-phase FIR filter with an 8 ms impulse response and a 4 ms delay to the signal as it passes from the input to the output. Its length in taps for the supported sample rates is shown in Table 6-4. The adaptive modifier also includes a system of generating FIR filter coefficients from the set of band specific-gains generated by the multi-band gain calculator. The filter coefficients update at the signal sampling rate, while the set of band-specific gains are produced at the analysis sampling rate of 250 Hz or every 4 ms.

Sampling rate	Filter length
(kHz)	(taps)
8	64
16	128
24	192
32	256
48	384

Table 6-4. *Adaptive modifier filter lengths for the supported sampling rates.*

The method of coefficient generation is a combination of the:

- frequency sampling method of designing linear-phase FIRs from a sampled frequency specification using the inverse Fourier transform,
- windowing method of impulse response control,
- cross-fade method of interpolating intermediate sets of values between two sets of values.

The first two steps are techniques used in design of an arbitrary-magnitude, linear-phase FIR filter found in many books on digital signal processing (DSP) fundamentals such as 'Digital signal processing: principles, algorithms, and applications' by Proakis and Manolakis.[151] Performing this method on a regular basis to provide time-varying filtering with an arbitrary frequency response was discussed by Kates who referred to it, in the context of dynamic-range compression, as the 'side-branch compressor structure'.[143] Similar methods of using a linear-phase FIR filter with a time-varying frequency response being controlled by an analyser / gain calculator in parallel with it have been documented by Williamson et al.,[117] Fisher,[145] and Schaub.[152,153] The third step is a cross-fade method of interpolating between sets of filter coefficients described by Fisher. This cross-fade is effectively a low-pass filtering operation applied to interpolate intermediate sets of coefficients from sets of time-varying coefficients defined at a slower sampling rate. Allen and Rabiner when describing the overlap-add approach to short-time Fourier analysis/synthesis noted the equivalence of the overlap-add technique to an FIR filter with time-varying coefficients that have been low-pass filtered by the windowing employed in the short-time Fourier Analysis/Synthesis.[141] Although the technique is akin to the overlap-add approach it does not suffer from either time or frequency aliasing nor from the compounding of quantization errors.

The adaptive modifier is shown in Figure 6-1. Details of the filtering technique are now described. A set of gains is supplied by the multi-band gain calculator. These are defined at equispaced frequencies from 0 Hz to half the sample frequency. These gains are mirrored around the half sampling frequency and applied to the magnitude input of an inverse fast Fourier transform (IFFT) with the phase input set to zeros. The IFFT produces an impulse response at its real output. Due to the input of the IFFT having a zero phase response, the real output of the IFFT also has a zero phase response and hence it is rotated by half its length in order to create a linear phase shift, a delay equal to half its length. While the impulse response gives the correct magnitude response at each of the frequencies at which it is defined in the frequency domain, it may contain deep dips in its response between these points due to the truncation of its impulse response by the rectangular windowing inherent in the IFFT. Application of a Hann window function to the impulse response will smooth this discontinuity and remove the dips in the frequency response. This means any abrupt changes in the specified frequency response will also be smoothed, which in most cases is beneficial.

The sets of coefficients are produced at the analysis sample rate, i.e. 250 Hz, while the FIR filter is operating at the signal's sample rate which can be as high as 48 kHz, i.e. up to 192 times the analysis rate. It would be undesirable to change the coefficients abruptly at the analysis sample rate as this would be equivalent to a 'square' wave with a frequency of 125 Hz modulating the signal. A smoother transition is needed, and there are several interpolation functions to choose from: linear (i.e. triangular) and cosine squared (i.e. Hann), and so forth. What is essential is that the interpolation is smooth and the overlapping interpolation weights add to unity. In SRL MKI a linear interpolation function was chosen to cross fade between the current and previous sets of coefficients. The interpolation or cross-fade length, M is given by the following equation.

$$M = \frac{Fs_{signal}}{Fs_{analysis}} \quad \textbf{(6-1)}$$

where:

M is the interpolation length in samples

Fs_{signal} is the sampling rate of the signal

$Fs_{analysis}$ is the sampling rate of the frequency analysis

The cross-fade operation combined with the convolution operation of the FIR is given by the following equation.

$$y(n) = \sum_{i=0}^{I-1} \left((1 - \alpha(m)) \cdot b_{prev}(i) + \alpha(m) \cdot b_{curr}(i) \right) \cdot x(n-i) \quad \textbf{(6-2)}$$

where:

α(m) is the interpolation weight, which is

α(m) = m / M for m = 0, 1, 2, ... M-1

with m incrementing at the signal sampling rate, Fs_{signal} and repeating at the analysis sampling rate, $Fs_{analysis}$ (i.e. every M samples)

b_{prev} is previous set of coefficients

b_{curr} is current set of coefficients

i is the convolution index

I is the number of coefficients

m is the interpolation index

M is the interpolation length

n is the sample number

x(n) is the input sample

y(n) is the output sample

Table 6-5 lists the interpolation factor or cross-fade length, M for the range of sampling rates supported.

Input sampling rate	Interpolation factor, M
(kHz)	points
8	32
16	64
24	96
32	128
48	192

Table 6-5. *FIR coefficient interpolation lengths for the supported sample rates*

In relation to a non-adaptive FIR filter which involves N multiplication and N-1 addition operations per sample, the direct implementation of the adaptive FIR filter in Equation 6-2 involves 3N multiplications and 2N-1 additions, although, it can be performed more efficiently using 2N+2 multiplications and 2N-1 additions. This filter, however, results in the most artefact-free filtering with neither time nor frequency aliasing and the least amount of computational noise. When there is no limiting, the only impact of the filter is simply a delay.

6.7 Multi-band gain calculator

The multi-band gain calculator is the core of the SRL processing. It is here that the current signal loudness and the speech signal loudness are estimated and compared and a gain is produced that will control the signal so that noise exceeding the speech reference levels is reduced. One band of the multi-band gain calculator is shown in Figure 6-4. The bands are all identical to the one shown. The variables are shown as functions of the band number, k. All processing is done without conversion to a logarithmic representation.

The current signal's specific loudness is estimated by the current loudness estimator; this is simply a short-term envelope detector applied to the signal's band power. The time constants for these have been varied during the development but for the formal evaluation of SRL MKI the attack and the release were set to 80 ms. This is slightly faster than the typical time constant for loudness integration of 100 ms discussed in Chapter 2. The attack and release time constants are both slower than those used in the short-term loudness model of Glasberg and Moore[48] but faster than those used in the long-term loudness model of Glasberg and Moore[48] and the dynamic loudness model of Chalupper and Fastl.[49] However, as discussed later in this section, a slower decay is applied to this estimate for speech.

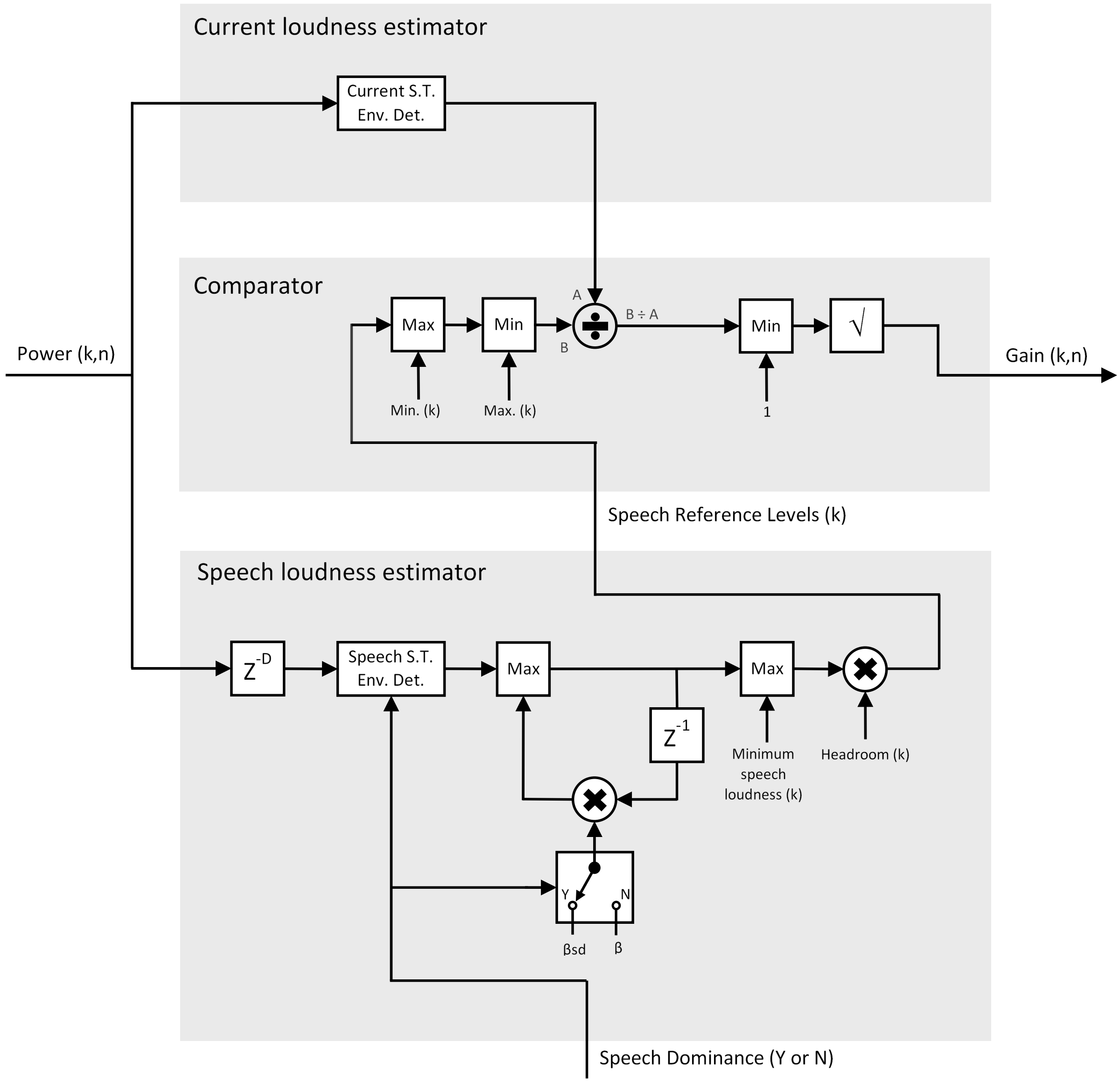


Figure 6-4. *One band of the multi-band gain calculator.*

A detailed schematic diagram of the envelope detector is shown in Figure 6-5. It is a 1st order IIR low-pass filter with switchable coefficients for representing different time constants. These time constants are switched depending on whether the enveloped detector is in 'attack' mode or 'release' mode. Attack mode occurs when the input sample p(n) is greater than the previous output sample e(n-1).

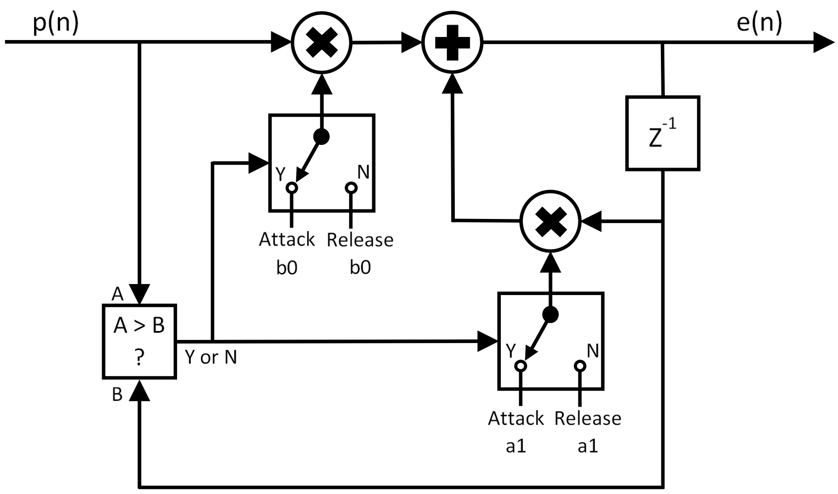


Figure 6-5. *Envelope detector.*

When this occurs, the comparator A > B ? produces a Yes or 'Y' output. Otherwise, it is in release mode and it produces a No or 'N' output. When in attack mode, the attack coefficients are selected for the b0 and a1 coefficients otherwise the release coefficients are selected.

The operation is given by the following equation.

$$e(n) = b0 \times p(n) + a1 \times e(n-1) \tag{6-3}$$

where:

$e(n)$ *is the envelope*

$p(n)$ *is the input signal* $(i.e. power)$

n *is the sample number*

$$b0 = \begin{cases} attack\ b0, & p(n) > e(n-1) \\ release\ b0, & otherwise \end{cases}$$

$$a1 = \begin{cases} attack\ a1, & p(n) > e(n-1) \\ release\ a1, & otherwise \end{cases}$$

The relationship between the time constants and their coefficients is given by the following equation.

$$a1 = e^{-\left(\frac{1}{Fs_{analysis} \times \tau}\right)} \qquad b0 = 1 - a1 \tag{6-4}$$

where:

τ *is the engineering time constant, i.e.* $80\ ms$

$Fs_{analysis}$ *is the analysis sampling rate, i.e.* $250\ Hz$

The speech specific loudness is estimated by the speech loudness estimator. Its short-term envelope detector is identical to the current signal's detector with the exception that it only updates its estimate when speech is dominant, i.e. when the speech dominance control signal applied to it is true. Its output is set to zero when the speech dominance control signal is false so that its value does not further influence the memory of speech loudness that follows it. As discussed in the earlier section on the speech dominance detector, the detection process updates every 20 ms and its decision is typically delayed by 40 ms plus an extra 2 ms if there is sample rate conversion involved. The power signals from which the loudness is estimated need to be aligned with this delayed speech dominance decision. This is achieved by delaying the power signals by the same amount. The value of the delay D, in samples, is 9. This is calculated using the following equation.

$$D = truncate\{(Delay_{speech.\ dom.\ det.} - Delay_{analyser}) \times Fs_{analyser}\} \quad \textbf{(6-5)}$$

where:

$$Delay_{speech.\ dom.\ det.} = \begin{cases} 0.04\ \mathrm{s\ for\ Fs} = 8 \mathrm{\ and\ } 16\ \mathrm{kHz} \\ 0.042\ \mathrm{s\ for\ Fs} = 24, 32\ \&\ 48\ \mathrm{kHz} \end{cases}$$

$$Delay_{analyser} = 0.004\ s$$

$$Fs_{analyser} = 250\ Hz$$

As discussed in Chapter 5, the perception of speech loudness is largely determined by the current loudness and fades with time.

The memory of the speech loudness is modelled by the 1st order decay function, comprising: a unit delay, a multiplier with the switchable decay coefficients, β and β_{sd} and a max function as shown in Figure 6-4. The memory of the speech loudness estimate is given by the following equation.

$$speech\ loudness\ memory(k,n) \ldots \quad \textbf{(6-6)}$$

$$= max \begin{cases} speech\ loudness\ estimate(k,n) \\ speech\ loudness\ memory(k,n-1) \times \begin{cases} \beta_{sd}, speech\ dominant \\ \beta, \quad otherwise \end{cases} \end{cases}$$

where:

n is the sample number

k is the band number

β is the decay rate

When the dominant speech is louder than the decayed speech loudness memory the model immediately adopts this louder estimate, otherwise it decays the speech loudness estimate. The rate of decay depends on whether dominant speech is present at a lower loudness. In the absence of dominant speech, the modelled memory decays the speech loudness slowly. In this model, this decay rate was set to 0.2 dB/s resulting in a decay coefficient, β = 0.99998158, see the following equation for its calculation. When dominant speech of a loudness lower than the decayed speech loudness appears, the model quickly adopts this new loudness estimate at a rate of 2 dB/s using a decay coefficient β_{sd} = 0.998159, see the following equation for its calculation.

$$\beta = \left(10^{-decay\ rate\frac{dB}{s}/10}\right)^{\frac{1}{Fs_{analysis}}} \qquad \textbf{\textit{(6-7)}}$$

where:

$Fs_{analyser} = 250\ Hz$

$decay\ rate = \ 0.2\ dB/s\ or\ 2\ dB/s$

The expected minimum loudness of speech is simulated by taking the maximum of the decayed speech loudness and a pre-determined minimum speech level. This is applied using a max function as shown in Figure 6-4 and expressed by the following equation.

$$speech\ loudness\ memory\ (k,n) = Max\begin{cases} speech\ loudness\ memory\ (k,n) \\ minimum\ speech\ level\ (k) \end{cases} \qquad \textbf{\textit{(6-8)}}$$

where:

n is the sample number

k is the band number

As discussed in Chapter 5, it can be beneficial to allow a small margin to accommodate transitional speech that is slightly louder than the established loudness of speech. This is achieved by including a headroom allowance when forming the final speech reference levels from the speech loudness memory estimates. This headroom is applied through multiplication as shown in Figure 6-4 and expressed the following equation.

$$SRL(k,n) = Headroom(k) \times speech\ loudness\ memory(k,n) \qquad \textbf{(6-9)}$$

where:

n is the sample number

k is the band number

In the formal evaluation of SRL MKI, the headroom multiplication factor was set to unity.

The final speech reference levels are passed to the comparator. Before comparing the current loudness estimates with the speech reference levels the speech reference levels are checked against maximum and minimum bounds. As discussed in Chapter 5, this ensures that the speech reference levels (limits) are neither so low that they result in excessive distortion or so high that they would potentially expose a listener to high-level noise. The bounding operation of the comparator is shown in Figure 6-4. The speech reference levels are compared with the set of minimum levels. If any fall below these values, they are clamped to them. The comparator then compares the speech reference levels with a set of maximum levels and if they rise above these values, they are clamped to them. The bounded speech reference levels are then compared (divided) by the current loudness levels. The quotient is the power gain, which is capped at one so as to only provide gain reduction. The amplitude gain is simply the square root of the power gain. This and the other operations of the comparator are shown in Figure 6-4 and expressed by the following equation.

$$Gain(n,k) = \sqrt{min\begin{Bmatrix} \frac{bounded\ SRL(n,k)}{CLE(n,k)} \\ 1 \end{Bmatrix}} \qquad \textbf{(6-10)}$$

$where$:

$CLE(n,k)\ are\ the\ current\ loudness\ estimates$

$$bounded\ SRL(n,k) = min\begin{cases} Max\ levels\ (k) \\ max\begin{cases} SRL\ (n,k) \\ Min\ levels\ (k) \end{cases} \end{cases}$$

and

$Max\ levels(k)\ are\ the\ maximum\ bounds\ for\ the\ SRL\ levels$

$Min\ levels(k)\ are\ the\ minimum\ bounds\ for\ the\ SRL\ levels$

$SRL(n,k)\ are\ the\ speech\ reference\ levels$

n is the sample number

k is the band number

6.8 Delay

It is desirable that the signal being modified by the adaptive modifier and the modification being applied to it are time-aligned. There is also an advantage in further advancing the modification being applied (by delaying the signal being modified) to provide what is termed 'look-ahead'. Look-ahead compensates to some extent for the time it takes for a change in the signal level to be estimated and a change in gain to be generated in response. An early description of the look-ahead being applied in the dynamic control of an audio signal appeared in the patent by Thomas in 1988.[154] When the control of the signal is slow, the signal can 'overshoot' its final controlled value. Look-ahead is most beneficial in reducing 'overshoot' caused when the estimate is slow to respond to a sharp increase in signal level due to a slow attack time.

The analysis filter bank and the adaptive modifier introduce a delay of 4 ms each into the gain applied to the signal. However, the adaptive modifier also introduces a delay of 4 ms into the signal. Although, when there are sharp changes in the frequency response of the adaptive modifier, there is a wide spread of its impulse response, and this means that the earlier samples that enter the filter contribute to its output, and so the benefit of the linear phase delay of 4 ms is to some extent negated. It is therefore not a clear cut decision as to what additional delay should be employed. The delay has been made adjustable from 0 ms to 16 ms. It has been set to a range of values during this research, with the longer delay periods offering an advantage in the control of transients through providing a look-ahead function. The trade-off is that delay in the processing may cause problems, particularly in applications where the user's own voice is presented to them after processing in real-time. In the formal evaluation of SRL MKI described in Chapter 7 the delay was set to 4 ms.

6.9 Broadband gain calculator

The objective behind the broadband gain calculator is to generate a gain to control the broadband peak level of the signal relative to the broadband peak level of the speech. There are input signals whose total peak power exceeds the typical peak power of past speech but whose specific loudness does not exceed that of past speech enough to reduce the total peak power to that of past speech. The multi-band gain calculator does not contain any form of loudness summation or power summation to control the total loudness or power of the signal. It is therefore necessary to specifically control it in the broadband signal. This is achieved by performing largely the same estimation and gain calculation processes that were applied to the band signals to the broadband signal, but using faster time constants. As shown in the system schematic in Figure 6-1, the broadband gain calculator receives the limited signal produced by the adaptive modifier, for which it produces a gain that is used to control the limited signal on a broadband basis using a multiplier. It therefore deals with a pressure signal rather than a power signal. The broadband gain calculator is shown in Figure 6-6.

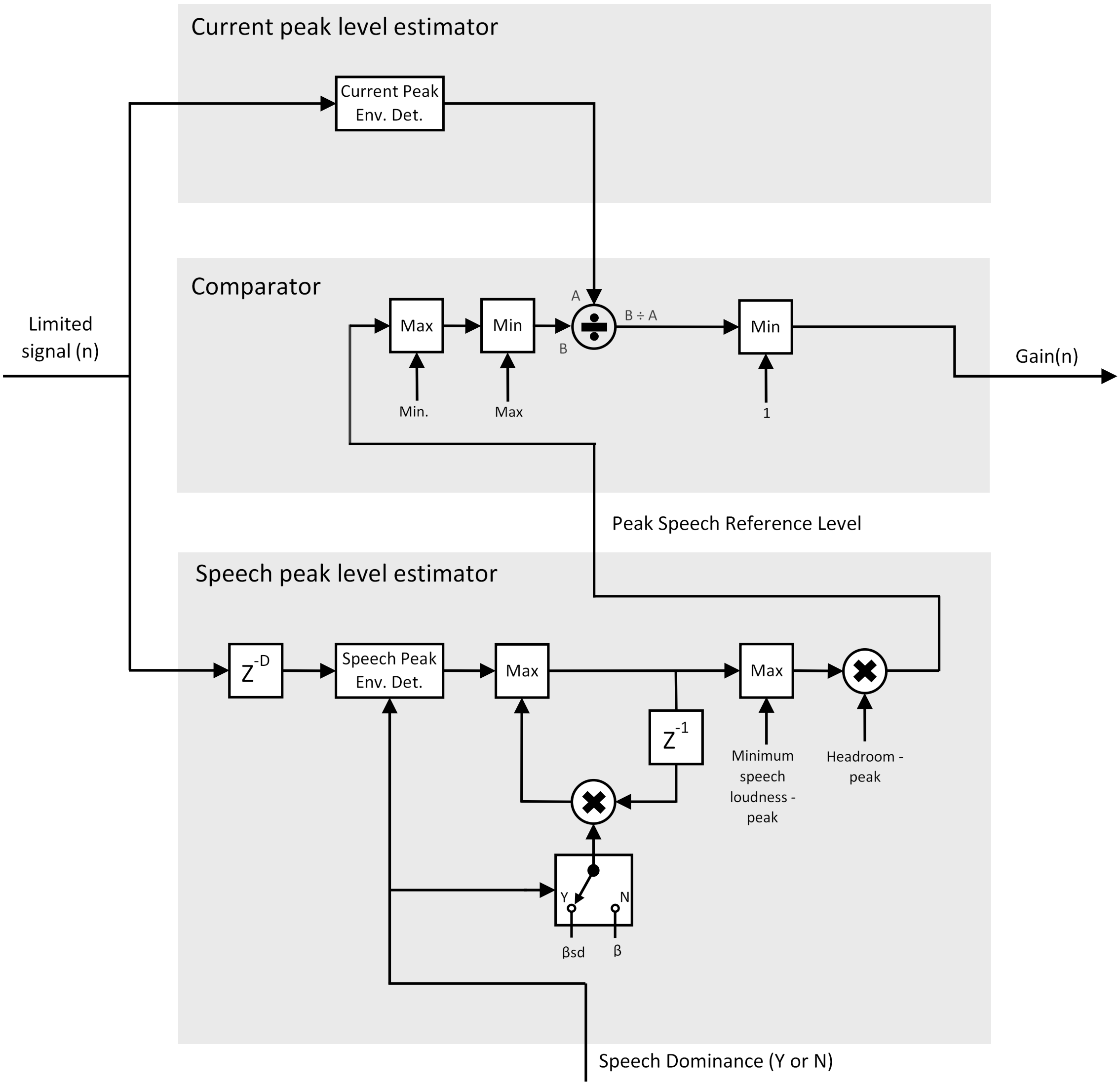


Figure 6-6. *Broadband gain calculator.*

The current and speech peak envelope detectors in Figure 6-6 are the same as those described for the multiband gain calculator with the exception of the following:

- The envelope detectors take the magnitude of their input signal prior to any other processing.
- The attack time constant is 0.25 ms.
- The release time constant is 10 ms.

The speech dominance detector alignment delay within the broadband speech peak level estimator need not be as long as the alignment delay in the multi-band speech loudness estimator. This is because the signal it receives is delayed by 4 ms by the 'compensation' delay preceding the adaptive modifier and by another 4 ms within the adaptive modifier. However, there is no analyser delay of 4 ms involved so the required delay is only 4 ms shorter than the alignment delay used in the multi-band speech loudness estimator.

The value of the delay D, in samples, is calculated as follows.

$$D = truncate\left\{\left(Delay_{\,spe.dom.det.} - Delay_{comp.\ delay} - Delay_{adapt.\ mod.}\right) \times Fs\right\} \quad \textbf{\textit{(6-11)}}$$

$where$:

$$Delay_{\,speech\ dominance\ detector} = \begin{cases} 0.04\,\mathrm{s}\ \ for\ \ Fs = 8\ and\ 16\ kHz \\ 0.042\,\mathrm{s}\ \ for\ \ Fs = 24, 32\ \&\ 48\ \ kHz \end{cases}$$

$$Delay_{compensation} = 0.004\ s$$

$$Delay_{adaptive\ modifier} = 0.004\ s$$

$$Fs = 8, 16, 24, 32\ and\ 48\ kHz$$

The resulting compensation delays for the supported sampling rates are given in Table 6-6.

Input sampling rate	Delay D
(kHz)	(samples)
8	256
16	512
24	816
32	1088
48	1632

Table 6-6. *Compensation delay lengths for the supported sample rates.*

The speech peak level estimator employs the same method of memory simulation and uses the same memory decay rates as used by the multi-band speech loudness estimators. Like the multi-band speech loudness estimators, the speech peak level estimator also applies a minimum restriction on the memory of speech loudness and provides an allowance of headroom. In the formal evaluation of SRL MKI, the minimum speech level was set to a very low level so that it would not influence the results of the assessment and the headroom multiplication factor was set to unity.

The final gain calculation and the other operations of the comparator are shown in Figure 6-6 and expressed by the following equation.

$$Gain(n) = min\begin{cases} \frac{bounded\ peak\ SRL(n)}{CPPE(n)} \\ 1 \end{cases} \qquad \textbf{(6-12)}$$

$where$:

$CPPE(n)\ is\ the\ current\ peak\ pressure\ estimate$

$$bounded\ peak\ SRL(n) = min\begin{cases} Max\ level \\ max\begin{cases} peak\ SRL\ (n) \\ Min\ level \end{cases} \end{cases}$$

and

$Max\ level\ is\ the\ maximum\ bound\ for\ the\ peak\ SRL\ level$

$Min\ level\ is\ the\ minimum\ bound\ for\ the\ peak\ SRL\ level$

$peak\ SRL(n)\ are\ the\ speech\ reference\ levels$

n is the sample number

Because the signal is a pressure signal rather than a power signal the final square root to produce the gain is not required.

6.10 Applications and coding

Two versions of SRL MKI have been created:

- A real-time version running on a TMS320C6727 DSP processor
- A sound file processing application running under Windows

6.10.1 Real-time DSP Version for informal evaluation

A real-time version of SRL MKI running on a TMS320C6727 Digital Signal Processor (DSP)[148] was created in ANSI C code using the TI's Code Composer development software.[155] The code was run in real-time on a Lyrtech Professional Audio Development Kit (PADK).[156] This development device is shown in Figure 6-7. It was used to listen to the SRL MKI processing of audio in real-time and to refine the SRL MKI algorithm based on informal listening tests.

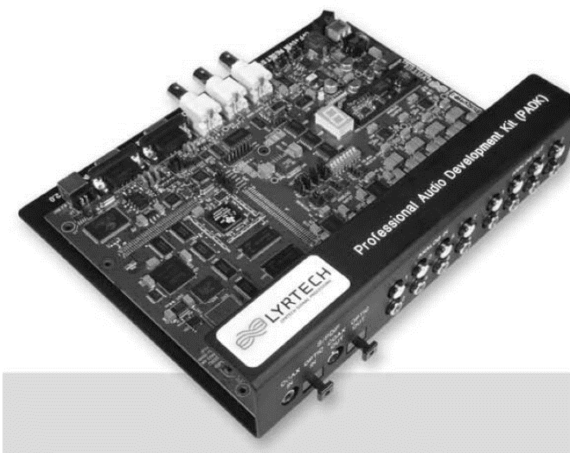


Figure 6-7. *Development device.*

6.10.2 Application Windows Audio File Processing

A sound file processing Windows application was developed to enable evaluation of SRL MKI by others as well as to quickly process sound files. The application was created as both a command line application and with a Windows graphical user interface. The SRL MKI user interface is shown in Figure 6-8. In addition to enabling the user to select a sound file, process it with SRL, and save the SRL-processed sound file, the application allows the user to playback the original and SRL-processed sound files. The application also provides a waveform display of the original and SRL-processed sound files.

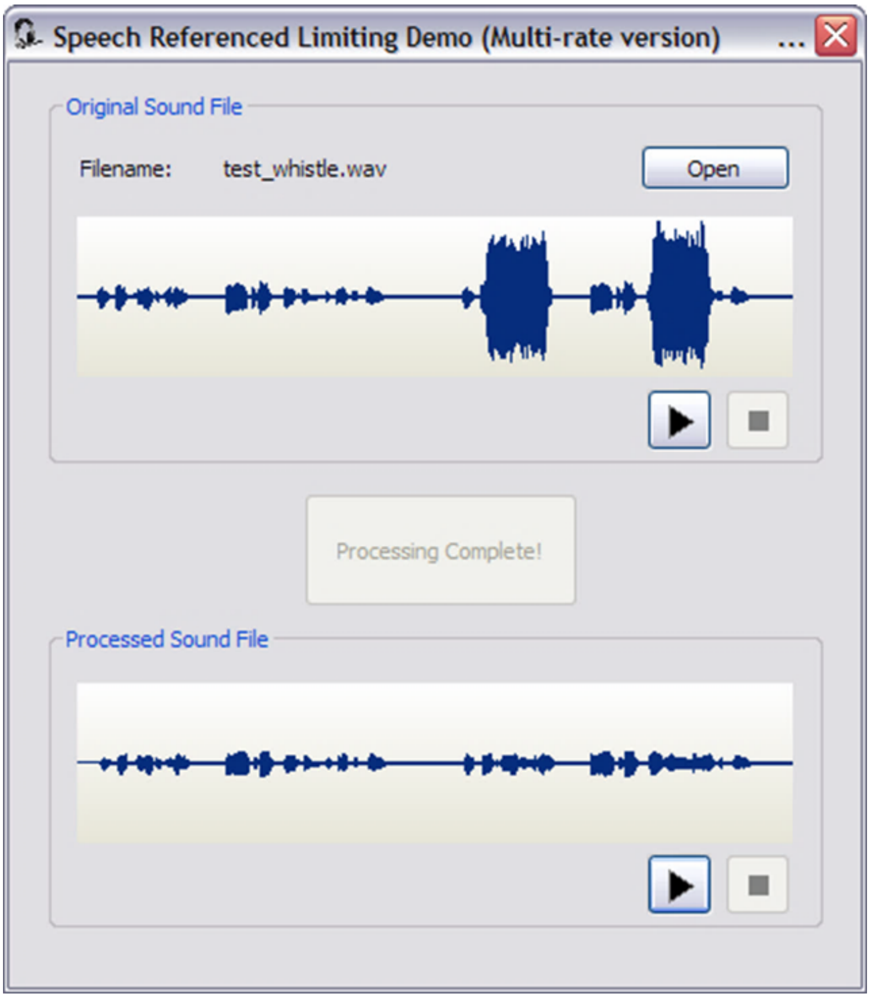


Figure 6-8. *SRL MKI Windows application user interface.*

6.11 Summary

This chapter has presented the SRL MKI scheme. This scheme was originally designed with a focus on telecommunications applications and used a sampling rate of 8 kHz. The speech dominance detector was initially optimised to deal with telephone-filtered speech with a bandwidth of less than 4 kHz which generally has a good speech-to-noise ratio. The scheme was later enhanced to deal with wideband telephony and other potential applications such as level-dependent hearing protectors and hearing aids. Additional sampling rates of 16 kHz, 24 kHz, 32 kHz and 48 kHz were introduced. The scheme performs analysis of the signal in both the time and frequency domains but controls the signal in the time domain using an adaptive linear-phase FIR filter followed by a fast broadband gain control. The algorithm was coded in ANSI C and a real-time DSP version and a Windows audio file processing application were developed.

Chapter 7

Subjective evaluation of SRL MKI

7 Subjective evaluation of SRL MKI

7.1 Introduction

This chapter reports on the subjective evaluation of the SRL MKI scheme.

It was hypothesised that, when listening to a speech signal, listening comfort would be improved through limiting the loudness of non-speech sounds with reference to the loudness of the speech to which the listener was acclimatised. It was also hypothesised that this processing would have minimal effect on the speech quality. These hypotheses rely on there being times when speech is sufficiently dominant within the signal for its loudness to be estimated.

Three laboratory experiments were conducted to assess the efficacy of the SRL MKI scheme in three applications, these being:

- hearing aids,
- level-dependent hearing protectors, and
- telephone headsets.

The aim was to obtain subjective data on the loudness control and the speech quality provided by the method. The hypotheses to be assessed were:

- that the method controls the loudness of non-speech sounds relative to speech sounds for higher level non-speech sounds, and
- that the method does not degrade speech quality.

7.2 Method

The experiments comprised: collecting stimuli that might be encountered in the three applications; processing these stimuli using the SRL MKI scheme described in Chapter 6; presenting the processed and unprocessed stimuli to subjects in the laboratory; collecting the subject’s responses and analysing them. The settings for SRL MKI processing are as shown in Table 7-1.

Parameter	Value
Analysis scale	Linear
Analysis filter bandwidth	125 Hz
Analysis window duration	8 ms
Analysis rate	250 Hz
Specific loudness integration time	80 ms
Fast memory decay rate	2 dB / s
Slow memory decay rate	0.02 dB / s
Headroom	1 (0 dB)
Maximum limit	1 (0 dB)
Minimum limit	1e-5 (-100 dB)
Total delay	8 ms
Peak level attack time	0.25 ms
Peak level release time	10 ms

Table 7-1. *Processing settings.*

In all three experiments, the stimuli comprised four speech phrases (short sentences), each followed by a noise, which might be encountered in the respective application.

The assessment tasks common to the evaluation for each of the applications were to:

- rate the loudness of the speech and the loudness of the noise on a continuous scale with seven category labels shown, and
- rate the quality of the speech on a continuous scale with five category labels shown.

The category labels used for the loudness rating were those defined by Cox et al.:[157]

7. Uncomfortably loud

6. Loud, but OK

5. Comfortable, but slightly loud

4. Comfortable

3. Comfortable, but slightly soft

2. Soft

1. Very soft

The speech quality rating relative to the reference had a range from 0 to 100% with the category labels recommended by the International Telecommunications Union, Recommendation ITU-R BS.1534-3, 'Method for the subjective assessment of intermediate quality level of audio systems',[158] as follows:

Excellent	(90%)
Good	(70%)
Fair	(50%)
Poor	(30%)
Bad	(10%)

Each test was performed by 16 normal-hearing subjects (normal-hearing criteria: hearing thresholds ≤ 20 dB HL at standard audiometric frequencies). The experiment was conducted in a soundproof booth within the National Acoustic Laboratories. The stimuli were presented to the subjects from Beyerdynamic DT 990 Pro headphones. The stimuli applied to the headphones were reproduced from sound files stored on a computer. A computer program developed using MATLAB[146] provided instructions to the subjects via a computer screen in a textual format with interactive graphics. The subjects responded using a computer mouse. The presentation order of the stimuli was counterbalanced using a Graeco-Latin square. The experiments were double-blind and included a hidden reference. A hidden reference is defined by the

International Telecommunications Union as a 'reference not identified to the test subject'.[159] In these experiments the hidden reference was the unprocessed sound. It was always presented to the subjects for assessment along with the processed sound. When rating the speech quality, the subjects rated the hidden reference against an identified reference. As will be seen in the results section, there were some small differences in the subjects' speech quality ratings of the hidden reference compared to the identified reference although the two sounds were identical.

Figure 7-1 shows the screen presented to the subjects to perform the loudness ratings. The subjects were provided with a pair of buttons on the screen to play two stimuli, labeled 'A' and 'B'. Below the respective buttons was a pair of slider controls, one for speech loudness rating and the other for noise loudness rating, which the subjects adjusted using a computer mouse. They could listen to the stimuli as many times as they needed to in order to be satisfied with their ratings before moving on to the next stimulus or test. The unprocessed and SRL-processed stimuli were randomly assigned to the two buttons.

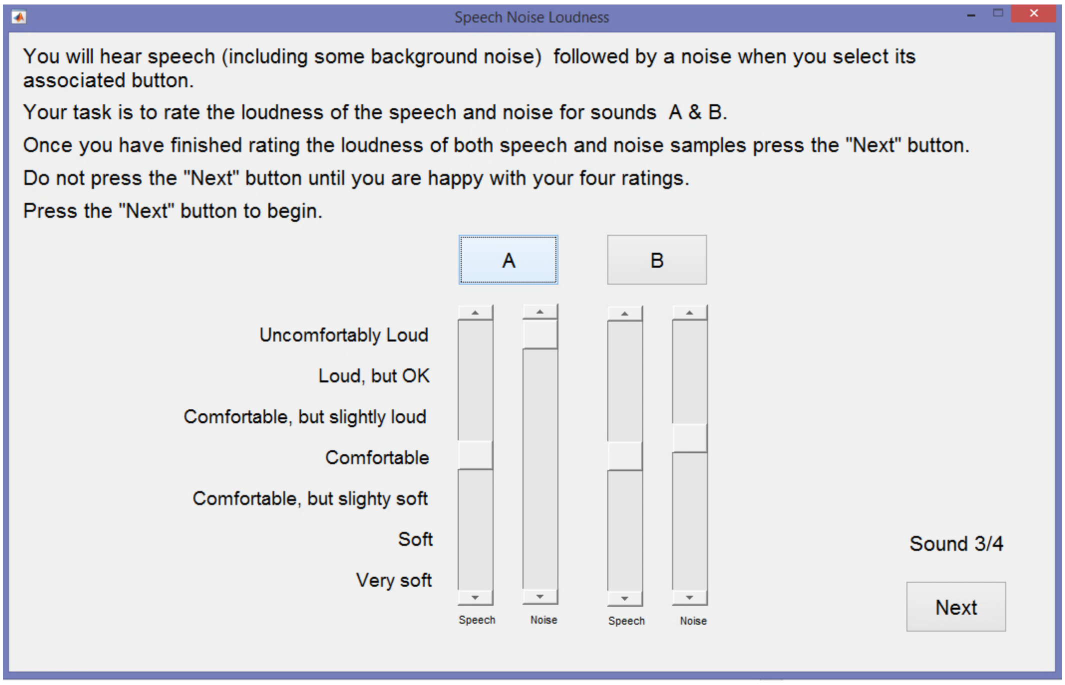


Figure 7-1. *Loudness rating screen.*

Figure 7-2 shows the screen presented to the subjects to perform the speech quality ratings. The subjects were provided with three buttons on the screen to play the reference (unprocessed) stimuli and stimuli A and B. Below each of the A and B buttons were slider controls to rate the speech quality of stimuli A and B, which the subjects adjusted using a computer mouse. The subjects were asked to compare each of the A and B stimuli with the reference and rate the speech quality. Only the speech portion of the recording was played. The subjects could listen to the stimuli as many times as they needed to in order to be satisfied with their ratings before moving on to the next stimuli or test. The unprocessed and SRL-processed stimuli were randomly assigned to buttons A and B.

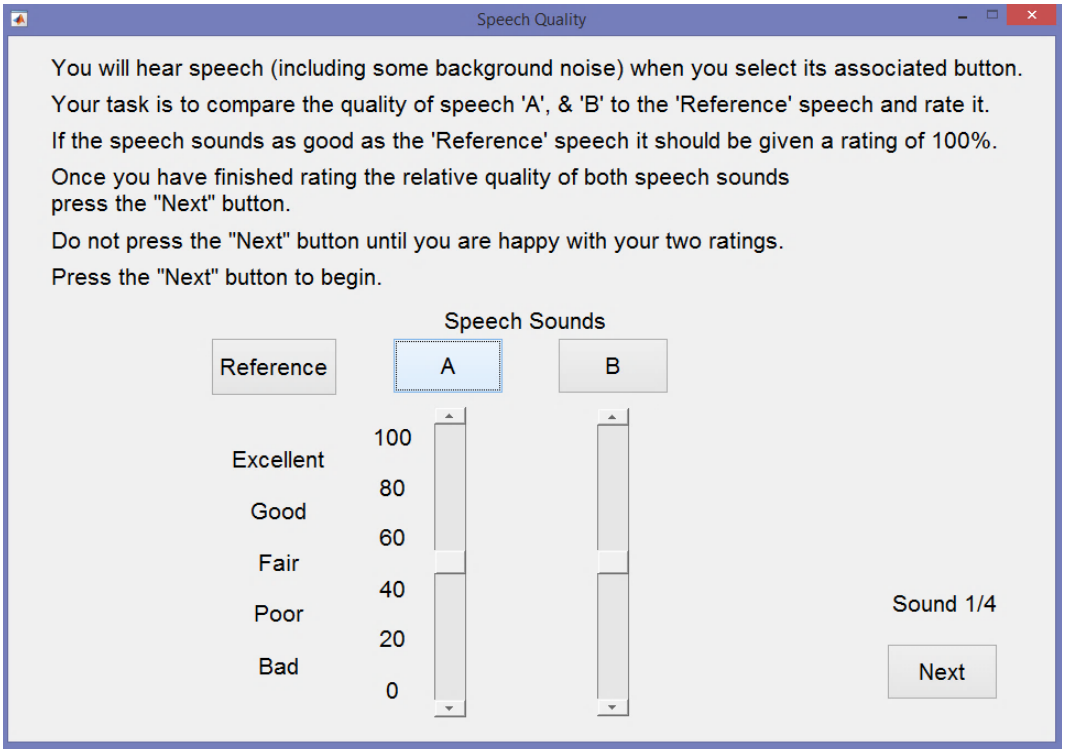


Figure 7-2. *Speech quality rating screen.*

7.2.1 Experiment 1: Hearing aid application

The stimuli comprised four recordings. Each recording comprised speech followed by a noise. Three of the four recordings were produced using an in-the-ear-canal microphone, such as used in an in-the-canal hearing-aid. The sampling rate was 48 kHz and resolution was 16 bits. The microphone was in the left ear of a person interacting with the person producing the phrase of speech. The fourth recording was made over the telephone. The stimuli were selected to be indicative of sounds encountered by hearing aid wearers. These comprised:

1. **Baby crying**: A woman on a phone talking and holding a baby who suddenly cries loudly into the telephone.

Speech level: 72 dB SPL RMS $_{\text{diffuse field equivalent,}}$

Noise level: 99 dB SPL RMS $_{\text{diffuse field equivalent.}}$

2. **Plates clanging**: A teenage male talking in the kitchen followed by the listener putting plates away in a cupboard.

   Speech level: 66 dB SPL RMS $_{\text{diffuse field equivalent,}}$

   Noise level: 87 dB SPL RMS $_{\text{diffuse field equivalent.}}$

3. **Vacuum cleaning**: A teenage male talking in the kitchen followed by the listener lending over close to a vacuum cleaner and starting it.

   Speech level: 72 dB SPL RMS diffuse field equivalent,

   Noise level: 93 dB SPL RMS diffuse field equivalent.

4. **Umpire's whistle**: A man shouting in a basketball court followed by a referee blowing a whistle.

   Speech level: 76 dB SPL RMS $_{\text{diffuse field equivalent,}}$

   Noise level: 81 dB SPL RMS $_{\text{diffuse field equivalent.}}$

The recordings were band-limited to 4 kHz to produce the unprocessed (reference) versions of the stimuli using Adobe Audition software.[160] The band limiting was employed to simulate the range of frequencies commonly made audible by hearing aids. The unprocessed stimuli were processed by the SRL MKI algorithm to produce SRL-processed versions of the stimuli. The resulting eight stimuli (four unprocessed and four SRL-processed sound files) were presented to the 16 subjects (7 females and 9 males) for loudness and quality assessment. The stimuli were presented to only one ear of the subjects to replicate listening with one hearing aid. Half the subjects received the presentation to the right ear and the other half to the left ear. The average presentation level of the speech component of the unprocessed stimuli was 71 dB SPL when measured at the eardrum reference point on a Head and Torso Simulator (HATS)[125] and converted to the diffuse field.

7.2.2 Experiment 2: Hearing protector application

As with the previous experiment, the stimuli comprised four recordings with each recording comprising speech followed by a noise. The four recordings were produced using a microphone mounted on the outside of a passive ear-muff worn by the person making the recording. The microphone location was about 40 mm out from the temple of the person making the recording. This microphone location was chosen to simulate the position of a microphone in a level-dependent hearing protector (earmuff). Recording techniques were otherwise as for Experiment 1. The dominant noises being recorded were produced by tools operated by the person wearing the microphone. The stimuli were selected to be indicative of sounds commonly encountered by hearing protector wearers using tools outdoors in a domestic environment. This comprised:

1. **Drill**: A man talking at a distance of about 1.5 metres from the microphone wearer immediately followed by the microphone wearer using a cordless power drill, in hammer mode, to drill into a brick.

   Speech level: 64 dB SPL RMS $_{\text{diffuse field equivalent,}}$

   Noise level: 90 dB SPL RMS $_{\text{diffuse field equivalent.}}$

2. **Hammer**: A man talking at a distance of about 1.5 metres from the microphone wearer immediately followed by the microphone wearer using a hammer to repeatedly strike a piece of wood. Note this recording has been reduced by 10 dB to accommodate the peak of the hammer's impact sound.

   Speech level: 53 dB SPL RMS $_{\text{diffuse field equivalent,}}$

   Noise level: 83 dB SPL RMS $_{\text{diffuse field equivalent.}}$

3. **Mower:** A man talking at a distance of about 1 metre from the microphone wearer immediately followed by the microphone wearer starting a lawn mower and running it over grass and twigs.

   Speech level: 65 dB SPL RMS $_{\text{diffuse field equivalent,}}$

   Noise level: 87 dB SPL RMS $_{\text{diffuse field equivalent.}}$

4. **Pressure cleaner**: A man talking at a distance of about 2 metres from the microphone wearer immediately followed by the microphone wearer starting a high-pressure cleaner and directing the pressurised water towards some paving stones.

   Speech level: 65 dB SPL RMS $_{\text{diffuse field equivalent,}}$

   Noise level: 90 dB SPL RMS $_{\text{diffuse field equivalent.}}$

The unprocessed stimuli (recordings) were processed by the SRL MKI algorithm to produce SRL-processed versions of the stimuli. The resulting eight stimuli (four unprocessed and four SRL-processed sound files) were presented to the 16 subjects (8 females and 8 males) for loudness and speech quality assessment. The stimuli were presented to both ears of the subjects. The average presentation level of the speech component of the unprocessed stimuli (excluding the speech associated with the hammer recording) was 65 dB SPL when measured at the eardrum reference point of a HATS and converted to the diffuse field. Due to the high peak level of the hammering sound, the gain needed to be lower to avoid clipping in the hammer recording, and the resulting presentation level of the speech was 54 dB SPL when measured at the eardrum reference point of a HATS and converted to the diffuse field.

7.2.3 Experiment 3: Telephone headset application

As with the previous two experiments, the stimuli comprised four recordings, with each recording comprising speech followed by a noise. The four recordings were sampled either directly from the telephone network signal or from the electrical signal applied to the receiver of a telephone. The speech samples were selected to be indicative of speech commonly encountered by call centre operators. The noises, while less often encountered, were typical of those that occur over the telephone network. The samples were edited so that the noise immediately followed the speech. Only in one case did the speech and noise come from the same recording. The sampling rate was 8 kHz and the resolution was 16 bit. This sampling rate is standard for narrow-band digital telephone networks. The edited recordings comprised:

1. **Fax machine**: A woman talking followed by a fax machine's initial tone sequence.

   Speech level: 70 dB SPL RMS $_{\text{diffuse field equivalent,}}$

   Noise level: 90 dB SPL RMS $_{\text{diffuse field equivalent.}}$

2. **Feedback**: A man talking followed by a feedback whistle produced by a cordless telephone.

   Speech level: 69 dB SPL RMS $_{\text{diffuse field equivalent,}}$

   Noise level: 82 dB SPL RMS $_{\text{diffuse field equivalent.}}$

3. **Music on hold:** A woman talking followed by "on hold" music being generated at a high level.

Speech level: 68 dB SPL RMS $_{\text{diffuse field equivalent}}$,

Noise level: 87 dB SPL RMS $_{\text{diffuse field equivalent}}$.

4. **Fault noise**: A man talking followed by a plethora of tones and noise produced by the telephone network when a fault occurred.

Speech level: 66 dB SPL RMS $_{\text{diffuse field equivalent}}$,

Noise level: 87 dB SPL RMS $_{\text{diffuse field equivalent}}$.

The unprocessed stimuli (recordings) were processed by the SRL MKI algorithm to produce SRL-processed versions of the stimuli. The resulting eight stimuli (four unprocessed and four SRL-processed sound files) were presented to the same 16 subjects as in Experiment 2, using the same procedures as for Experiment 2 with the exception that the stimuli were presented to only one ear of the subjects to replicate listening with a monaural headset. Half the subjects received the presentation to the right ear and the other half to the left ear. The average presentation level of the speech component of the unprocessed stimuli was 69 dB SPL when measured at the eardrum reference point of a HATS and converted to the diffuse field.

7.3 Results and discussion

The results from the three experiments were statistically analysed using the Statistica™ software package by Dell™. The mean ratings across subjects and 95% confidence intervals of the means were produced for all the measures. Further statistical analysis was performed using the analysis of variance (ANOVA) method.[161] The raw data and basic statistics in shown in Appendix B.

7.3.1 Experiment 1: Hearing aid application

7.3.1.1 Loudness

The results of the loudness ratings of speech and noise, encountered in hearing aid applications, are shown in Figure 7-3. The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the speech and noise in the unprocessed and SRL-processed conditions for each noise type.

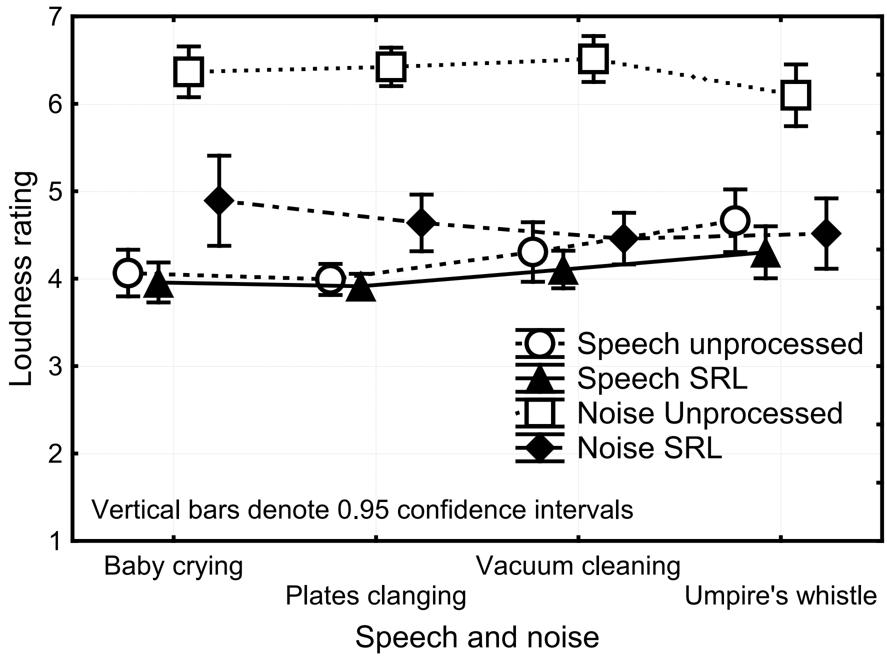


Figure 7-3. *Loudness of speech and noise encountered in hearing aid applications.*

7.3.1.2 Speech loudness

Both the unprocessed and SRL-processed speech was rated on average as being ‘comfortable’ to ‘comfortable, but slightly loud’ (categories 4 to 5). The difference between the loudness ratings of the unprocessed speech and the SRL-processed speech, averaged across the four noise types, was 0.2 loudness categories (i.e. 4.3 for unprocessed speech compared to 4.1 for SRL-processed speech).

A two-way repeated-measures ANOVA was performed on the speech loudness ratings with processing condition and noise type as factors. The main effect of processing condition was found to be statistically significant (p=0.011), although, the estimated mean difference was limited as processing reduced loudness by only 0.2 loudness categories on average, which was too small to be of any practical significance. Post hoc analysis was performed using the Bonferroni method[162] to compare the speech loudness ratings within each type of noise. No significant difference was found in the loudness ratings, due to processing condition, for the speech paired with the baby crying, plates clanging and vacuum cleaning noises. However, there was a significant difference in the loudness ratings, due to processing condition, for the speech paired with the umpire’s whistle noise (p=0.003). Although, statistically significant, there was only a 0.4 loudness category difference between the unprocessed and SRL-processed speech. The mean loudness ratings were 4.7 and 4.3 respectively, placing them both in the mid-range between ‘comfortable’ to ‘comfortable, but slightly loud’ (categories 4 and 5). The speech recording associated with the umpire’s whistle was made in a reverberant indoor basketball court and there was a significant level of noise present in the recording of the speech. Suppression of the reverberation and noise and possibly some speech by the SRL MKI algorithm may explain the small but statistically significant difference in the loudness rating of the speech.

7.3.1.3 Noise loudness

In contrast to the speech loudness results, there were highly significant differences in the noise loudness as a result of processing condition. All unprocessed noises were rated as being between ‘loud, but O.K.’ to ‘uncomfortably loud’ (categories 6 and 7). All the SRL-processed noises were rated as being between ‘comfortable’ to ‘comfortable, but slightly loud’ (categories 4 to 5). The difference between the loudness ratings of the unprocessed noise and SRL-processed noise, averaged across the four noise types, was 1.8 loudness categories (i.e. 6.4 for unprocessed noise compared to 4.6 for SRL-processed noise). A two-way repeated-measures ANOVA was performed on the loudness ratings of the noises with processing condition and noise type as factors. The results showed a highly significant effect of the processing condition ($p<0.001$). The main effect of noise type was not significant ($p=0.093$). However, the interaction between the processing condition and noise type was significant ($p=0.026$). Post hoc Bonferroni tests revealed that the effect of processing condition was highly significant for all noise types ($p<0.001$). The interaction between the processing condition and the noise type can be seen in Figure 7-3 where the loudness reduction as a result of processing for the baby crying was less than for the other three noises. This may be because the subjects were more sensitive to the sound of a baby crying than to other noises. It may also be because the SRL scheme favours sounds with speech-like spectral characteristics and a cry was spectrally similar to speech. Furthermore, it may be because the spectral components were all of a high level due to the baby’s cry being clipped as it passed through the telephone system. I speculate all these aspects play a part.

7.3.1.4 SRL-processed speech and noise loudness

The SRL method aims to reduce the loudness of non-speech sounds to a loudness similar to that of the immediately preceding speech sounds. An analysis of the loudness of the SRL-processed noise in relation to the SRL-processed speech was therefore of interest, i.e. how close was the SRL-processed noise to the SRL-processed speech? The average loudness rating of the SRL-processed noise was 4.6 and the SRL-processed speech was 4.1, a difference of 0.5 loudness categories. A two-way repeated-measures ANOVA was performed on the loudness ratings of SRL-processed stimuli with stimulus type (speech or noise) and noise type as factors. The results showed a highly significant effect of the stimulus type ($p<0.001$). The noise type was not significant ($p=0.414$). However, the interaction between the stimulus type and noise type was significant ($p=0.008$). Post hoc Bonferroni tests revealed that the loudness of the SRL-processed noise was not significantly different to the loudness of the SRL-processed speech for the vacuum cleaner ($p=0.831$) and the umpire's whistle ($p=1.000$). However, there were highly significant differences for the baby crying ($p<0.001$) and plates clanging ($p<0.001$) although the loudness of these sounds were within one loudness category range of the speech.

7.3.1.5 Speech quality

The results of the quality ratings of speech, encountered in hearing-aid applications, are shown in Figure 7-4. The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the unprocessed and SRL-processed speech for each noise type.

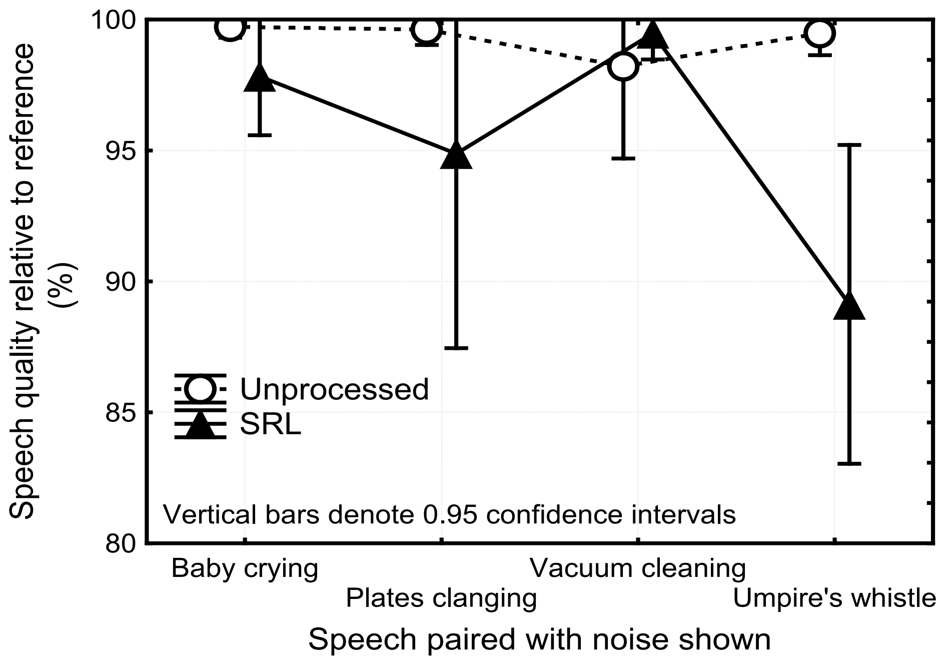


Figure 7-4. *Quality of speech encountered in hearing aid applications.*

All the speech was rated as being 'excellent' or better (90% or higher relative to the reference speech) except for the SRL-processed speech paired with the umpire's whistle which was rated just short of 90%. A two-way repeated-measures ANOVA was performed on the speech quality ratings with processing condition and noise type as factors. The results showed a significant effect of the processing condition ($p=0.009$). The noise type was also significant ($p=0.021$) as was the interaction between the processing condition and noise type ($p=0.008$). Post hoc Bonferroni tests revealed that the quality of the SRL-processed speech was not significantly different to the quality of the unprocessed speech paired with the baby crying ($p=1.000$), plates clanging ($p=1.000$) and vacuum cleaning ($p=1.000$).

However, the processing caused a statistically significant decrease in speech quality for the speech paired with the umpire's whistle ($p=0.002$). As discussed previously, in relation to speech loudness, the speech recording associated with the umpire's whistle was made in a reverberant indoor basketball court and there was a considerable level of noise (which includes competing speech) present in the recording of the speech.

The suppression of the reverberation and noise by the SRL MKI algorithm may explain this statistically significant difference in the speech quality rating. The variability in mean, as shown by the confidence intervals was larger for the SRL-processed speech compared to the unprocessed speech. This was quite evident for the speech paired with the plates clanging and paired with the umpire's whistle. Clearly the evaluation of changes in the speech as a result of SRL processing differs between subjects. The less than 100% average scores for the unprocessed stimuli (which was identical to the reference, i.e. it was the hidden reference) and the variability in the mean shows that the subjects were not perfect in their ability to accurately detect changes in the stimuli. This was clearly evident for the speech paired with the vacuum cleaner.

7.3.1.6 Summary

Overall, the most striking feature of the loudness ratings was the combination of a 1.8 loudness-category reduction in noise loudness with only a 0.2 loudness-category reduction in speech loudness as a result of SRL processing. This reduction in the loudness resulted in no statistically significant change in the speech quality in three out of the four cases and in the case where this was statistically significant the speech was rated just short of 90%.

7.3.2 Experiment 2: Hearing protector application

7.3.2.1 Loudness

The results of the loudness ratings of the speech and noise, in the hearing protector application, are shown in Figure 7-5. The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the speech and noise in the unprocessed and SRL-processed conditions for each noise type.

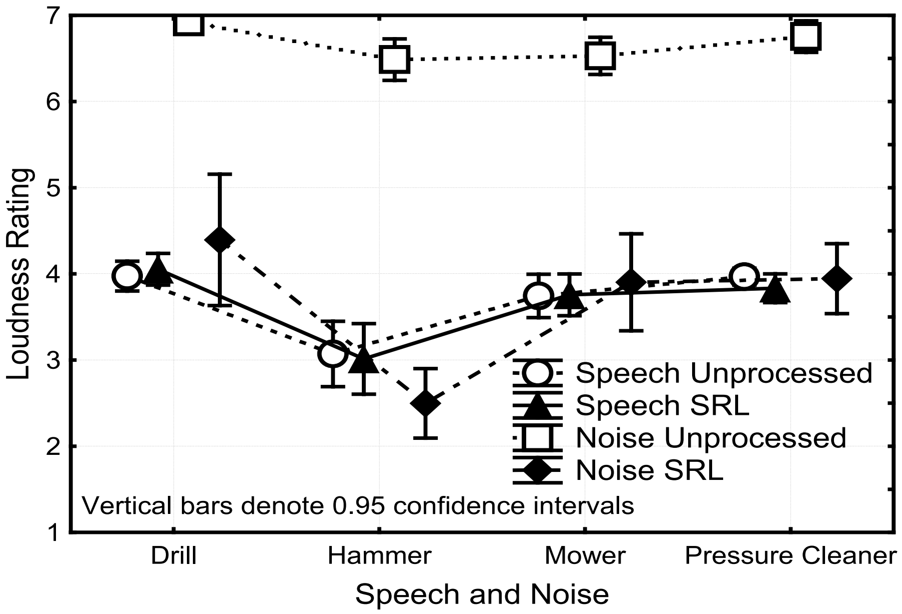


Figure 7-5. *Loudness of speech and noise encountered in hearing protector applications.*

7.3.2.2 Speech loudness

All speech was rated as being 'comfortable, but slightly soft' to 'comfortable' (categories 3 and 4). There was no difference in the average loudness rating of the unprocessed and SRL-processed speech; both had an average loudness rating of 3.7. A two-way repeated-measures ANOVA was performed on the speech loudness ratings with processing condition and noise type as factors. As expected, the effect of processing condition was not significant ($p=0.345$). The effect of noise type, however, was highly significant ($p<0.001$). However, the interaction between the processing condition and noise type was not significant ($p=0.052$). Post hoc Bonferroni tests comparing the loudness ratings of the SRL-processed speech with the unprocessed speech revealed that only the speech paired with the hammer noise was significantly different (p<0.001). This was because the unprocessed speech paired with the hammer noise was at a lower level so that the peak level of the hammer noise could be accommodated within the dynamic range of the recording

7.3.2.3 Noise loudness

In contrast to the speech loudness results, there were highly significant differences in the noise loudness as a result of processing condition. All unprocessed noises were rated as being between 'loud, but O.K.' to 'uncomfortably loud' (categories 6 to 7). All the SRL-processed noises were rated as being between 'soft' to 'comfortable, but slightly loud' (categories 2 to 5). The difference between the loudness ratings of the unprocessed noise and SRL-processed noise, averaged across the four noise types, was 3.0 loudness categories (i.e. 6.7 for unprocessed noise compared to 3.7 for SRL-processed noise). A two-way repeated-measures ANOVA was performed on the loudness ratings of the noises with processing condition and noise type as factors. As expected, the results showed a highly significant effect of the processing condition (p<0.001). The noise type was also highly significant (p<0.001) and the interaction between the processing condition and noise type was highly significant (p<0.001). Post hoc Bonferroni tests revealed that the effect of processing condition was highly significant for all noise types (p<0.001). Post hoc Bonferroni tests also revealed that the loudness of hammer noise was significantly different to the other noises (p<0.001), which was not unexpected given that the hammer was recorded at a lower level. This effect of noise type can be seen in Figure 7-5 where the loudness ratings of the unprocessed and SRL-processed hammer noise are less than corresponding loudness ratings for the other three noises. The interaction of the processing condition for the hammer noise is also evident in Figure 7-5. The reduction in the loudness rating of the hammer due to SRL-processing was 4.0 loudness categories (i.e. 6.5 for unprocessed hammer noise compared to 2.5 for SRL-processed hammer noise). The peak power control produced by the broadband gain calculator was very effective at controlling the hammer noise, possibly too effective, as it has reduced the hammer's loudness to below speech loudness.

7.3.2.4 SRL-processed speech and noise loudness

The average loudness rating of both the SRL-processed noises and speech was 3.7. A two-way repeated-measures ANOVA was performed on the loudness ratings of SRL-processed stimuli with stimulus type (speech or noise) and noise type as factors. As expected, the loudness of the SRL-processed noise was not significantly different to the loudness of the SRL-processed speech for all the noises (p=0.906). The noise type, however, had a highly significant effect on loudness (p<0.001) although the interaction between the stimulus type and noise type was not significant (p=0.052). Post hoc Bonferroni tests revealed that difference in loudness due to noise type was due to the lower level of the hammer noise and its associated speech. As discussed previously, the recording gain for the hammer stimulus was lower than for the other stimuli to accommodate the high peak level produced by the hammer, and hence the hammer noise and its associated speech both had lower RMS sound levels.

7.3.2.5 Speech quality

The results of the quality ratings of the speech, encountered in the hearing protector application, are shown in Figure 7-6. The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the unprocessed and SRL-processed speech for each noise type. There was no significant difference in the speech quality as a result of processing condition; all the speech was rated as being 'excellent' or better (90% or higher relative to the reference speech).

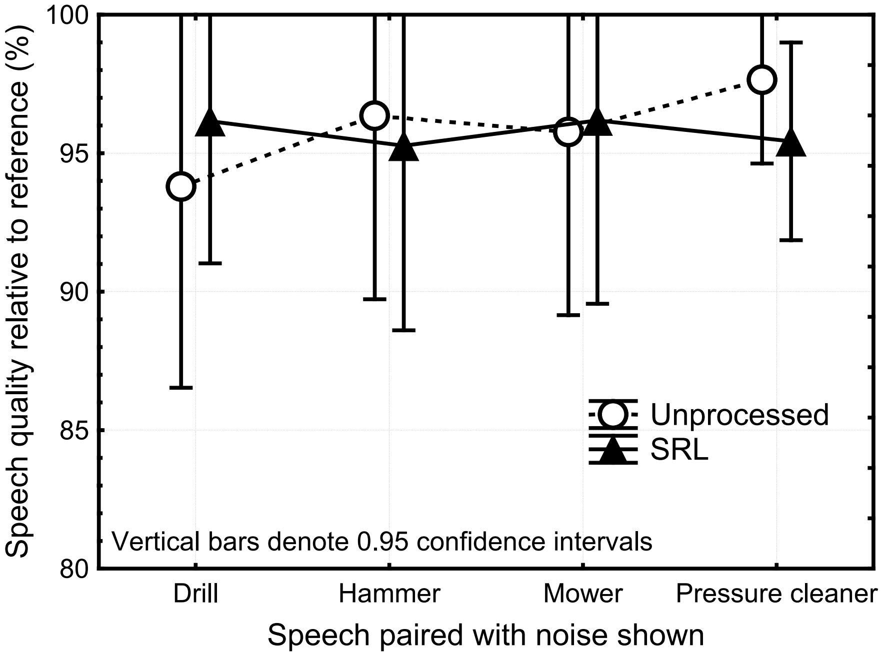


Figure 7-6. *Quality of speech encountered in hearing protector applications.*

A two-way repeated-measures ANOVA was performed on the speech quality ratings with processing condition and noise type as factors. As expected, the quality of the SRL-processed speech was not significantly different to the quality of unprocessed speech ($p=0.887$). The effect of noise type was not significant ($p=0.973$) and nor was the interaction between the processing condition and noise type ($p=0.278$). There was considerable variation in the subject's rating of the speech quality for both the SRL-processed stimuli and the unprocessed stimuli. This was consistent across the noise types. The lower than 100% rating for the hidden reference (i.e. the unprocessed speech compared to the unprocessed speech) indicates that some subjects were not reliable in their rating and/or the speech was hard to rate.

7.3.2.6 Summary

Overall, the most striking feature of the loudness ratings was the three loudness-category reduction in noise loudness combined with no reduction in speech loudness as a result of SRL processing. This processing resulted in no significant change in the speech quality.

7.3.3 Experiment 3: Telephone headset application

7.3.3.1 Loudness

The results of the loudness ratings of the speech and noise, in the telephone headset application, are shown in Figure 7-7. The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the speech and noise in the unprocessed and SRL-processed conditions for each noise type.

7.3.3.2 Speech loudness

All speech was rated as being 'comfortable' (category 4). There was no difference in the average loudness rating of the unprocessed and SRL-processed speech; both had an average loudness rating of 3.8. A two-way repeated-measures ANOVA was performed on the speech loudness ratings with processing condition and noise type as factors. As expected, the effect of processing condition was not significant ($p=0.598$). The effect of noise type was also not significant ($p=0.555$) nor was the interaction between the processing condition and noise type significant ($p=0.415$).

7.3.3.3 Noise loudness

In contrast to the speech loudness results, there were significant differences in the noise loudness as a result of processing condition. All unprocessed noises were rated as being between 'loud, but OK' to 'uncomfortably loud' (category 6 to 7). The SRL-processed noises were rated as being ‘comfortable, but slightly soft’ or ‘comfortable’ (categories 3 and 4).

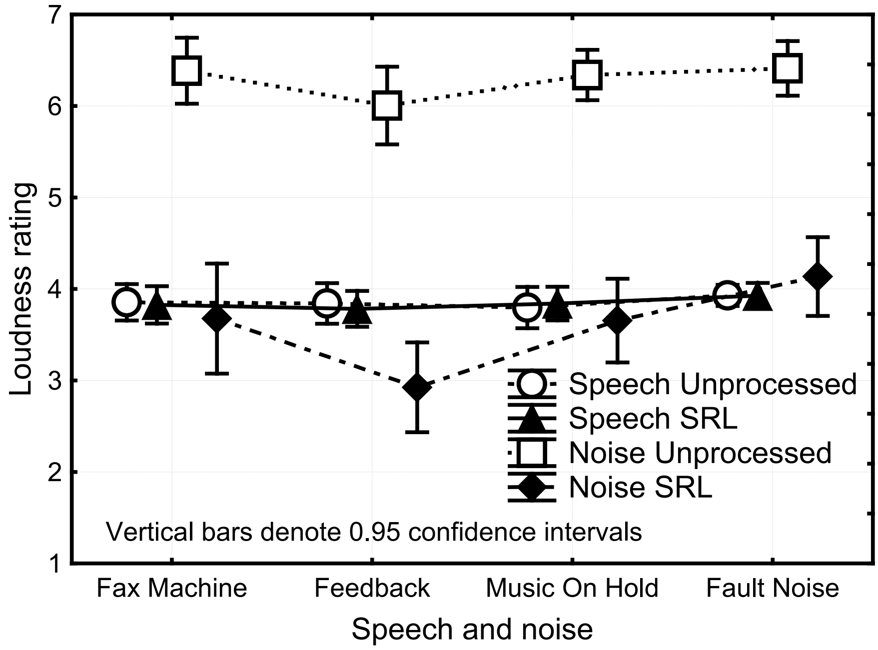


Figure 7-7. *Loudness of speech and noise encountered in telephone headset applications.*

The difference between the loudness ratings of the unprocessed noise and SRL-processed noise, averaged across the four noise types, was 2.7 loudness categories (i.e. 6.3 for unprocessed noise compared to 3.6 for SRL-processed noise). A two-way repeated-measures ANOVA was performed on the loudness ratings of the noises with processing condition and noise type as factors. As expected, the results showed a highly significant effect of the processing condition ($p<0.001$). The noise type was also highly significant ($p<0.001$) and the interaction between the processing condition and noise type was significant ($p=0.02$). Post hoc Bonferroni tests revealed that the effect of processing condition was highly significant for all noise types ($p<0.001$). The tests also revealed that the loudness rating for the feedback noise was significantly different to the other noises ($p<0.021$). This effect of noise type can be seen in Figure 7-7 where the loudness ratings of the unprocessed and SRL-processed feedback noise are less than corresponding loudness ratings for the other three noises.

7.3.3.4 SRL-processed speech and noise loudness

The average loudness rating of the SRL-processed noises was 3.6 and for SRL-processed speech it was 3.8, a difference of 0.2 loudness categories. A two-way repeated-measures ANOVA was performed on the loudness ratings of SRL-processed stimuli with stimulus type (speech or noise) and noise type as factors. The loudness of the SRL-processed noises was not significantly different to the loudness of the SRL-processed speech ($p=0.149$). The effect of the noise type on the loudness rating was highly significant ($p<0.001$) as was the interaction of the stimulus type and the noise ($p=0.001$). The effect of noise type was discussed in the preceding section on noise loudness. Post hoc Bonferroni tests of the interaction of stimulus type and noise type revealed that the significance was due to the feedback noise which was made significantly softer than the speech by the SRL processing ($p<0.001$). The reason for this was that this particular version of the SRL processing method uses frequency bands with linear spacing, rather than frequency bands more logarithmically spaced such as critical bands[45] or the equivalent rectangular bandwidth of auditory filters.[46] As a result, the speech energy in the high frequencies was lower and therefore a narrow-band signal, such as feedback noise, was suppressed to below the perceived speech level. For telephone headset applications, where acoustic shock is often reported to be caused by high-frequency narrow-band signals,[163] [2] [164] [165] this extra suppression may be desirable.

Like the hearing aid application and the hearing protector application, the most striking difference was the large (2.7 loudness-category) reduction in the loudness of the noises with no reduction in the speech loudness as a result of the SRL processing.

7.3.3.5 Speech quality

The results of the speech quality ratings are shown in Figure 7-8. The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the unprocessed and SRL-processed speech for each noise type. There was no significant difference in the speech quality as a result of processing condition; all speech was rated as being ‘excellent’ or better (90% or higher relative to the reference speech).

A two-way repeated-measures ANOVA was performed on the speech quality ratings with processing condition and noise type as factors. As expected, the quality of the SRL-processed speech was not significantly different to the quality of the unprocessed speech for all noise types (p=0.149). The effect of noise type was not significant (p=0.614) and nor was the interaction between the processing condition and noise type (p=0.574).

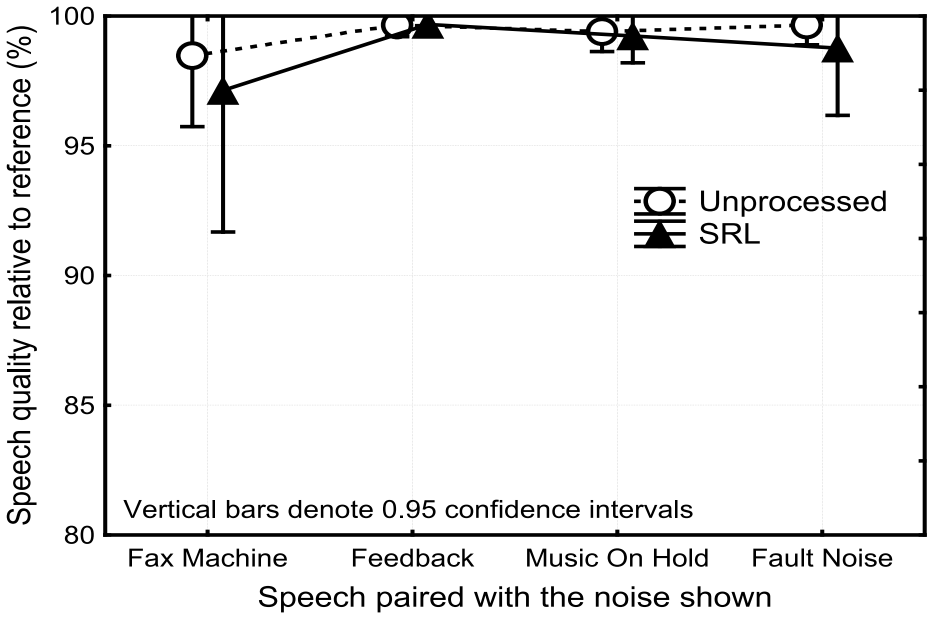


Figure 7-8. *Quality of speech encountered in telephone headset applications.*

7.3.3.6 Summary

Like the hearing aid application and the hearing protector application, the most striking difference was the large (2.7 loudness-category) reduction in the loudness of the noises with no reduction in the speech loudness as a result of the SRL processing. This occurred without a statistically significant change in the speech quality.

7.4 Conclusion and recommendations

From the experiment results, we can derive the following conclusions. First, the hypothesis that SRL processing controls the loudness of non-speech sounds relative to the loudness of speech sounds was true for higher level non-speech sounds. On average there was a 2.5 loudness category reduction in the loudness of louder non-speech sounds. In all three experiments, the loudness of higher-level non-speech sounds was brought closer to the loudness of speech and in most cases (9 out of 12) the loudness of the non-speech sounds was not significantly different to the speech loudness. In the three cases, where the non-speech sounds differed in loudness from the speech, two were less than one loudness category higher than the speech (i.e.

approaching ‘comfortable, but slightly loud’ when the speech was ‘comfortable’) and one was less than one loudness category below the speech loudness (i.e. ‘comfortable but slightly soft’ when the speech was just below ‘comfortable’). Interestingly, one of the two cases in which the loudness of the non-speech sound exceeded the speech was for a baby crying.

Only in one case out of twelve was the loudness of the SRL-processed speech significantly different (at a confidence level of $\alpha<0.05$) to the loudness of the unprocessed speech. This exception is considered likely to be a result of stronger reverberation and background noise (and possibly interfering speech) being present in the recording which was reduced by the SRL processing.

Second, the hypothesis that SRL processing does not degrade speech quality was found to be true for all samples of speech assessed except one. In all three experiments, the rated quality of the SRL-processed speech was not statistically different to the rated quality of the reference speech, with the exception of one stimulus, a male voice recorded in a basketball court, which was rated as being just under ‘excellent’. This was the same SRL-processed stimulus that was rated as having a difference in loudness compared to the unprocessed stimulus.

Other observations were:

1. The stimulus with the highest peak to RMS ratio (i.e. the hammer) was reduced in level to such an extent that its perceived loudness was lower than the preceding speech. This demonstrates the effectiveness of the broadband gain calculator / gain control at doing this without affecting the speech. However, it may have been too effective as it reduced the hammer’s loudness to below speech loudness. From a speech masking perspective this strong suppression is desirable, however, from a loudness perspective the hammer should be reduced to a comfortable loudness but not made to sound unnaturally weak.
2. The stimulus with the narrowest frequency width (i.e. the feedback) was also reduced in level to such an extent that its perceived loudness was lower than the preceding speech. This demonstrates the effectiveness of the multi-band gain calculator / adaptive modifier at doing this without affecting the speech. However, it too may be too effective in some applications where it is important to hear narrow-band sounds at speech levels or higher, e.g. alarm sounds.

Further evaluation of the SRL processing method using embodiments that employ more auditory-based frequency analysis is desirable to address the issue of over-suppression of narrow-band sounds in applications such as hearing aids and hearing protectors with amplification schemes. The concern is that the SRL MKI may over-suppress warning signals and therefore, while improving listening comfort and protecting the hearing of the wearer, it could endanger the wearer through over suppression of these important sounds. This appears to be a consequence of the narrow-band frequency analysis employed which, although not auditory based, has, in general, closely matched the loudness of non-speech sounds to that of immediately preceding speech sounds.

One question that could be asked is whether such results could have been obtained with a multi-band fixed-reference limiter. In order for similar results to be achieved, the fixed limits would need to be very carefully set just above the maximum speech levels in each case. Doing so would prevent the limiting of speech so that there was no change in the perceived loudness or quality of the speech and would allow the noises to be limited so that their perceived loudness was similar to the speech loudness. This would be possible with fixed limits but only if the speech level remained constant. If the speech level increased, then both its loudness and quality would be reduced and, if it decreased, then the loudness of the noises would exceed the speech loudness. Certainly, one could adjust the limit levels manually but this would be impractical. In contrast, the SRL processing automatically does this for the listener.

One could also ask if such results could have been obtained with a limiting scheme that adaptively adjust its limiting thresholds based on signal statistics, such as setting limiting thresholds in multiple bands to the 95% cumulative histogram level of the signal. As discussed in Chapter 4, a multiband scheme using percentile estimates of the input signal was developed for hearing aids by Ludvigsen (1997)[119] and another scheme, using the percentile estimates of the output signal, called Adaptive Dynamic Range Compression (ADRO) was developed by Blamey et al. (2005).[120,166] These approaches would appear preferable to fixed threshold limiting systems for controlling extraneous noises, however, they would not, in theory, be as effective in controlling extraneous noises while preserving speech as a scheme for which the limiting thresholds were adaptively determined by the level of the speech itself, as is the case with the SRL scheme.

While these experiments clearly demonstrate the ability for SRL to reduce the loudness of non-speech sounds and preserve the quality of speech, they do not adequately demonstrate SRL's adaptability to a variety of speech levels. It was therefore decided that future evaluation should include a greater variety of speech levels. Furthermore, the excellent results found in these experiments, might not be indicative of SRL's performance in situations where high-level background noise was present with the speech. It was therefore decided that future evaluation of SRL needed to include high-level noise simultaneously presented with the speech as well as sequentially presented.

Chapter 8

SRL MKII scheme

8 SRL MKII scheme

8.1 Introduction

The SRL MKII (mark 2) scheme is the second version of the SRL scheme, designed to address issues that arose during the implementation and evaluation of SRL MKI as well as to perform with greater reliability in a diversity of applications. The SRL MKI scheme was originally developed with narrow-band telecommunications applications in mind and was later extended for use in wideband telecommunications applications, hearing aids and level-dependent hearing protectors. However, it became apparent that the SRL MKI scheme had limitations in non-telecommunications applications. This was not unexpected as its single, non-frequency-specific, speech dominance detector was not suited to applications in which the speech to background noise ratio could vary strongly across frequency.

Evaluation of the SRL MKI scheme with its single, non-frequency-specific, speech dominance detector revealed that two types of errors were occurring. The first was that noise that dominated speech at some frequencies was included along with speech that dominated noise at other frequencies, leading to corruption of the speech level estimates at those frequencies where noise dominated. The second was that speech that dominated noise at some frequencies was not used to update the speech level estimates because noise was dominating speech at other frequencies. It was therefore deemed necessary to implement the frequency specific approach to speech dominance detection, as envisaged when the scheme was first proposed (see Chapter 5).

The formal subjective evaluation of the SRL MKI scheme documented in Chapter 7 identified two further issues that needed to be addressed. First, there was greater suppression of narrow-band high-frequency sounds than was needed for most applications. Second, while the length of the signal delay introduced by the SRL MKI scheme was acceptable for telecommunications applications, in which there was good echo cancellation, it was longer than desirable for the extended applications, such as in hearing aids and level-dependent hearing protectors, which allowed the user to hear their own voice in real-time through delayed processing.

Other issues that needed to be addressed also became clear. While the peak energy control worked very effectively in the SRL MKI evaluation, its effectiveness was dependent on the phase and magnitude relationship of the frequency components making up the peak. If the phase relationship of the components changed after being controlled, then the peak control was less effective. Phase changes could occur to the signal after the control had been applied, for example, in earphones and other reproduction equipment. Furthermore, although peak control indirectly reduced the effect of loudness summation resulting from brief peaks, it did not address loudness summation for longer duration sounds, nor did it take into account the relative loudness or level of the signal compared with the speech on a frequency-specific basis.

Feedback from both researchers and industry in response to presentations of the SRL MKI scheme also indicated that a more flexible architecture would enable incorporation of SRL into their testing platforms and enable further research to be performed in a broader range of applications.

To address these issues, SRL MKII was coded from the ground up to incorporate new methods of speech dominance detection and gain calculation and to be more flexible.

8.2 Overall architecture

The overall architecture of SRL MKII is the same as that presented in Figure 5-6, in Chapter 5. This is redrawn in Figure 8-1. The processes that are unique to the SRL scheme are the speech dominance detector and the speech-referenced gain-calculator. The input signal is applied to the analyser which produces a set of band signals, Signal (i). The speech dominance detector analyses the input signal, from which the band signals were obtained, and generates a set of speech dominance signals. The gain calculator accepts the band signals and the speech dominance signals and produces a set of band-specific gains, Gain (i). The adaptive modifier accepts the input signal and modifies it in accordance with the gains it receives from the gain calculator to produce the output signal.

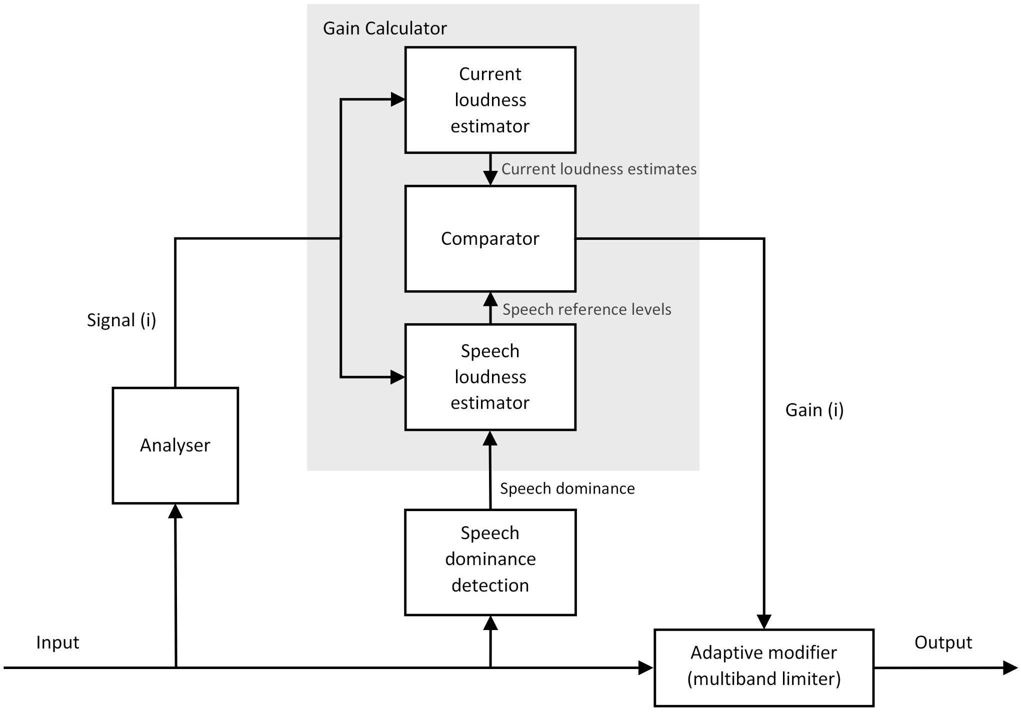


Figure 8-1. *The simplified SRL scheme with a separate analyser.*

As discussed in earlier chapters, there are many methods by which the analysis and adaptive modification processes can be performed. The SRL MKII gain calculator is compatible with most of these methods. For example, the SRL MKII gain calculator and speech dominance detector have been successfully incorporated into a level-dependent, sound-restoring hearing protector that uses the short-term filter-bank method[134] to perform analysis and adaptive modification, this is an alternative to the method described in this chapter and in Chapter 6.

The analyser and the adaptive modifier may share common processes. For example, the adaptive modifier may use frequency modification techniques such as overlap-add[141] or short-term filter bank methods[134] both of which require overlapped short-term fast Fourier transform (FFT) analysis which can also be used by the analyser.

The interface between the speech dominance detector and the speech loudness estimator within the gain calculator is substantially different in the SRL MKII scheme, compared with the SRL MKI scheme. Rather than one speech dominance signal, it comprises the nine signals shown in Table 8-1 and explained in Section 8.4.

Number	Speech dominance signal
1	Voice* present
2	Voice* lower frequency
3	Voice* upper frequency
4	Voice* commit
5	Voice* commit max upper frequency
6	Sibilance present
7	Sibilance lower frequency
8	Sibilance upper frequency
9	Sibilance commit

**Voice refers to voicing*

Table 8-1. *Speech dominance signals.*

The incorporation of the SRL MKII gain calculator and speech dominance detector into the implemented SRL MKII scheme is shown in Figure 8-2. The analyser and spectral modifier have the same structure as those in SRL MKI, they support a variety of sample rates and resolutions, as discussed later in this chapter. The most obvious difference between the SRL MKII and MKI schemes is the removal of the broadband gain calculator and multiplier from the MKI scheme. Its function has been replaced by a loudness summation feature within the multi-band gain calculator along with appropriate time aligning of the signal using the delay preceding the adaptive modifier.

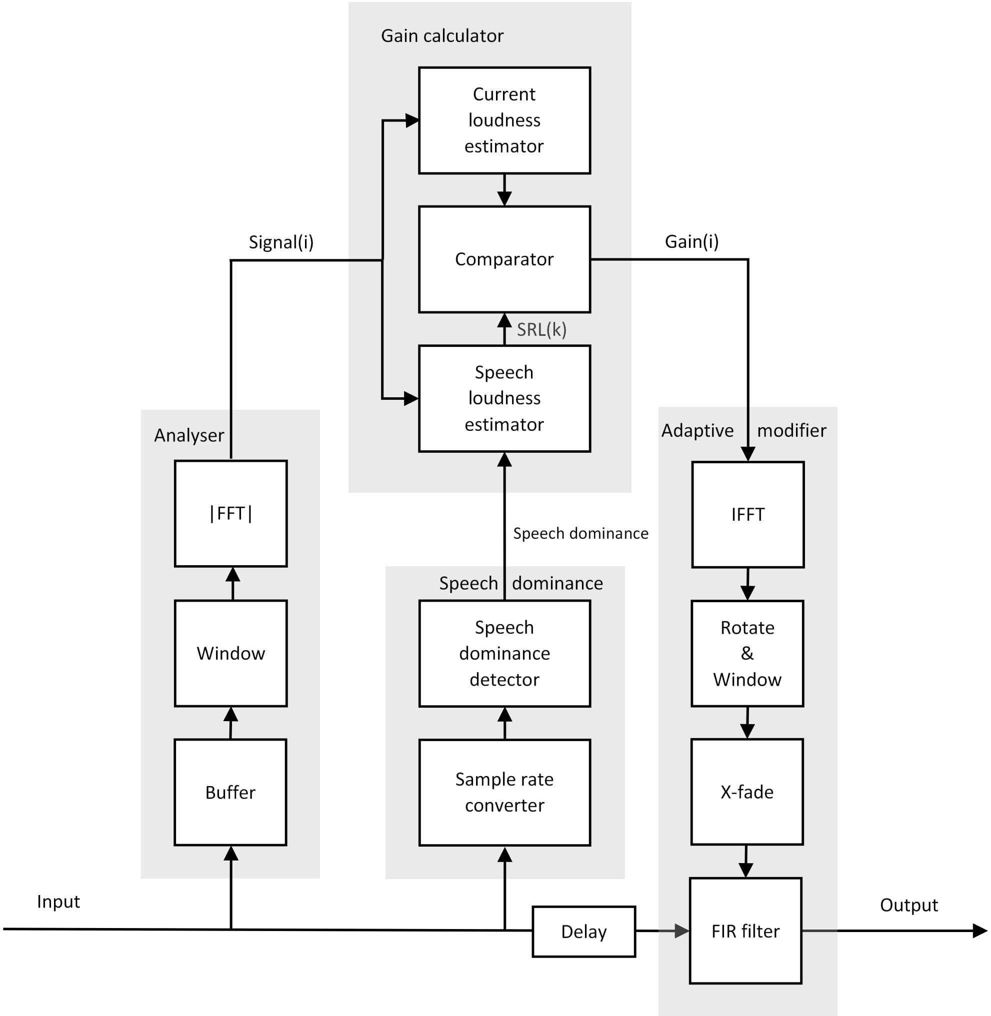


Figure 8-2. *SRL MKII scheme.*

8.3 Flexibility and parameter selection

The implemented scheme supports the system parameters shown in Table 8-2. In addition to these parameters are a number of thresholds and initialisation levels.

The scheme performs processing at the five standard sampling rates shown in Table 8-2. In addition to these, the Windows application (see later in this chapter) also supports the 44.1 kHz, 22.05 kHz, 12 kHz and 11.025 kHz sampling rates. A signal with one of these additional sampling rates is converted to the next highest sampling rate out of the five standard SRL sampling rates for processing and then back to its original sampling rate after processing. This covers the standard sampling rates as defined by the Audio Engineering Society,[167] and the European Broadcast Union,[168] and employed in telecommunication systems.[169,170]

The analysis rate is flexible. At a rate of 1000 Hz (i.e. an analysis period of 1 ms) the analysis is reaching the limit of the monaural temporal acuity of the human auditory system discussed by Durrant and Lovrinic.[171] Higher or lower rates can be employed although, with a decreased rate, the ability of the scheme to control transients reduces and the delay required to prevent overshoot increases meaning a separate broadband control of the signal is required as was the case with the SRL MKI scheme. An analysis sampling rate of 1000 Hz was chosen for the SRL MKII evaluation to enable fast control over transients.

The analysis duration is also flexible, although, with the window function chosen, it needs be at least twice the duration of the analysis period so that the energy of the signal is entirely captured. A duration of less than twice the analysis period would result in less than 50% overlap, the minimum overlap for a signal to be completely captured when using a Hann window.[150] There are other aspects of resolution and efficiency that are traded in the analysis, as discussed in earlier chapters. The selected analysis duration of 4 ms has, to some extent, sacrificed low-frequency spectral resolution for finer temporal resolution and higher efficiency with a 75% overlap of analysis windows. Compared with the 8 ms window used in the SRL MKI, the temporal resolution is twice as fine, making the multi-band system far more responsive to transients and improving its capacity to control peaks with less than half the delay of the SRL MKI. The cost of this higher temporal resolution is poorer frequency resolution, which largely manifests as an inability of the system to achieve auditory-based filter bandwidths (e.g. Bark or ERB) or third-octave filter bandwidths in the low frequencies.

Parameter	Restriction	Selected	
Sampling Rate	8, 16, 24, 32 & 48 kHz		
Analysis rate	no more than half the sampling rate	1000	Hz
Analysis period	at least twice the sampling period	1	ms
Analysis duration	at least twice the analysis period	4	ms
Analysis overlap	at least 50%	75	%
Discrete Fourier analysis resolution (DFT BW)	determined by analysis duration	250	Hz
Band structures (scales)		Bark	
Bark	BW >= min DFT BW if using DFT analysis		
ERB	"		
Linear	same as DFT BW		
Third octave	BW >= min DFT BW if using DFT analysis		
Fast-limiter active	yes/no	yes	
Slow-limiter active	yes/no	yes	
Fast-limiter attack time	≥ 2 times analysis period	2	ms
Fast-limiter release time	≥ analysis duration	250	ms
Slow-limiter attack time	≥ 2 times analysis period	100	ms
Slow-limiter release time	≥ analysis duration	500	ms
Fast to slow limit threshold relationship	≥ 0 dB	8	dB
Loudness summation		Bark	
Bark	all band structures except ERB		
ERB	ERB band structure		
Loudness-summation max. reduction - fast	≥ 0 dB	10	dB
Loudness-summation max. reduction - slow	≥ 0 dB	8	dB
Speech envelope release rate	≥ 0 dB / sec	12	dB /sec
Time period for a new speech update	> analysis period	200	ms
Rate of convergence to new speech	> analysis period	100	ms
Band linking: maximum inter-band difference	≥ 0 dB	30	dB
Filter duration	equal to the analysis duration	4	ms
Filter delay	equal to half the analysis duration	2	ms
Alignment delay	typically equal to the analysis period	1	ms
Total signal delay	filter delay + alignment delay	3	ms

Table 8-2. *SRL MKII parameters.*

Two loudness summation scales were implemented, the Bark and the ERB. The ERB scale is used when the ERB frequency scale is used for the specific loudness estimation and the Bark scale is used otherwise. These are discussed in the loudness summation section later in this chapter.

The adaptive modifier length and delay are not independently selectable parameters as they are a consequence of the analysis period and analysis duration selected. The alignment delay prior to the adaptive modifier has been selected to match the input buffer's duration (i.e. the analysis period). This results in the gain modification coefficients generated within the adaptive modifier being applied to the cross fade process within the adaptive modifier (see Chapter 6 for details) at the same time as the signal enters the filter. These coefficients reach their full effect 1 ms before the filter produces its maximum output in response to the signal. When the filter's impulse response possesses a low kurtosis, such as when the gain function is smooth, the modification is typically 1 ms in advance of the signal. However, when the gain function is more discontinuous, the filter's impulse response is broad and its output therefore also includes signal energy up to 1 ms in advance of the modification peak and up to 3 ms behind it. Due to the spread of up to 4 ms in the impulse response, there is a corresponding spread of input energy contributing to the output. A slightly advanced modification in relation to the signal is therefore beneficial and achieves greater control of transient sounds such as impact noise. The alignment delay is therefore 1 ms.

8.4 Speech dominance detection and its performance

The speech dominance detector (SDD) is the director of the speech loudness estimator within the SRL scheme. It operates with a sampling rate of 16 kHz and is preceded by the sampling rate converter (see Chapter 6). The SDD analyses the signal every 20 ms and provides the control signals as shown in Table 8-3 to the gain calculator at this rate. These control signals are overlaid on the waveform and spectrogram of speech followed by noise in the example shown in Figure 8-3.

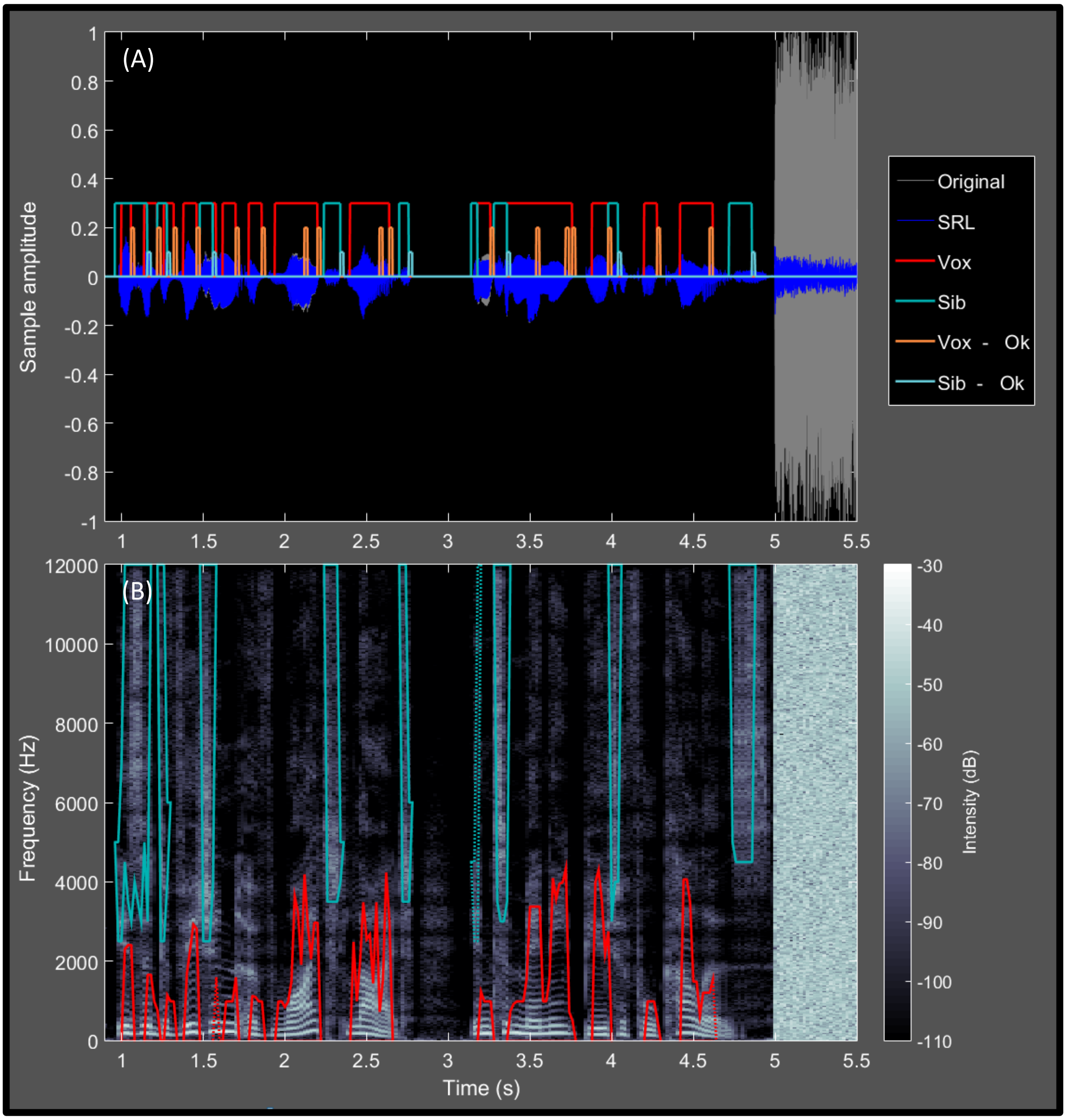


Figure 8-3. (A) *Waveform and* ***(B)*** *spectrogram of 4 seconds of speech followed by a burst of white noise overlaid with speech dominance detector signals.*

The upper plot, (A) of Figure 8-3 shows the waveform of a 4 second segment of speech followed by a burst of white noise. The original signal is shown in grey and the SRL-processed signal in blue. Overlaid on this waveform are the 'voicing' present (also referred to as the 'voice' present) signal, which is labelled as 'Vox', and coloured red and the sibilance present signal, which is labelled as 'sib', and coloured light blue. These are binary signals in which a high level represents the presence of voicing (red trace) or sibilance (light blue trace) and a low level (i.e. 0) represents its absence.

Number	Speech dominance signal	Colour	Figure type
1	Voice present (labelled as 'Vox')	Red	Waveform
2	Voice lower frequency	Red	Spectrogram
3	Voice upper frequency	Red	Spectrogram
4	Voice commit (labelled as 'Vox – Ok')	Orange	Waveform
5	Voice commit max upper frequency*		
6	Sibilance present (labelled as ' Sib')	Light blue	Waveform
7	Sibilance lower frequency	Light blue	Spectrogram
8	Sibilance upper frequency	Light blue	Spectrogram
9	Sibilance commit (labelled as 'Sib – Ok')	Bright blue	Waveform

* The voice max upper frequency is not visible by itself in this figure, its effect, however, is incorporated into limiting the maximum voice upper frequency.

Table 8-3. *Speech dominance signals and their identification features for the displays in Figure 8-3.*

Also overlaid on the waveform in Figure 8-3 is the voice commit signal, which is approximately two thirds of the height of the voicing (voice) present signal, it is labelled as 'Vox – Ok', and coloured orange. It appears briefly after an occurrence of uncorrupted voicing that is present for 200 ms or that ceases. Also shown on this waveform is the sibilance commit signal, which is about one third the height of the sibilance present signal, it is labelled as 'Sib – Ok', and coloured bright blue. It appears briefly when an occurrence of uncorrupted sibilance ends.

The lower plot, (B) of Figure 8-3 is a spectrogram of the original (unprocessed) signal. Overlaid on this is the voicing sampling region from the voice lower frequency to the voice upper frequency, which is shown in red. Also overlaid on this spectrogram is the sibilance sampling region from the lower sibilance frequency to the upper sibilance frequency, which is shown in light blue. If the voicing or the sibilance is not committed, then it is shown as a dotted line on the spectrogram as is the case for a sibilance detection located approximately midway during the speech segment.

The effect of the voice commit max upper frequency can be seen in the voicing overlay on the spectrogram where the highest frequency of voicing has a consistent maximum and appears flattened on occasions.

In this example, the voicing does not extend far above 4000 Hz and the sibilance does not extend below 2500 Hz. In the crossover region, between 2500 Hz and 4000 Hz, both voicing and sibilance contribute to the estimation of the speech reference levels. Below 2500 Hz, the speech reference levels are predominantly a function of the voicing and above 4000 Hz the speech reference levels are predominately a function of sibilance.

Also evident in the waveform of Figure 8-3 is a point at which the unprocessed speech, in grey, is revealed from behind the processed speech, in blue. In this short time region, the speech signal has been slightly attenuated due to the SRL processing. This may occur when new speech follows an absence of speech during which the speech reference level may have decayed a little. This is an example of how close the limiting levels come to the maximum speech levels. The effect of this small amount of attenuation is, in most cases, inaudible. The perceptual effect of this, on the loudness and quality of speech, was assessed in the evaluation of SRL MKII reported in Chapter 9.

Also evident in the waveform of Figure 8-3 is the substantial attenuation of the noise burst following the speech to approximately the same level as the speech. During the noise there is no false triggering of the SDD.

As mentioned in previous chapters, the exact details of the method of speech dominance detection cannot be disclosed within this thesis. However, the input and output signals have been described in detail. Furthermore, in Chapter 5, it was disclosed that a rule-based scheme was developed that analysed the time-varying harmonic content of the voiced spectra using both spectral and cepstral analysis. Also, in Chapter 5, it was disclosed that the detection of dominant sibilance involved a rule-based scheme that analysed the time-varying spectrum.

Figure 8-4 further demonstrates SRL MKII's performance. For this demonstration, the initial speech reference levels were set low so that the speech dominance detection and adaptation speed of SRL to new higher-level speech could be seen. Sub-plot (D) shows that the initial 200-300 ms of speech was attenuated pending the first 200 ms of voicing being committed after which there was up to 100 ms during which the limit was increased to this new speech level. It can also be seen in this initial speech segment that the sibilance detected by the SDD was not considered good enough by the SDD to commit. This can be noted by the lack of a sibilance commit signal following the brief periods of detection on the waveform sub-plot (A) and also observed as dotted blue lines on the spectrogram sub-plot (B). This, however, did not result in the sibilance being attenuated. As will become apparent in Section 8.5.3.11, the scheme extrapolates its high-frequency speech reference levels, based on the slope of the voicing spectra, in the absence of sibilance being detected.

Figure 8-4 also demonstrates the heavy attenuation a power drill underwent, as shown in sub-plot (D), bringing it down to levels similar to the preceding speech as shown in sub-plots (A) and (C). The heavy attenuation, however, was completely released when the low-level speech appeared after the power drill. Although not attenuated, it took two voiced utterances before this low-level speech was detected and the speech reference levels started to converge to this lower speech level. Following this, the SDD commits six out of the possible six occurrences of voicing and both occurrences of sibilance which were at a low level. The hammering that followed this low-level speech was heavily suppressed as shown in sub-plot (D) to levels similar to the preceding speech as shown in sub-plots (A) and (C). This suppressed hammering level, however, corresponded to a perceptual level below that of the speech, i.e. an over-suppression. This perceived over-suppression became apparent during the subjective evaluation of SRL MKII documented in Chapter 9. A higher minimum bound on the speech reference level would limit this over-suppression.

It can also be seen in sub-plot (B) that there is a gap between the frequency regions in which there was voicing (voice) and sibilance detection during the low-level speech. This gap sometimes occurs when the speech is low level or there is noise present in this upper mid-frequency region. This absence of speech dominance detection in this region is addressed in Section 8.5.3.12 which describes how the scheme interpolates speech reference levels within this gap region.

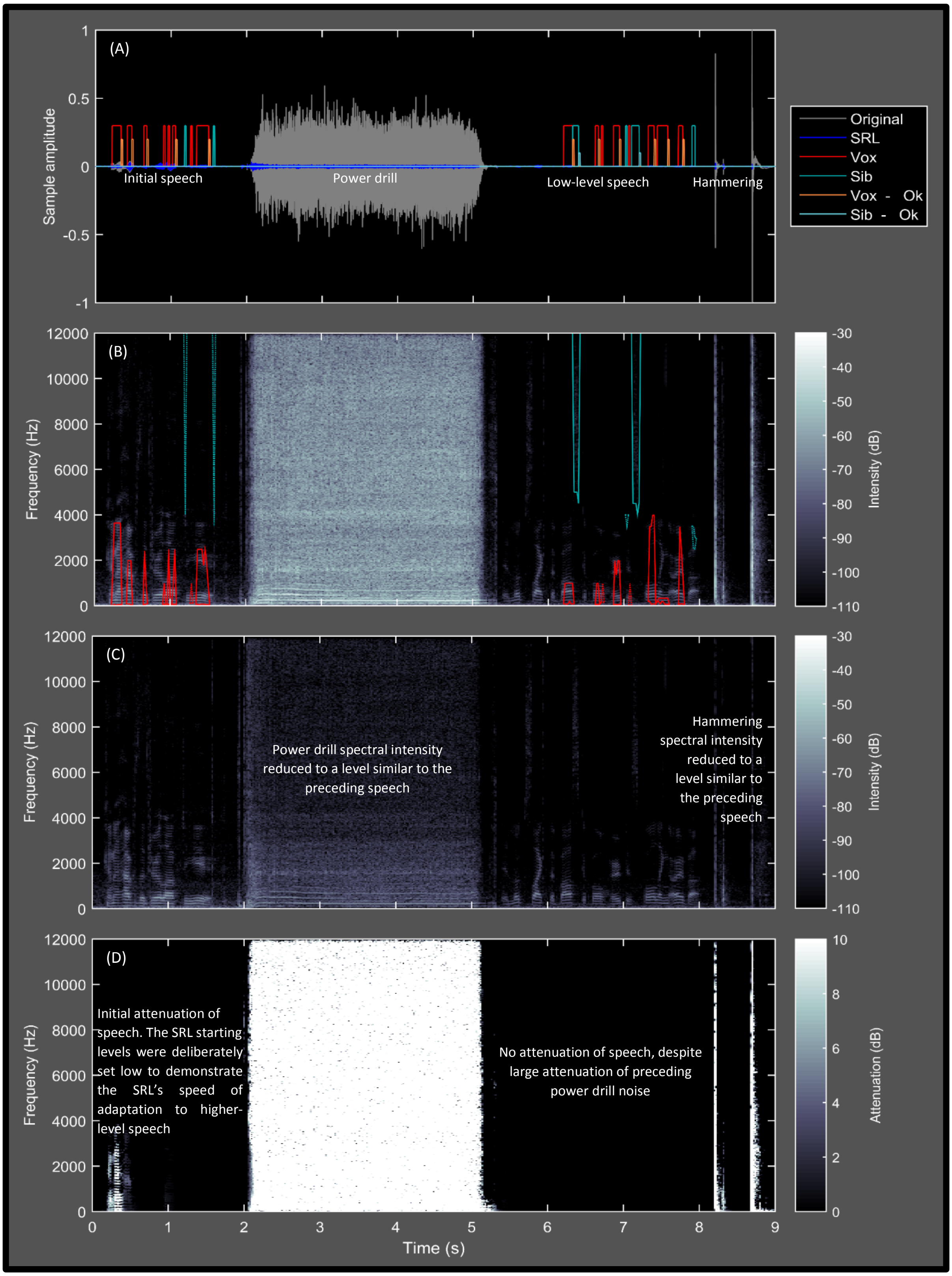


Figure 8-4. (A) *Waveforms (original – grey, SRL-processed - blue) of speech, followed by: a power drill, low-level speech and hammering, overlayed with SDD signals,* ***(B)*** *spectrogram of the original signal overlayed with SDD regions,* ***(C)*** *spectrogram of SRL-processed signal and* ***(D)*** *spectrogram of attenuation.*

Figure 8-5 provides another demonstration of SRL's performance and particularly that of the SDD. In this case, the signal was a recording of speech from a wideband telephone headset signal which occurs simultaneously with some high-level telephone dialling tones (i.e. dual tone modulated frequency, DTMF). Sub-plot (A) shows the attenuation of the waveform (blue versus grey waveform) with a strong level decrease in the tone bursts and minimal decrease in the speech. As shown in sub-plot (B), the speech was sampled both between and during the tone bursts in frequency regions in which the tones were not present. The scheme attempts to get all the samples it can of the speech-dominated regions to create the speech reference levels. The spectral intensity of the resulting SRL-processed signal is shown in sub-plot (C). It is apparent that the tones have been reduced to a similar spectral intensity to the speech in the same regions. Sub-plot (D) shows a spectrogram of the attenuation produced by the scheme. The strong reduction in the tones is clearly evident as is the minimal reduction in the speech. Only when the speech was simultaneously present with the noise in the same frequency region did it undergo substantial attenuation and only in that spectro-temporal region. The attenuation of speech diminished above and below the frequency region of the simultaneously present tones. Also, despite the considerable attenuation of the third presentation of tones, there was no sign of attenuation of the speech immediately following it.

The plot also shows that the SRL scheme was processing quickly and that it smoothly attenuated the tones and did not enhance either their onsets or their offsets.

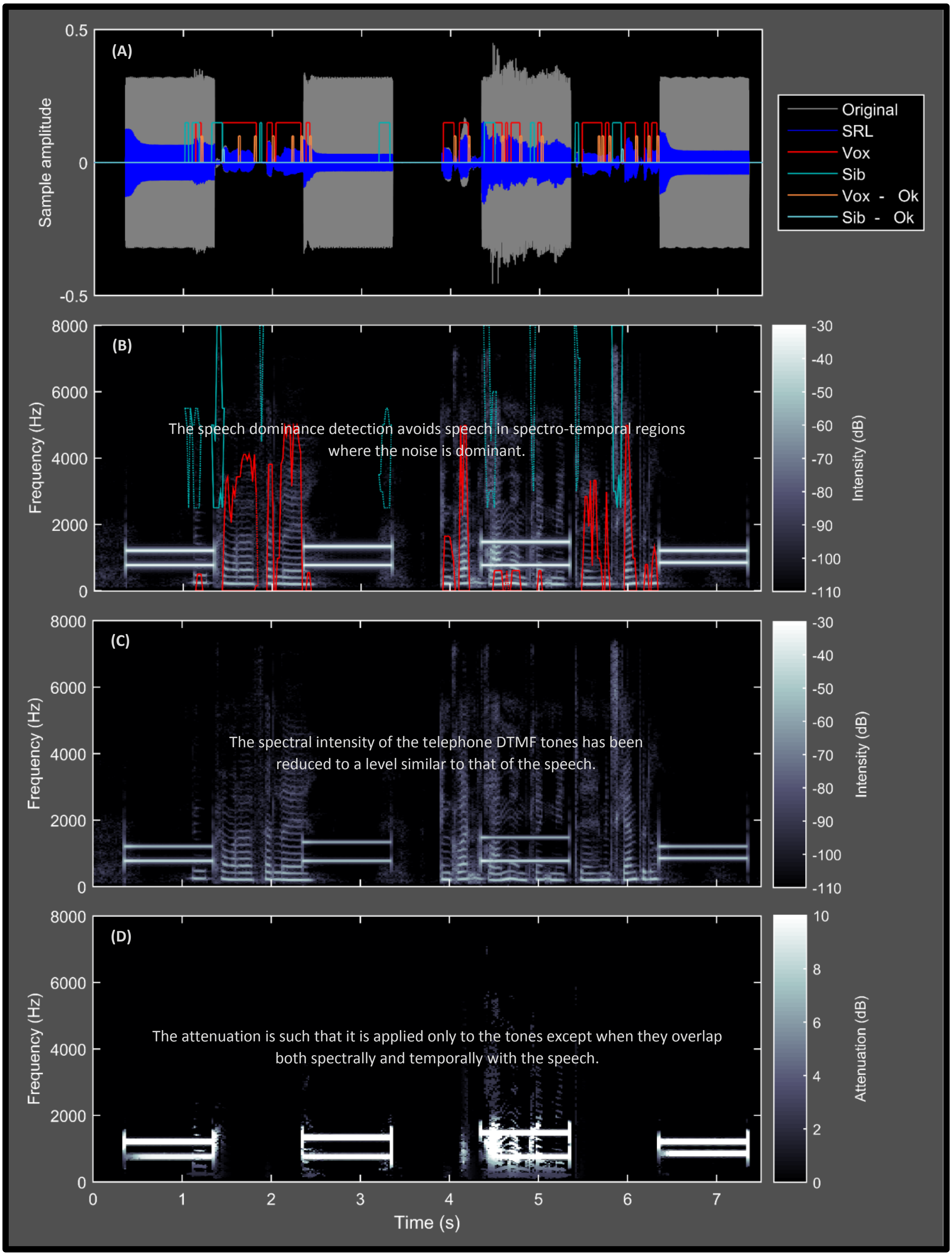


Figure 8-5. (A) *Waveforms (original – grey, SRL-processed - blue) of speech and series of telephone DTMF tones overlayed with SDD signals,* ***(B)*** *spectrogram of the original signal overlayed with SDD regions,* ***(C)*** *spectrogram of SRL-processed signal and* ***(D)*** *spectrogram of attenuation.*

8.4.1 Performance with multiple languages

The scheme was tested using speech material in multiple languages to check that the speech detector, which had been developed using material in English, would perform equally well when used with other languages. The test material came as an annexe to the International Telecommunications Union Recommendation ITU-T P.501 Test signal for use in telephonometry.[172] The languages were English (American accent), English (British accent), Chinese (Mandarin), Dutch, Finnish, French, German, Italian, Japanese, Polish and Spanish. The performance for five of these languages is shown in Figure 8-6 and Figure 8-7. Testing sequences were constructed from this speech material interspersed with the bursts of white noise.

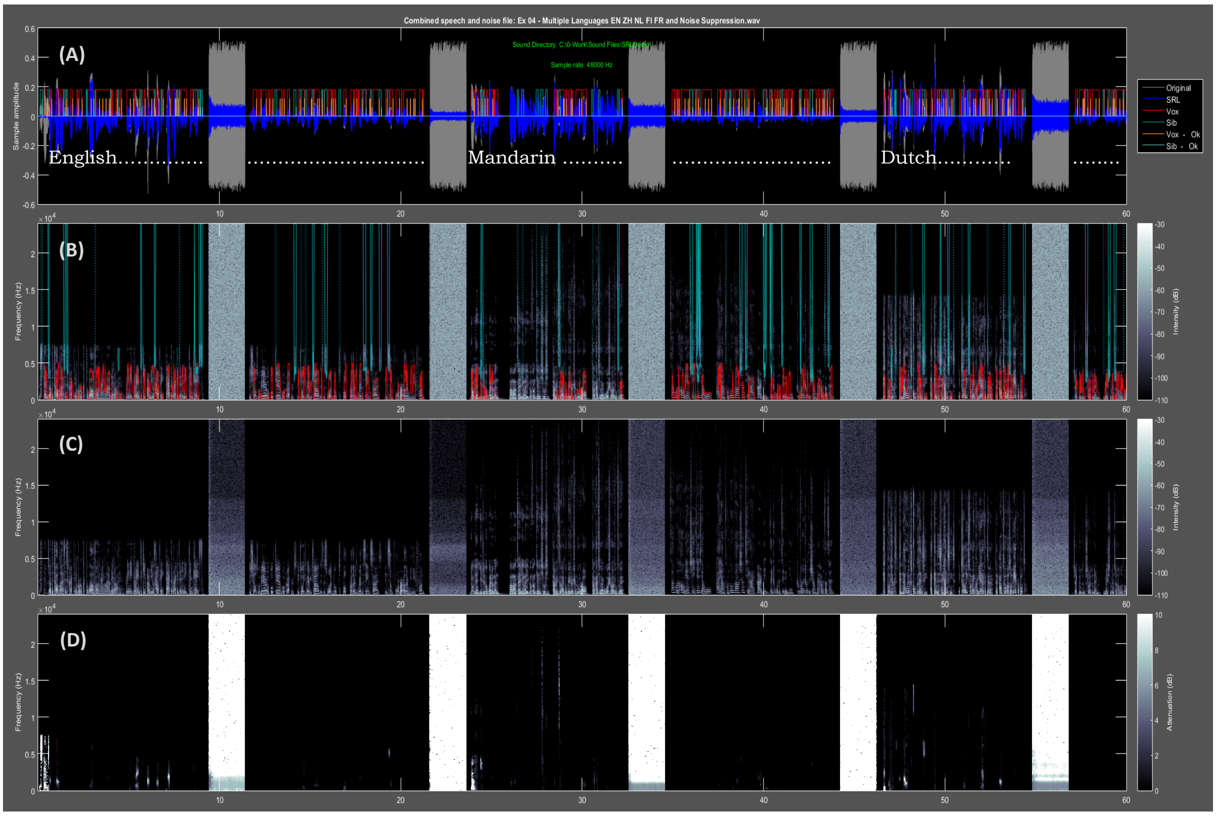


Figure 8-6. *Performance with multiple languages.* ***(A)*** *Waveforms (original – grey, SRL-processed - blue) of speech in multiple languages of varying level interspersed with bursts of white noise overlayed with SDD signals,* ***(B)*** *spectrogram of the original signal overlayed with SDD regions,* ***(C)*** *spectrogram of SRL-processed signal and* ***(D)*** *spectrogram of attenuation produced by SRL.*

The speech comprised of both male and female speech from each language. For each language the speech was presented at one level prior to the first noise burst, and then different speech from that language was presented at a 10 dB reduced level prior to the second noise burst in order to check that the limiting of the higher level noise burst was following the varying speech level. This was confirmed by observing the preservation of the speech and the attenuation of the noise that followed speech to a level similar to that of the preceding speech.

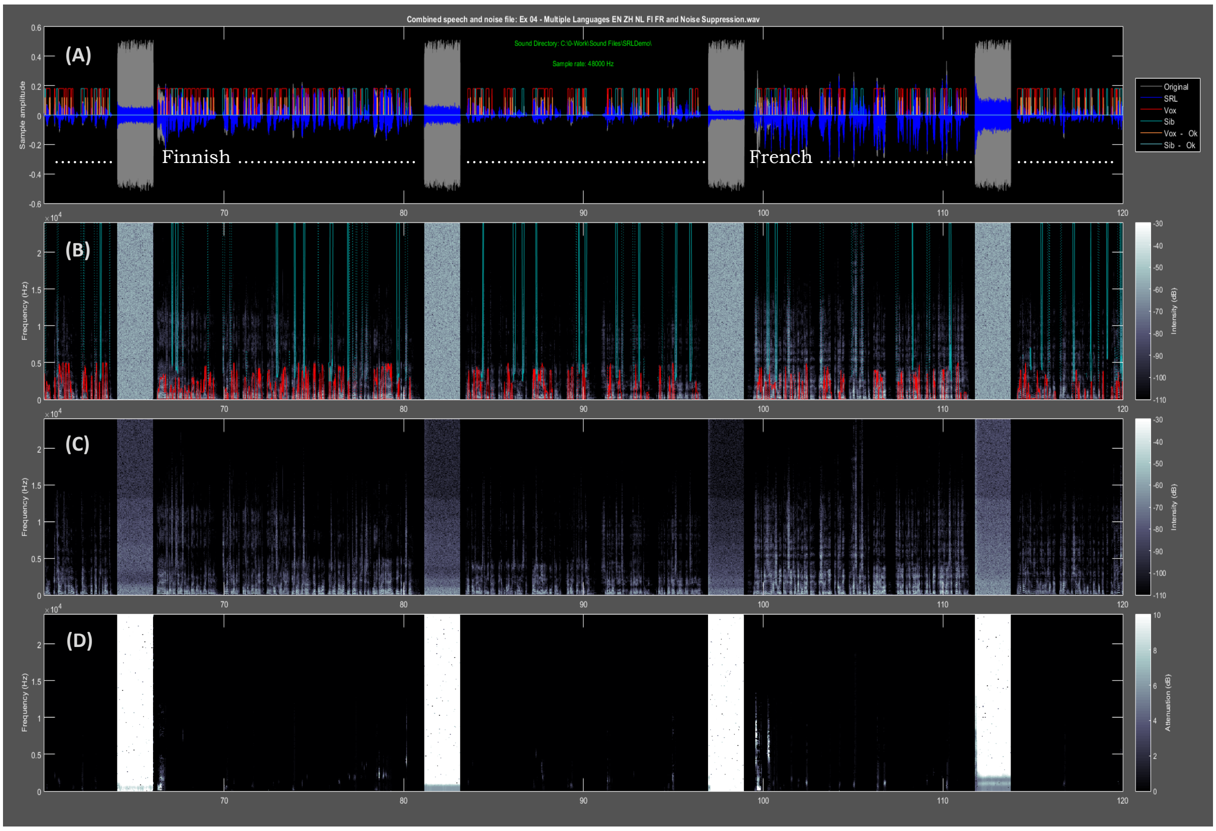


Figure 8-7. *Performance with multiple languages continued.* ***(A)*** *Waveforms (original – grey, SRL-processed - blue) of speech in multiple languages of varying level interspersed with bursts of white noise overlayed with SDD signals,* ***(B)*** *spectrogram of the original signal overlayed with SDD regions,* ***(C)*** *spectrogram of SRL-processed signal and* ***(D)*** *spectrogram of attenuation produced by SRL.*

For all the language material provided with the ITU-T P.501 the SRL MKII performance was equal to its performance when tested with English. As can be seen in the above two figures, the bursts of white noise are attenuated to approximately the level of the preceding speech and other than on a few occasions, mainly just following an increase in the speech level, the speech was not attenuated.

8.5 Gain Calculation

The overall objective of SRL is to control non-speech sounds that have an estimated loudness and/or low-level neurophysiological stimulation level exceeding that of recent speech by modifying the signal from which the sounds are reproduced. The gain calculator produces and compares estimates of the current signal loudness and power with estimates of the past speech loudness and power which it has derived from the analysed signal in the I frequency bins, Signal (i), in order to determine the frequency-specific gains, Gain (i), needed to reduce any excess estimated loudness and power in the current signal compared with the past speech signal.

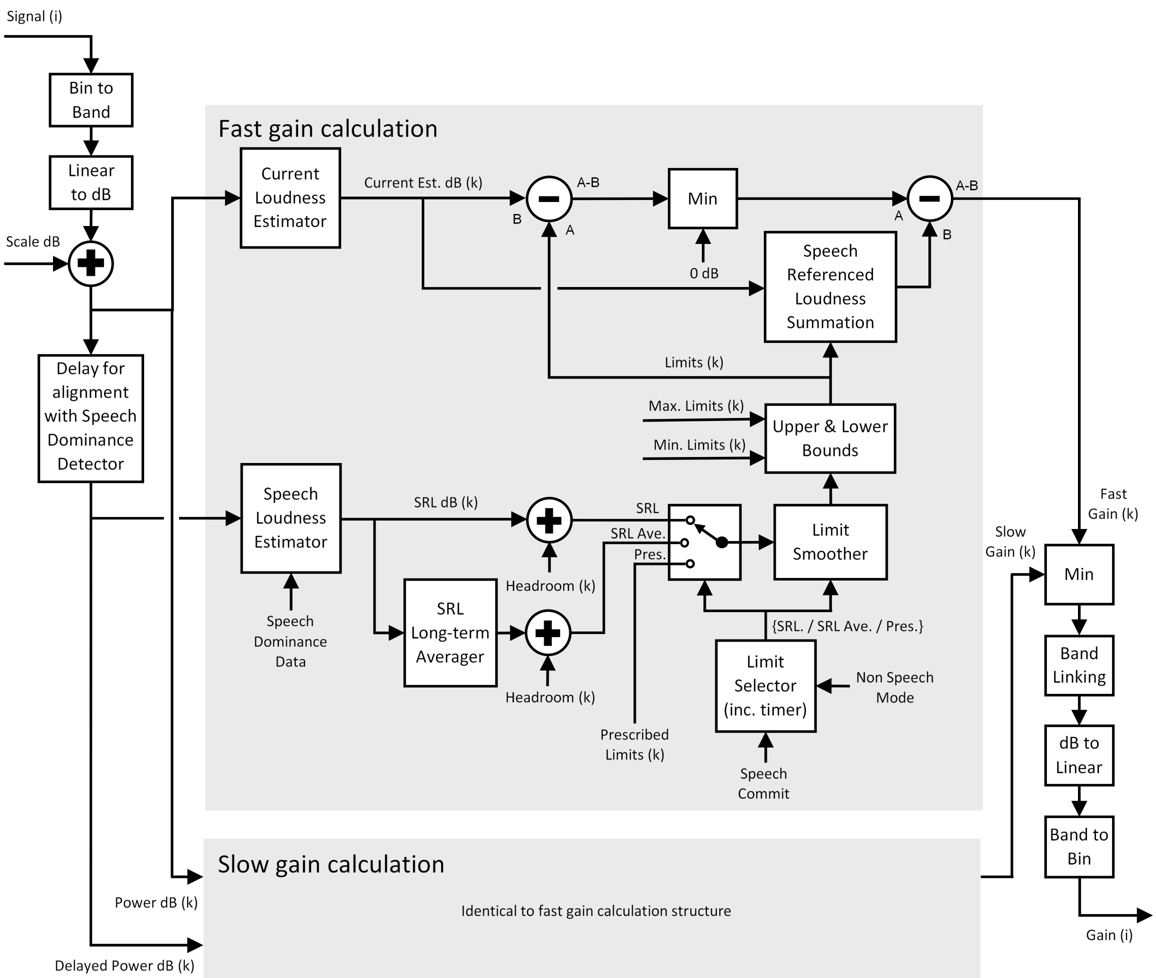


Figure 8-8. *The SRL MKII gain calculator.*

Figure 8-8 is a schematic diagram of the SRL MKII gain calculator. It comprises a fast gain calculation and a slow gain calculation in parallel. These are aimed at controlling the low-level neurophysiological response (i.e. somatic response) and loudness perception respectively. The fast gain calculation is identical to the slow gain calculation in structure, however, its parameters differ. In particular, its fast attack time constant is considerably faster than typical loudness integration rates. Although it is referred to as a loudness calculation, it is a multiband, peak-power calculation. These two calculators are preceded by a common input signal process comprising:

- band forming,
- conversion to decibels,
- scaling, and
- delaying (for the speech loudness estimators only),

and followed by a process of:

- combining the fast and slow gains they produce,
- linking the individual band gains, and
- converting the combined gain to the appropriate format for the adaptive modifier.

The gain calculator receives the power signal, Signal (i), for the I frequency bins from the analyser shown in Figure 8-2. It combines the power in these frequency bins to form the K frequency bands (bin to band). A number of band scales are supported by the SRL MKII scheme, including Bark,[45] ERB,[46] third-octave[132] and linear (with 250 Hz bandwidth for an analysis window length of 4 ms). The band powers are converted to dB (linear to dB) and an offset (scale dB) is added. This offset enables the power to be in dB SPL if required (for applications such as hearing aids and level-dependent hearing protectors). It may also be set so the power, in dB, is relative to digital saturation (for applications such as processing sound files on a computer). The power dB (k) values are delayed by approximately 40 ms (delay for alignment with speech dominance detector) to produce the delayed power dB (k) values for the speech loudness estimator that are time-aligned with the decisions from the speech dominance detector. These powers are supplied to the fast and slow gain calculators described in Section 8.5.1.

The gains from the fast and slow gain calculators (fast gain (k) and slow gain (k)) are combined using a minimum operation (min) such that the combined gain is the minimum gain produced by either. These are applied to the band linking function (band linking) that restricts the maximum gain of all the bands to be within X dB of the minimum gain of any band. The value of X (band-linking: maximum inter-band difference) was set to 30 dB for the MKII evaluation documented in Chapter 9. The resulting gains in dB are converted to linear scaling factors (dB to linear) and then converted to bin gains (Gains (i)) for the I bins required by the adaptive modifier (band to bin).

8.5.1 Fast and slow gain calculators

The fast and slow gain calculators are identical in structure, differing only in their parameters. The following description applies to both calculators.

The calculator receives Power dB (k) and the Delayed Power dB (k) which are supplied to the current loudness estimator and the speech loudness estimator respectively. These estimators produce estimates of the current loudness (Current Est. dB (k)) and the speech loudness (SRL dB (k)) respectively. Section 8.5.2 gives a detailed description of the operation of the current loudness estimator and those aspects of the speech loudness estimator that are shared with the current loudness estimator. Section 8.5.3 provides a detailed description of how this complex speech loudness estimator operates to produce the speech reference levels.

The following 6 subsections, 8.5.1.1 to 8.5.1.6 describe the remaining operations of the calculator.

8.5.1.1 Speech reference level – long-term averager

In applications that involve processing constantly changing speech levels, as occurs with telephone headsets when users take many calls, it is desirable to have the speech reference levels converge to the average of the past speech reference levels when speech is absent for some time. To create this average, the speech reference levels produced by the speech loudness estimator are applied to a long-term averager. The long-term averager is shown in Figure 8-9.

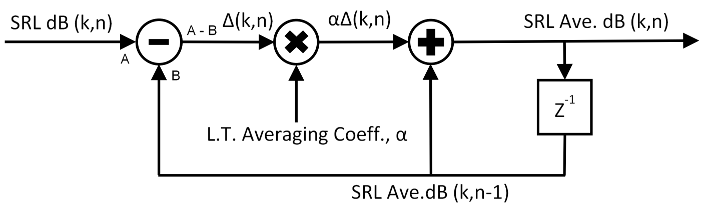


Figure 8-9. *Speech reference level long-term averager.*

The operation of this averager is given by Equation 8-1.

$$SRL\ Ave\ dB(k,n) = \alpha \times \left(SRL\ dB(k,n) - SRL\ Ave\ \ dB(k,n-1)\right) + SRL\ Ave\ dB(k,n-1) \qquad \textbf{(8-1)}$$

$$\begin{aligned} where\colon\ & \alpha\ is\ the\ coefficient\ described\ by\ equation\ below \\ & k\ is\ the\ band\ number \\ & n\ is\ sample\ number \end{aligned}$$

The relationship between the time constant and its coefficient, α, is given by Equation 8-2.

$$\alpha = \ min \begin{Bmatrix} \frac{2.5}{Fs_{analysis} \times TC} \\ 1 \end{Bmatrix} \qquad \textbf{(8-2)}$$

$$\begin{aligned} where\colon\quad & TC\ is\ the\ time\ constant \\ & Fs_{analysis}\ \ is\ the\ sampling\ rate\ of\ the\ analysis \end{aligned}$$

The time constant is typically set to average over several minutes. This, however, was not relevant during the evaluation of SRL MKII reported in Chapter 9, as the "In the absence of speech" parameter was set to "Maintain last Speech Reference Limits".

8.5.1.2 Headroom

In order for new speech of a slightly higher loudness level than the existing speech reference levels to be passed without attenuation, prior to the speech reference level being updated to include it, an allowance in terms of a headroom factor is provided. This has been previously discussed in Chapter 5 in relation to the theory behind the SRL scheme, and in Chapter 6, in relation to the SRL MKI scheme. In the SRL MKII scheme, headroom factors, headroom (k) (in dB), have been added to both the SRL dB (k) and SRL Ave. dB (k) values. In the evaluation of SRL MKII, as described in Chapter 9, the headroom (k) factors were all set to 3 dB. This value allowed the new speech and noise to be slightly louder than the past speech, reducing the amount of attenuation applied to new speech of a slightly greater loudness and allowing noise, such as an alarm, to be slightly louder than the past speech. However, due to the loudness summation reduction, described in Section 8.5.1.5, high-level broadband noises, e.g. white noise bursts, were suppressed to a degree that appears to counteract the headroom for this noise type, at least this was the perceptual result in the evaluation of SRL MKII reported in Chapter 9.

8.5.1.3 Limit selection and smoothing

There may be considerable time during which speech is not regularly presented to the SRL scheme. For example, this may occur when SRL is included in a hearing aid that is used by a person most of the day. In such situations, there are three options:

1. The scheme can maintain the last speech reference levels, offering added protection but possibly too much limiting for some situations such as listening to music at a high sound level.
2. The scheme can drift to the average of the past speech reference levels. This is very appropriate for speech of possibly unpredictable level, such as is experienced when receiving multiple telephone calls, but again not so appropriate for situations such as listening to music at a high sound level.
3. The scheme can drift to prescribed or other fixed limit levels. These may be standard limit levels prescribed using a formula, such as the NAL-SSPL,[113] or other fixed levels set by a clinician.

The SRL MKII scheme has been created with these three options available. The time before a drift to the average speech reference levels, or prescribed limits, in the absence of speech, is a parameter that may be set. For example, in hearing aid application, this can be set to one minute. Under this arrangement, the speech referenced limiting automatically becomes active whenever dominant speech is present and for up to a minute after it becomes absent. Following this, the aid smoothly transitions to the prescribed or other pre-set fixed limits. This can work in the same manner for hearing protector users and (tele)communications headset users, although, the default, in the case of headset users will most likely be the average of the past speech reference values, rather than prescribed limits, and the time before drifting shorter. Figure 8-8 shows a switch to select between the three sets of limits. This is operated by a timer programmed for a given drift period for modes other than maintaining the speech reference limits in the absence of speech. When switching from one set of limits to another, it is important that the transition sounds smooth. Figure 8-10 shows a schematic of the limit smoothing within the scheme. The speed at which the speech reference levels are adopted when dominant speech becomes present is in the order of a tenth of a second. The speed at which the drift to the average speech reference level occurs in headset applications may be 10 seconds and for the drift to prescribed limits in hearing aids and level-dependent hearing protectors may be around a minute.

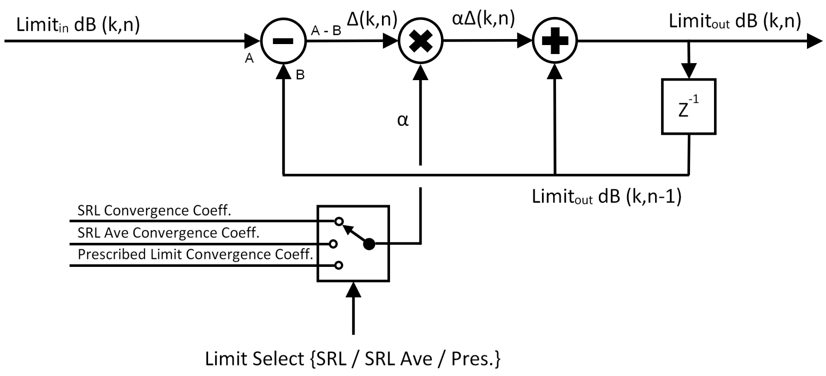


Figure 8-10. *Limit smoother.*

The operation of this smoother is given by Equation 8-3 and the coefficients it uses by Equation 8-4.

$$limit_{out}dB(k,n) = \alpha \times \left(limit_{in}dB(k,n) - limit_{out}dB(k,n-1)\right) + limit_{out}dB(k,n-1) \qquad \textbf{(8-3)}$$

where:

α *is the coefficient described by equation below*

k *is the band number*

n *is sample number*

and

$$\alpha = \begin{cases} SRL\ Convergence\ Coeff. \\ SRL\ Ave\ Convergence\ Coeff. \\ Prescribed\ Limit\ Convergence\ Coeff. \end{cases} \quad \textit{According to the limit mode} \qquad \textbf{(8-4)}$$

and each $\alpha = \min\left\{\begin{matrix} \frac{2.5}{Fs_{analysis} \times TC} \\ 1 \end{matrix}\right\}$ *with a TC for the relevant limit selection*

and $Fs_{analysis}$ *is the sampling rate of the analysis*

8.5.1.4 Maximum and minimum limits

In order to prevent extreme limits being adopted in unusual circumstances the variable limits produced by the limit smoother are bounded within a specified range of maximum and minimum values (Max. Limits (k) and Min. Limits (k)) by a bounding operation (Upper & Lower Bounds). The maximums prevent the adoption of high-level limits in cases where high-level noise has mistakenly been sampled as speech. The minimums prevent the adoption of low-level limits in cases where speech at a low level has been detected (e.g. speech at some distance from the listener). The overall setting of maximum limits for the evaluation of SRL MKII, documented in Chapter 9, was done individually by each subject with the adaptive (i.e. SRL limits) disabled. The frequency balance of these was based on the National Acoustic Laboratories' saturated sound pressure level (NAL-SSPL) prescribed real-ear saturated response (RESR).[113,173] The minimum limits used in the evaluation were set to 28 dB below those prescribed by the NAL-SSPL.

8.5.1.5 Speech referenced loudness summation

As discussed in Chapter 2, it is well known that the sensation of loudness increases with the bandwidth of the stimulus. This effect is known as loudness summation, as the summation of specific loudness across frequency produces total loudness. There, however, remains considerable debate about the degree to which this occurs and how it should be modelled (see Chapter 2 for a discussion of this). Hearing aids and other audio processing devices, to my knowledge, currently do not directly address the effect of loudness summation for signals with bandwidths broader than that of speech. The prescriptions for hearing-aid fitting, such as the National Acoustic Laboratories' non-linear amplification prescription (NAL-NL2),[110] have taken into consideration the amount of loudness summation that occurs for speech when determining the gain to prescribe. Assuming a multi-band device has been set for overall loudness normalisation of speech and contains no additional compensation for loudness summation, it would, in theory, produce greater than normal loudness for sounds with bandwidths greater than speech for those listeners with some functioning loudness summation, and for which the hearing aid is capable of providing audibility outside the speech bandwidth.

The SRL MKII addresses the issue of loudness summation resulting from sounds with a bandwidth exceeding that of speech, using a speech referenced loudness summation estimator and reducing the level of sound in response to this estimated excess. Figure 8-12 shows a schematic diagram of the speech referenced loudness summation estimator. The first part of this estimation process is counting the number of critical bands (Bark or ERB) in which the current loudness estimate (Current Est. dB(k,n)) exceeds the loudness summation adjusted speech referenced level (Limit dB (k, n)). Figure 8-11 shows a flow chart of this counting process. This process accommodates a band structure that can be different to the critical band (Bark or ERB) structure employed, such as a linear band spacing of 250 Hz.

From this count is subtracted the number of critical bands with excess loudness that may simultaneously result from speech. This is shown by the second subtraction from the left-hand side of Figure 8-12. If the resulting value is zero or lower the signal produces no more loudness summation than typical speech and therefore does not require further attenuation other than that provided on a band-specific basis by the SRL scheme. If, however, the resulting value is greater than zero then there is additional loudness summation from the signal compared to that of speech which needs to be attenuated.

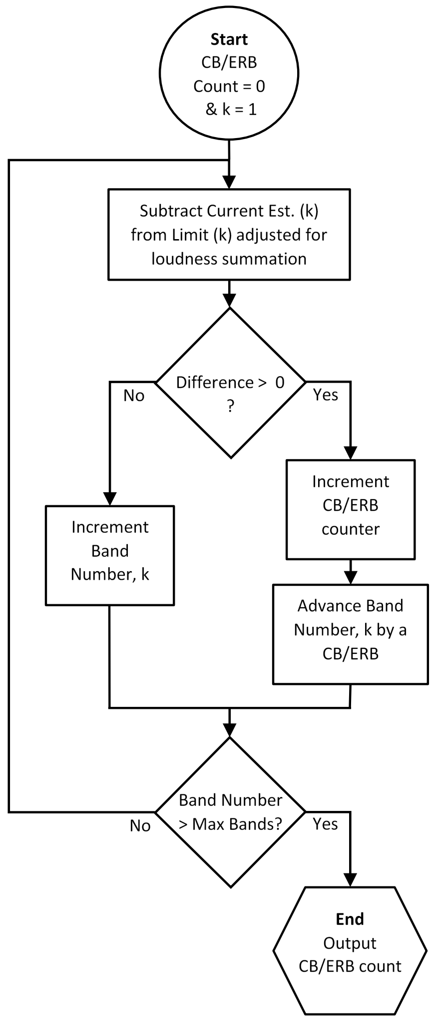


Figure 8-11. *Flow chart of speech referenced loudness summation estimation.*

The result is multiplied by a maximum loudness summation factor in dB to produce an appropriate compensatory attenuation in dB. This attenuation is applied to a smoothing function comprising all the operations to the right-hand side of the application of the maximum loudness summation factor. This smoothing function has an instantaneous attack time and a release time set by the loudness summation release coefficient, α.

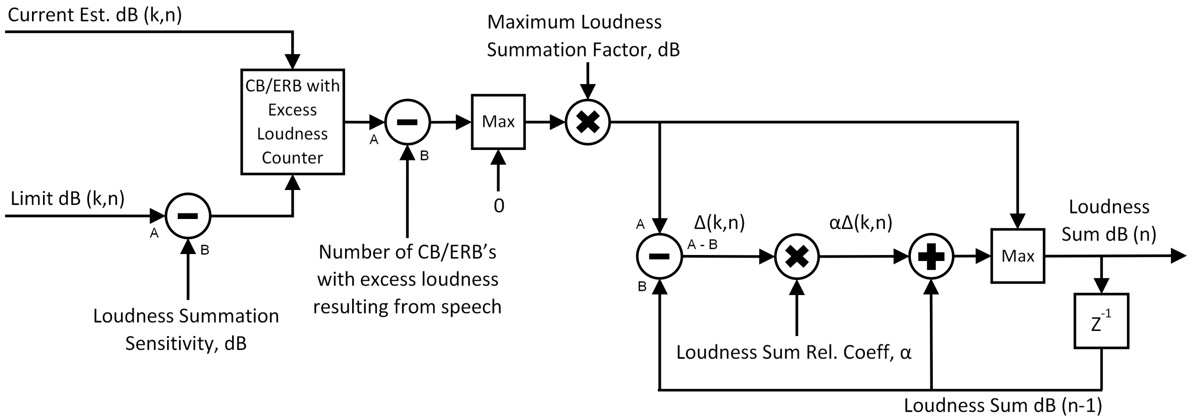


Figure 8-12. *Speech referenced loudness summation estimator.*

For the evaluation of SRL MKII, documented in Chapter 9 the following loudness summation values were used:

- Loudness summation sensitivity (dB): 0
- Loudness summation bandwidth: Bark scale
- Number of CB/ERB's with excess loudness resulting from speech (Fast): 4
- Number of CB/ERB's with excess loudness resulting from speech (Slow): 1
- Maximum loudness summation factor (fast): 10 dB / total number CBs
- Maximum loudness summation factor (slow): 8 dB / total number CBs
- Loudness summation release time constant: 125 ms

Correction for loudness summation is applied to all bands equally as shown in Figure 8-8 and described the following section on gain generation.

8.5.1.6 Gain generation

The final gain produced by each of the fast and slow gain calculators for each of the bands is the band-specific reduction gain (i.e. the limit less the current level) less the loudness summation, this is shown in Figure 8-8 and expressed in Equation 8-5.

$$Gain\ dB(k,n) = min\begin{Bmatrix} Limit\ dB(k,n) - Current\ Est.dB(k,n) \\ 0 \end{Bmatrix} - Loudness\ Sum\ dB(n) \qquad \textbf{(8-5)}$$

where:

k *is the band number*

n *is sample number*

The gains resulting for both the fast and slow gain calculations are combined by taking the minimum of the two as described at the beginning of Section 8.5.

8.5.2 Loudness estimators

The aim of the SRL scheme was to preserve the fidelity of speech when the signal-to-noise ratio was good and suppress noise when it exceeded the estimated speech loudness. It was therefore important that the loudness estimate of speech, when speech was at a good signal-to-noise ratio, at least equalled the loudness estimate of the current signal, excluding the time delay in estimating the speech loudness, to avoid suppression of the speech. It was also important that sounds that were perceptually louder than speech were estimated as being louder so that they were suppressed. Clearly, in order to achieve this, the same method had to be applied to estimating both the current loudness and the past speech loudness. Hence the current and speech loudness estimators are the same and use the same parameters.

The envelope detectors used in SRL MKI to estimate the frequency-specific loudness of the signal operated on the power of the signal. An alternative approach, used in the SRL MKII scheme, is an envelope detector that operates on the logarithm of the power signal. Envelope detectors that track the logarithm of the signal's power in decibels, or other logarithmic units, rather than the signal's pressure or power signal, are used in some hearing aid algorithms. A general approach to this method was described by Schaub[152] and high-order variants of this method were described by Schaub and Leber.[174] As these envelope detectors operate on the logarithm of the power signal they do not have to deal with the large dynamic signal range faced by an envelope detector that operates on the absolute value of the pressure or the power of the signal. They are therefore less susceptible to quantisation noise and are more robust when used with low-precision, fixed-point processors in devices such as hearing aids and level-dependent hearing protectors.

The frequency-specific loudness estimator developed for SRL MKII is a first-order tracker of the logarithm of the signal's power. This is shown in Figure 8-13. It uses the scaled difference between the input power in dB (Power dB (k,n)) and its previous estimate (Est. dB (k,n-1)) to update its current estimate (Est. dB (k,n)). It switches the scaling coefficient (α) to provide separate attack and release rates. Like the conventional envelope detector described in Chapter 6, it uses the difference between the input power and its previous estimate to determine whether it is in attack or release mode, selecting an attack coefficient if the input exceeds the previous estimate and a release coefficient if it does not.

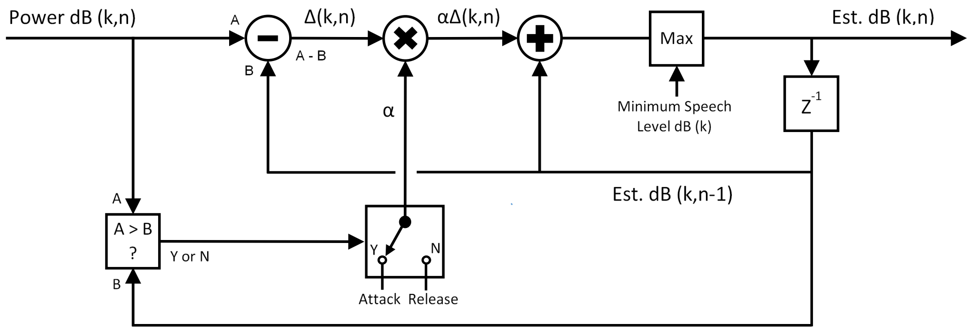


Figure 8-13. *1st order envelope detector in dB.*

Also incorporated into the frequency-specific loudness estimator is a minimum estimation value. This is included so that, when the power at the specific frequency is low, the estimator does not converge to a low value as this would delay recovery after the appearance of a high-level signal. This minimum value is set to the expected frequency-specific minimum loudness level of speech. The reference for this minimum may be the absolute input SPL if the level is referenced either to an input or output level in SPL. In the case of processing audio recordings or when used in telecommunications systems, where there is no defined relationship to acoustic levels, this minimum level may be referenced to digital saturation (e.g. to band levels corresponding to speech with, for example, a broadband level of -50 dB relative to digital saturation). The operation of the envelope detector is given by Equation 8-6.

$$Est.dB(k,n) = Max\begin{cases} \alpha \times \left(Power\ dB(k,n) - Est.\ dB(k,n-1)\right) + Est.dB(k,n-1) \\ Minimum\ Speech\ Level\ dB(k) \end{cases} \quad \textbf{(8-6)}$$

$$where\ \ \alpha = \begin{cases} attack\ \alpha & for\ Power\ dB\ (k,n) > Est.dB(k,n-1) \\ release\ \alpha & otherwise \end{cases}$$

and

k *is the band number*

n *is sample number*

The relationship between the time constants and their coefficients is given by Equation 8-2.

For the envelope detector to be stable the coefficient, α, cannot exceed 1. For the analysis sampling rate of 1,000 Hz this corresponds to a minimum time constant of 2.5 ms. However, this minimum time constant may become longer or shorter depending on the analysis window length. The total window length in this implementation is normally set to 4 ms. However, due to the use of a Hann window weighting, the effective window length determined by the 6 dB time width is only 2 ms. Therefore, the minimum time constant resulting from using an α coefficient of 1 is effectively 2 ms using the 6 dB window definition.

The parameters that distinguish the fast and slow estimators are described in more detail in the following two sub sections.

8.5.2.1 Fast estimator

The attack time constant of the fast estimator needs to be faster than the acoustic startle integration time constant, i.e. 3 ms, as discussed in Chapter 3. The minimum attack time constant, discussed in the preceding section, of 2 ms, is employed. The estimate is used to directly control the suppression of peaks that exceed the peaks of speech, and because of this, the release rate needs to be slow enough not to cause distortion of speech or excessive amplification of noise that follows a peak. For example, the release rate cannot be so fast that the reverberating sound in a typical room from an impact noise, such as a door slam or hammer hit, fails to be suppressed along with the initial impact sound. Yet it should be fast enough to release its suppression before it reduces the level of speech that follows. Unfortunately, there is no one rate that will satisfy both requirements. For the evaluation of the SRL MKII scheme, as documented in Chapter 9, a release rate of 250 ms was used for the fast estimator.

8.5.2.2 Slow estimator

The attack time for the slow estimator needs to be at least as fast as the loudness integration time constants discussed in Chapter 2. This is to ensure the signal is controlled at least as quickly as the listener integrates the loudness. However, the attack time should not be so fast that the loudness estimate of speech becomes unrealistically high as this would result in noises being limited to speech reference levels that are not consistent with the perceived speech loudness. As discussed in Chapter 2, the loudness integration rate for the average human is in the order of 100 ms. Some loudness estimators, such as the short-term loudness estimator of Glasberg and Moore,[48] use fast integration rates, while others, such as Chalupper and Fastl,[49] use slower rates. The SRL MKI scheme used an attack rate of 80 ms in the evaluation, documented in Chapter 7. For the evaluation of the SRL MKII scheme, documented in Chapter 9, an attack rate of 100 ms was used for the slow estimator.

The loudness perception of pulsed stimuli as a function of pulse density, documented by Zwicker and Fastl[23] (see Chapter 2), showed a similar loudness integration rate to that for signals that vary in duration. Based on this data, a release rate equal to the attack rate would be appropriate. However, this measure is based on simple stimuli presented under laboratory conditions rather than more complex real world sounds and does not consider the persistence of loudness within the memory of a listener. To address this effect, Glasberg and Moore, proposed a long-term loudness estimator in addition to a short-term loudness estimator.[48] Rennies et al.[60,61] found that this long-term loudness estimator provided a better match to subjective loudness data than the short-term estimator of Glasberg and Moore or Chalupper and Fastl (see Chapter 2). This would indicate that a longer release time would be more appropriate for estimates of loudness where the memory of loudness matters, such as when estimating the loudness of running speech. As will become evident later in this chapter, this loudness estimate is used directly to control the short-term loudness of sounds that exceed the loudness of past speech. As a result, the release rate needs to be slow enough not to introduce audible distortion into the signal. Many studies have shown that speech quality is least degraded with a slow release rate.[175-178] For the evaluation documented in Chapter 9, a release rate of 500 ms was used for the slow estimator.

8.5.3 Speech loudness estimator

The speech loudness estimator in SRL MKII is a considerably more sophisticated process than that used in SRL MKI. It aims to create sets of speech reference levels that are representative of the speech even when the speech at some frequencies is embedded in noise. The estimator uses the information provided by the speech dominance detector to determine when, and over what frequencies, the power of voicing or sibilance should be sampled in order to form estimates of voicing and sibilance loudness. It extrapolates and interpolates missing data from its estimates of voicing and sibilance loudness to form sets of speech reference levels.

The scheme forms two sets of speech loudness estimates: one using fast time constants and one using slow time constants. To be able to produce the same loudness estimates as the current loudness estimator for speech alone, a fundamental requirement of the SRL method, the methods and parameters of temporal integration used for the speech estimates are the same as those for the current estimates. However, these fast and slow estimates are formed separately from separate estimates for voicing and sibilance.

Figure 8-14 is an overall schematic diagram of one of the two speech loudness estimators. This could be either the fast or slow speech estimator, as they are identical in structure and differ only in their parameters.

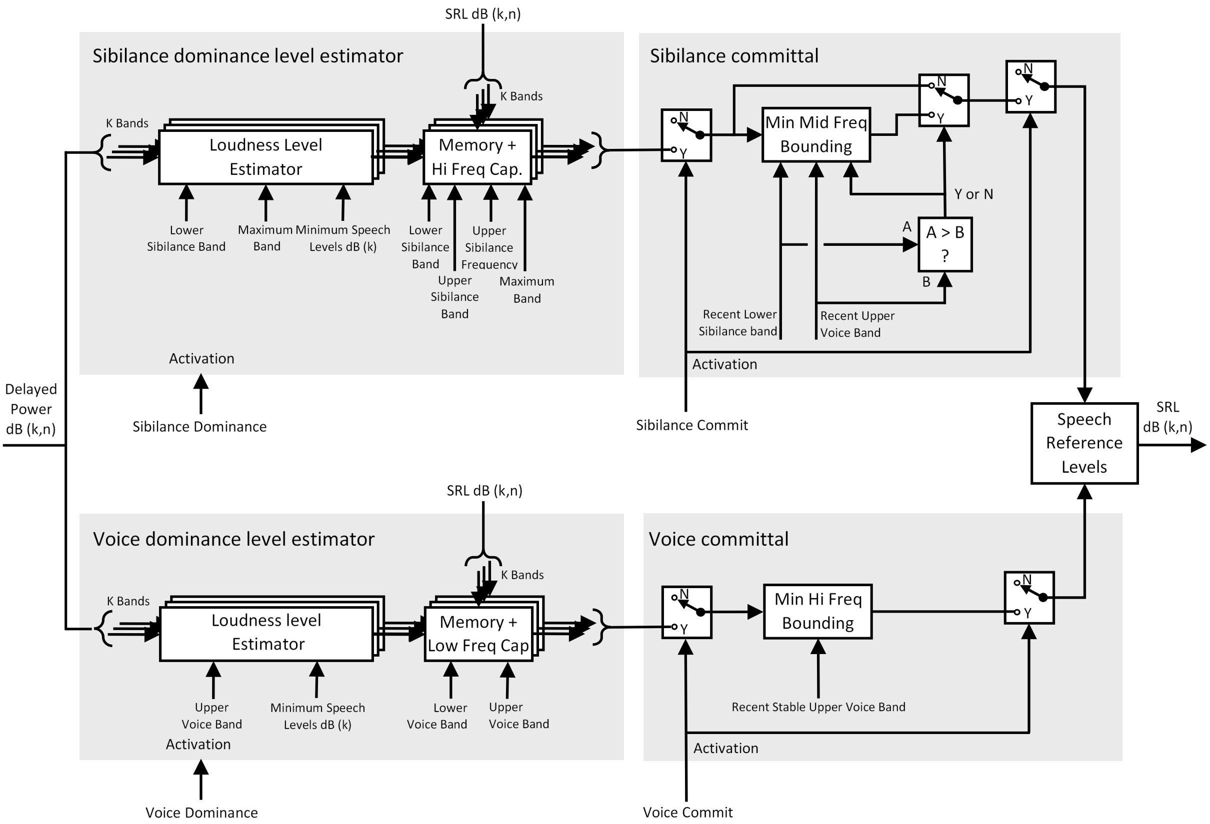


Figure 8-14. *Speech loudness estimator.*

The estimator receives a set of power estimates in dB, comprising one estimate for each of the K bands per analysis (n). These are labelled as Delayed Power dB (k,n). The estimates have been delayed so that they are time-aligned with the control signals from the speech dominance detector. The estimator produces a set of speech reference levels in dB, comprising one reference level for each of the K bands per analysis (n). These are labelled as SRL dB (k,n).

The estimator includes a means of updating the speech reference levels in response to dominant voicing as well as a means of updating the speech reference levels in response to dominant sibilance. Either of these functions can update the speech reference levels through their respective committal processes (labelled Voice Committal and Sibilance Committal). The committal of new data to the speech reference levels occurs when the speech dominance detector determines that the signal used to update the recent estimate of either voicing or sibilance was dominated by voicing or sibilance respectively within the frequency range it defines.

8.5.3.1 Updating the speech level estimates from voiced speech

The dominant energy of speech in the lower frequencies comes from voicing.[179] Every 20 ms, the speech dominance detector provides five signals about the voicing dominance within the signal as shown in Table 8-4.

Number	Speech sampling control signal
1	Voice dominance
2	Lower voice frequency
3	Upper voice frequency
4	Voice commit
5	Voice commit maximum upper frequency

Table 8-4. *Voicing sampling control signals.*

In brief, when the voice dominance signal is true, voicing is dominant in the frequency range from the lower voice frequency to the upper voice frequency. When the voice commit signal is true, then the signal from the time that the voice dominance signal became true is considered to be dominated by voicing within the time-varying frequency-range from the lower voice frequency to the upper voice frequency but not at any frequency greater than the voice commit maximum upper frequency.

The voicing loudness level is estimated by the loudness level estimators when the voicing is dominant (i.e. the voice dominance signal is true). The voice loudness level estimators are the same as the current loudness estimators, except that they are only active when voicing is dominant. The resulting voicing loudness level estimates are applied to a memory simulator (labelled as Memory + Low Freq Cap in Figure 8-14) to update the memory of speech loudness.

On starting the voice dominance level estimation, the estimates of past speech, (i.e. SRL dB values) are loaded into the voicing estimator's memory. The reason for doing this, rather than simply keeping the previous calculated voicing estimates, is that, during the intervening time between two episodes of voiced speech the speech reference levels may have been updated by the sibilance level estimator and these updates need to be included. There is, of course, a substantial region of cross-over between the frequencies of voicing and sibilance,[180] and the estimates in the cross-over bands are therefore a function of both.

8.5.3.2 Frequency range of voicing updates

The speech dominance detector provides estimates of the lower voice frequency and upper voice frequency. To accommodate different band structures, based on frequency scales such as ERB,[46] and Bark,[45] within the gain calculator, these frequency estimates are provided to the speech loudness estimator in units of Hertz and need to be converted to fit the band structure used. The frequencies correspond to different bands when different frequency scales are used. For example, the bands are different when using the ERB scale compared with the Bark scale. To ensure that there is no dominant noise present within the edge bands of the specified voicing region the edge bands are discarded. Therefore, the band corresponding to the lower voice frequency is ignored, and the voicing estimate starts at the next higher band, the Lower Voice Band (LVB). Likewise, the band corresponding to the upper voice frequency is ignored and the voicing estimate ends at the next lower band, the Upper Voice Band (UVB).

8.5.3.3 Voicing updates

As long as the voice dominance signal is true, the estimates of the voicing loudness level in the bands up to the Upper Voice Band are updated at the analysis rate (i.e. every 1 ms) and the voicing band range is updated at the speech dominance detection rate (i.e. every 20 ms).

8.5.3.4 Updating the memory of speech loudness with new voicing

During voicing dominance, the memory of the loudness of speech is updated in the memory block (labelled Memory + Low Freq. Cap in Figure 8-14). A detailed schematic diagram of this is shown in Figure 8-15.

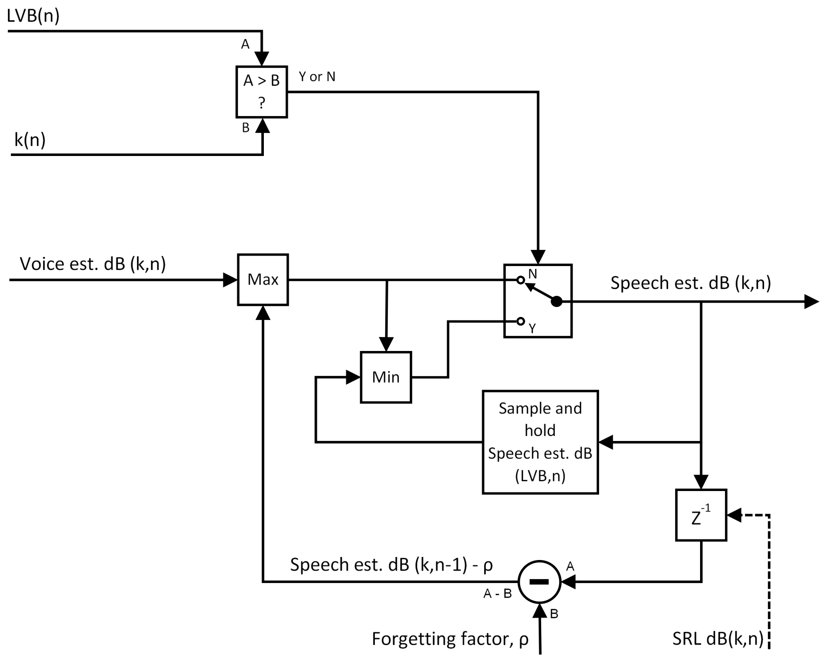


Figure 8-15. *Voicing memory and low frequency cap.*

At the beginning of a voiced segment, when voice dominance changes to true, the estimates of past speech reference levels (SRL dB (k,n)) are loaded into the voicing estimator's memory (dashed line into the unit delay in Figure 8-15). The memory simulator then updates its memory values for the voiced band range from the Lower Voice Band up to the Upper Voice Band (i.e. LVB(n) < k(n) < UVB(n)). The memory of speech is decreased by the forgetting factor, ρ, to produce new memory values (Speech est. dB (k,n-1) – ρ). Should a new estimate of the voicing level (Voice est. dB (k,n)) exceed the corresponding new memory of speech, then the new memory is immediately replaced with the new voicing level estimate. The operation of memory is given by Equation 8-7.

The 2 dB per second forgetting rate used in the SRL MKI evaluation, reported in Chapter 7, was found to be too slow when speech was present. A forgetting rate greater than 5 dB per second was found to better simulate the adaptation to new lower voicing levels. In the SRL MKII evaluation, described in Chapter 9, the forgetting rate is set to 12 dB per second.

$$Speech\ est.dB(k,n) = Max\begin{cases} Speech\ est.\ dB(k,n-1) - \rho \\ Voice\ est.\ dB(k,n) \end{cases} \tag{8-7}$$

$$for\ LVB(n)\ \leq k(n)\ \leq UVB(n)$$

$where:$ ρ $is\ the\ forgetting\ factor\ in\ dB\ per\ sample$
$LVB\ is\ the\ Lower\ Voice\ Band\ number$
$UVB\ is\ the\ Upper\ Voice\ Band\ number$
$k\ is\ the\ band\ number$
$n\ is\ sample\ number$

The relationship between the forgetting factor, ρ, and the forgetting rate is given by Equation 8-8.

$$\rho = \frac{forgetting\ rate\ in\ dB\ per\ sec}{Fs_{analysis}} \tag{8-8}$$

$where\ Fs_{analysis}\ is\ the\ sampling\ rate\ of\ the\ analysis$

The lowest voice band in the voice dominance range may not extend as low as the lowest frequency band. There are a few reasons why this may occur. First, the fundamental frequency of a speaker's glottal vibration can typically vary from 80 Hz to 700 Hz. If a child is speaking, then the fundamental is typically quite high, meaning there is no voicing energy in the low bands. Second, corruption of low-frequency speech by noise frequently occurs in everyday life due to sounds such as the high-level, low-frequency noise within a car.

The bands below the voice band range may contain low-level noise, low-level voiced speech, high-level noise or a combination of these. Updating the speech loudness estimates with low-level voiced speech will result in more accurate speech estimates. Updating the estimates with low-level noise will reduce these low-frequency speech estimates but only until speech at these frequencies reappears. As there is no speech at these frequencies in the meantime, the adverse effect is limited to the brief adaptation period during which new reference levels are produced. Updating with high-level noise, however, will corrupt the estimates and should be avoided. Therefore, the method of processing bands below the voicing band range is to update the speech level estimates with the voicing level estimates when voicing is dominant but cap them at the level of the voice estimate of the lowest voicing band. This ensures that high-level noise does not create estimates exceeding established voicing levels and controls the potential for low-frequency noise to create upwards spread of masking, which would adversely affect the intelligibility of low-frequency speech components.

The schematic in Figure 8-15 contains a switch that applies a cap equal to the speech estimate in the Lower Voice Band (Speech Est. dB (LVB, n)) to all bands below the Lower Voice Band. Equation 8-9 is an extension of Equation 8-7, which includes calculations for the bands below the Lower Voice Band.

$$Speech\ est.dB(k,n) =$$

$$\begin{bmatrix} Max\begin{Bmatrix} Speech\ est.\ dB(k,n-1) - \rho \\ Voice\ est.\ dB(k,n) \end{Bmatrix} for\ LVB(n) \le k(n) \le UVB(n) \\ Min\begin{Bmatrix} Max\begin{Bmatrix} Speech\ est.\ dB(k,n-1) - \rho \\ Voice\ est.\ dB(k,n) \end{Bmatrix} \\ Speech\ est.\ dB(LVB(n),n) \end{Bmatrix} for\ k(n) < LVB(n) \end{bmatrix} \tag{8-9}$$

$$\begin{aligned} where:\ & \rho\ is\ the\ forgetting\ factor\ in\ dB\ per\ sample \\ & LVB\ is\ the\ Lower\ Voice\ Band\ number \\ & UVB\ is\ the\ Upper\ Voice\ Band\ number \\ & k\ is\ the\ band\ number \\ & n\ is\ sample\ number \end{aligned}$$

8.5.3.5 Committing the dominant voicing estimates

It would, of course, be desirable to update the speech reference levels as the dominant speech was occurring, as was the case with the SRL MKI scheme (apart from a delay of approximately 40 ms). However, this would remove the capacity to change the outcome resulting from modified speech reference levels when the dominant signal was found not to be speech. Using this method, a high-level noise incorrectly identified as speech would corrupt the speech reference levels and would not be limited in level. This was a limitation of the SRL MKI scheme. Speech dominance detection is more accurate with increasing duration of the speech segment. This was addressed in the SRL MKII scheme by the speech dominance detector providing a voice commit control signal to the gain calculator when it is confident the signal, since the voice dominance became true, has been dominated by voicing within the voicing frequency range it specified. At the end of the voicing or after a given time (i.e. 200 ms), whichever comes first, the detector has sufficient confidence in the measures to make a decision as to whether the segment was voice-dominated and to set the voice commit signal status. The detector also determines the maximum frequency of uncorrupted voicing in the segment and sets a voice commit maximum upper frequency for the voicing segment.

If the voice commit signal becomes true, then the speech estimates from the voice dominance level estimator become the new speech reference levels following the application of a minimum bound on the high frequencies (Min Hi Freq Bounding) as shown in Figure 8-14. This bound is to ensure that, in the absence of sibilance updates, there are always reasonable minimum speech reference levels in the high frequencies. The bounds are extrapolated from the estimated levels of voiced speech, the details of which are discussed in Section 8.5.3.11. The speech estimates applied to the bounding process comprise the speech estimates from the lowest band to the band immediately below the voice commit maximum upper frequency, as well as the unmodified speech estimates in the bands above this following application of the minimum bound on the high frequencies. The limiter values are smoothly updated with the new speech reference levels.

Of course, the duration of voicing can often be far longer than 200 ms. When the voicing dominance persists for longer periods, there are further committals of the voice level estimates.

8.5.3.6 Updating the speech level estimates from sibilant speech

The dominant energy of speech in the high frequencies comes from sibilance.[180] The speech dominance detector provides four signals every 20 ms about the sibilance in speech as shown in Table 8-5.

Number	Speech sampling control signal
6	Sibilance dominance
7	Lower sibilance frequency
8	Upper sibilance frequency
9	Sibilance commit

Table 8-5. *Sibilance sampling control signals.*

Referring to Figure 8-14, in brief, when the sibilance dominance signal is true, sibilance is dominant in the frequency range from the lower sibilance frequency to the upper sibilance frequency. When the sibilance commit signal is true, then the signal since the sibilance dominance signal became true is considered to be dominated by sibilance within the time-varying frequency-range from the lower sibilance frequency to the upper sibilance frequency.

The approach for determining the speech level estimates from sibilant speech is very similar to that used in obtaining speech level estimates from voiced speech. The sibilance loudness level is estimated by the loudness level estimators when the sibilance is dominant, i.e. when the sibilance dominance signal is true. The sibilance level estimators are the same as the current loudness estimators, except that they are active only when sibilance is dominant. The resulting sibilance loudness level estimates are applied to a memory simulator to update the memory of speech loudness.

On starting the estimation, the estimates of past speech, (i.e. SRL dB values) are loaded into the sibilance estimator's memory. As with voiced speech, the reason for doing this is that during the intervening time between two episodes of sibilant speech the estimates may have been updated by the voicing estimator and these updates need to be included.

The approach is then to update the loudness estimates of the sibilant speech within the range of the identified sibilance-dominated frequencies and to extend this to higher frequencies when it appears that these will also contain sibilance.

8.5.3.7 Frequency range of sibilance updates

The speech dominance detector provides estimates of the lower sibilance frequency and upper sibilance frequency. As discussed previously in relation to the range of voiced frequencies, it is necessary to accommodate different band structures and so the frequency estimates are provided to the speech loudness estimator in units of Hertz and need to be converted to fit the band structure used. To ensure that there is no dominant noise present within the edge bands of the specified sibilance region, the edge bands are discarded and the sibilance band range commences with the band above the band corresponding with the lower sibilance frequency, the Lower Sibilance Band (LSB) and ends with the band below the band corresponding to the upper sibilance frequency, the Upper Sibilance Band (USB).

8.5.3.8 Sibilance updates

As long as the sibilance dominance signal is true, the estimates of the sibilance loudness level in the bands from the Lower Sibilance Band up to the Upper Sibilance Band are updated at the analysis rate (i.e. every 1 ms) and the sibilance band range being updated at the speech dominance detection rate (i.e. every 20 ms).

8.5.3.9 Updating the memory of speech loudness with new sibilance

During sibilance dominance, the memory of the loudness of speech is updated in the memory block (labelled Memory + Hi Freq Cap in Figure 8-14), a detailed schematic diagram of this is shown in Figure 8-16.

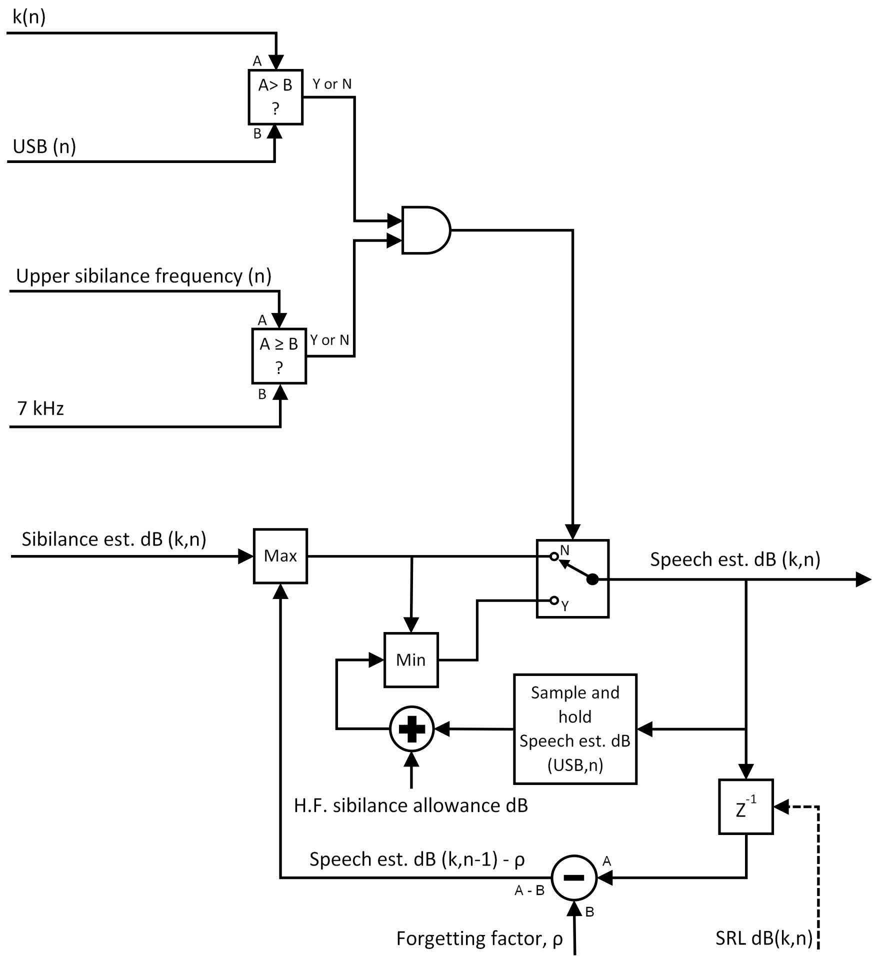


Figure 8-16. *Sibilance memory and high-frequency cap.*

At the beginning of a sibilance segment when sibilance dominance changes to true, the estimates of past speech values (SRL dB (k,n)) are loaded into the sibilance estimator's memory (dashed line into the unit delay in Figure 8-16). The memory simulator then updates its memory values for the sibilance band range from the Lower Sibilance Band up to the Upper Sibilance Band (i.e. LSB(n) ≤ k(n) ≤ USB(n)). The memory of speech is decreased by the forgetting factor, ρ, to produce new memory values (Speech est. dB (k,n-1) – ρ). Should a new estimate of the sibilance level (Sibilance est. dB (k,n)) exceed the corresponding new memory of speech, then the new memory is immediately replaced with the new sibilance level estimate. The operation of the memory simulation is given by Equation 8-10.

$$Speech\ est.dB(k,n) = Max\begin{Bmatrix} Speech\ est.\ dB(k,n-1) - \rho \\ Sibilance\ est.\ dB(k,n) \end{Bmatrix} \qquad \textbf{(8-10)}$$

$$for\ LSB(n)\ \leq k(n)\ \leq USB(n)$$

$$\begin{aligned} where:\ & \rho\ is\ the\ forgetting\ factor\ in\ dB\ per\ sample \\ & LSB\ is\ the\ Lower\ Sibilance\ Band\ number \\ & USB\ is\ the\ Upper\ Sibilance\ Band\ number \\ & k\ is\ the\ band\ number \\ & n\ is\ sample\ number \end{aligned}$$

The forgetting factor ρ is the same as the forgetting factor used for voiced speech.

The speech dominance detector operates with a sampling rate of 16 kHz and provides a maximum bandwidth of about 7.7 kHz in which to detect speech dominance. This bandwidth restriction is reasonable from an energy perspective as there is typically little speech energy above this frequency. The contribution to the total energy of speech from the band above 7.7 kHz is very small.[63] The benefit gained from capturing the signal at frequencies higher than 7.7 kHz, in terms of speech dominance detection, through using a higher sampling rate, would be minor compared with the greatly increased processing overhead required to do so. The SRL MKII scheme has been designed to process signals with sampling rates of up to 48 kHz, providing bandwidths of up to 24 kHz. According to the Speech Intelligibility Index, the contribution of speech above 7.7 kHz to intelligibility is around 1.5 %.[12] Despite the typically low contribution of speech above 7.7 kHz to speech energy and intelligibility there are speakers who produce /s/ with energy centred around 10 kHz. It would be desirable to capture the energy of sibilance above 7.7 kHz to preserve its naturalness and intelligibility. High-frequency sibilance does not occur without there also being significant sibilant energy below 7.7 kHz. An upper sibilance frequency of 7 kHz or more is therefore a very good indicator that there is strong sibilant energy extending to frequencies higher than 7.7 kHz. This can therefore be used as a trigger to extend the sampling of speech energy in bands above 7.7 kHz.

The method of treating bands above the speech dominance detection bandwidth is to:

- update the estimates of the sibilant loudness in these bands when there is sibilance dominance and the upper sibilance frequency is at least 7 kHz, and
- cap the level of the sibilance loudness estimate in these higher bands to the estimate in the highest sibilance band plus an allowance (labelled as H.F sibilance allowance dB in Figure 8-16) to ensure that high-level noise does not create estimates exceeding established sibilance levels by more than a defined allowance. In the evaluation of SRL MKII, described in Chapter 9, this allowance was set to 2 dB.

Equation 8-11 is an extension of Equation 8-10 which includes calculations for the bands above the Upper Sibilance Band.

$$Speech\ est.dB(k,n) =$$

$$\left[\begin{array}{l} Max\left\{\begin{array}{l} Speech\ est.\ dB(k,n-1)-\ \rho \\ Sibilance\ est.\ dB(k,n) \end{array}\right\} for\ LSB(n)\ \leq k(n)\ \leq USB(n) \\ Min\left\{\begin{array}{l} Max\left\{\begin{array}{l} Speech\ est.\ dB(k,n-1)-\ \rho \\ Sibilance\ est.\ dB(k,n) \end{array}\right\} \\ Speech\ est.\ dB(USB(n),n)+H.F.\ Allow \end{array}\right\} for\ \begin{array}{l} k(n) >\ USB(n)\ \& \\ USF(n)\ \geq\ 7\ kHz \end{array} \end{array}\right. \qquad \textbf{(8-11)}$$

$$\begin{array}{ll} where: & \rho\ is\ the\ forgetting\ factor\ in\ dB\ per\ sample \\ & LSB\ is\ the\ Lower\ Sibilance\ Band\ number \\ & USB\ is\ the\ Upper\ Sibilance\ Band\ number \\ & USF\ is\ the\ Upper\ Sibilance\ Frequency \\ & H.F.Allow\ is\ the\ High\ Frequency\ \ Sibilance\ Allowance\ dB \\ & k\ is\ the\ band\ number \\ & n\ is\ the\ sample\ number \end{array}$$

8.5.3.10 Updating and committing the dominant sibilance estimates

As discussed in the section on updating and committing voicing, it is preferable to only update the speech reference levels once there is good confidence the signal has been dominated by speech (i.e. voicing in that case and sibilance in this case). The method of updating estimates and then committing them used for voicing is also used for updating the sibilance estimates and committing them to the speech loudness estimates and hence the speech reference levels.

If the sibilance commit signal becomes true, then the speech estimates from the sibilance dominance level estimator become the new speech reference levels following the application of a minimum bound on the mid-frequencies (Min Mid Freq Bounding) as shown in Figure 8-14. This bound is to ensure that in the absence of updates covering the mid-frequencies there are always reasonable minimum speech reference levels in the mid-frequency region. These bounds are interpolated from the estimated levels of the voiced and sibilant speech, the details of which are discussed in Section 8.5.3.12. The speech estimates comprise the speech estimates from sibilance updates in the frequency range from the lowest sibilance band to the upper sibilance band or the maximum band, depending on the upper frequency of the sibilance, as well as the unmodified speech estimates in the bands below the minimum sibilance band. As with the voicing committal, the sibilance committal causes the limiter values to be smoothly updated with the new speech reference levels.

8.5.3.11 Minimum high-frequency levels extrapolated from voicing levels

There are some situations in which high-frequency noise is sufficiently intense that sibilance detection occurs infrequently or not at all. In such cases, it is necessary to extrapolate an appropriate set of high-frequency speech loudness levels from the voicing levels. The speech spectrum given in the SII[12] for speech spoken with normal vocal effort indicates that the slope of the speech spectrum above 1 kHz is fairly constant and falls at a rate of approximately 5 dB / octave. Third-octave band presentation of this data is shown in Figure 8-17.

While this figure shows the typical spectrum in the free field, the level, as a function of frequency, may vary due to electronic processing, such as hearing aid amplification and telephone amplification. It is therefore preferable to calculate the slope from analysis of the signal. Knowing this slope and the level in a band at 1 kHz makes it possible to better predict the high-frequency speech levels.

This prediction is achieved by determining the most recent stable upper voice band. The estimated speech loudness level in this band, or a band around 3 kHz, whichever is the higher, is compared with the estimated speech loudness level in a band around 1 kHz to obtain a slope in dB / band for the particular band structure.

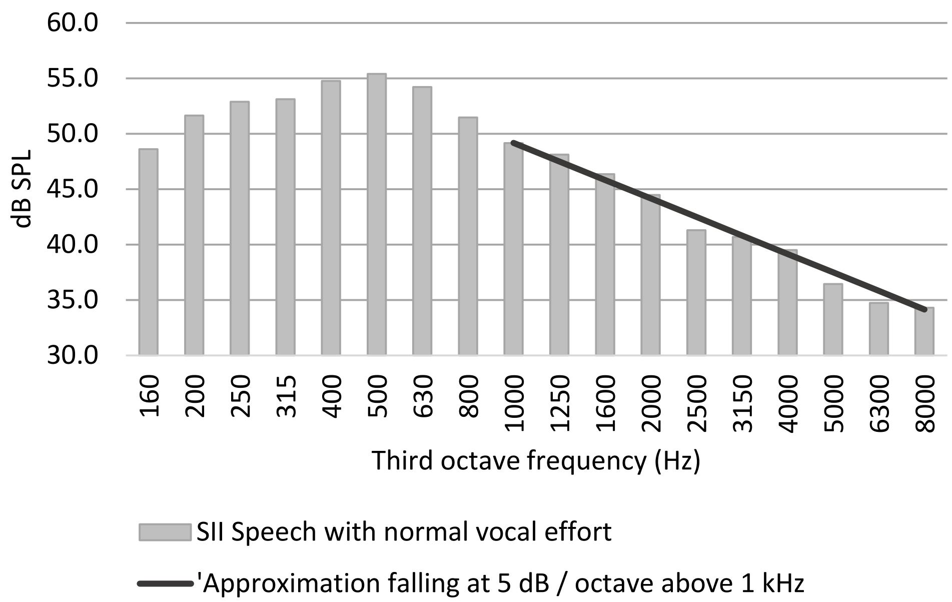


Figure 8-17. *Third-octave spectrum of speech, 'normal' vocal effort, and superimposed -5 dB / octave slope from 1 kHz.*

This slope is express in Equation 8-12.

$$slope(n) = min\left\{0, \frac{Speech\ est.dB\ (RSUVB(n)) - Speech\ est.dB\ (1k\ band\ number, n)}{RSUVB(n) - 1\ k\ band\ number}\right\} \qquad \textbf{(8-12)}$$

$$where\text{:}\quad RSUVB\ is\ the\ Recent\ Stable\ Upper\ Voice\ Band\ number$$
$$1k\ band\ number\ is\ the\ 1kHz\ band\ number$$

The slope is then used to extrapolate minimum values for the level estimates at high frequencies using the Recent Stable Upper Voice Band as a base. If the level estimates (Speech est. dB (k,n)) at high frequencies are higher than these minimum values, they remain unchanged. The extrapolated minimum values are essentially a fall-back. The extrapolation is expressed in Equation 8-13.

$$Speech\ est.\ dB(k,n) =$$

$$max\begin{Bmatrix} Speech\ est.dB(k,n) \\ Speech\ est.dB(RSUVB(n),n) + slope(n) \times (k(n) - RSUVB(n)) \end{Bmatrix} \qquad \textbf{(8-13)}$$

$$for\ RSUVB(n) < k(n) \leq K$$

$$
\begin{aligned}
where:\ & k \ is \ the \ index \ for \ the \ band \\
& n \ is \ the \ sample \ number \\
& RSUVB \ \ is \ the \ Recent \ Stable \ Upper \ Voice \ Band \ number \\
& K \ is \ the \ maximum \ band \ number
\end{aligned}
$$

A schematic diagram of the high-frequency speech loudness extrapolator is shown in Figure 8-18.

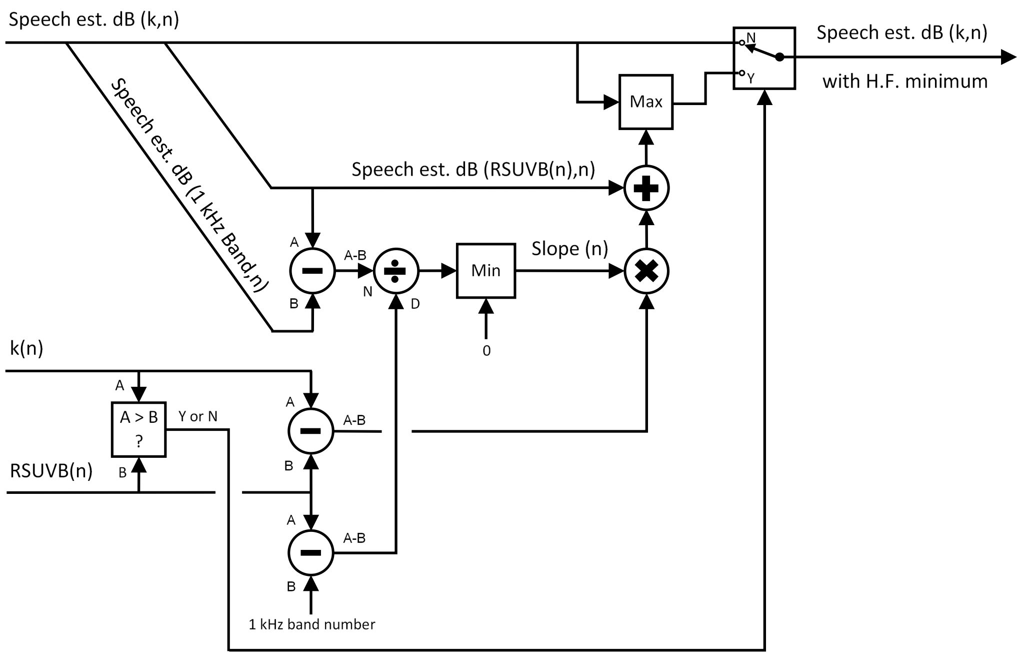


Figure 8-18. *High-frequency speech loudness extrapolator.*

8.5.3.12 Interpolating minimum mid-frequency speech level estimates

There is normally considerable overlap between the spectral energy of voicing and sibilance. However, there can be situations in which mid-frequency noise is present to a degree that reduces the number of occasions that speech in this region is found to be dominant. Despite regular updates of low-frequency voicing and high-frequency sibilance the mid-frequency speech level estimates will fall. It is desirable to interpolate minimum levels for bands in this mid-frequency region from the low-frequency voicing levels and the high-frequency sibilance levels. The missing mid-frequency updates of speech level estimates are detected by tracking the recent upper voicing frequencies and recent lower sibilance frequencies. If there is a gap between them, then there are mid-bands that have not been recently updated. The interpolation selected for use within the gap is a linear weighting of the levels on either side of the gap. If, however, the estimated speech level in a band in the gap exceeds the interpolated value, the estimated speech level is retained. The interpolation is expressed by Equation 8-14.

$$Speech\ est._{out}\ \ dB(k,n) = max \left\{ \begin{matrix} Speech\ est._{in}\ \ dB(k,n) \\ \frac{Speech\ est._{in} dB(LE(n)) \times (UE(n) - k(n)) + Speech\ est._{in}\ dB(UE(n)) \times (k(n) - LE(n))}{UE(n) - LE(n)} \end{matrix} \right\} \quad \textbf{(8-14)}$$

$$for\ LE(n)\ < k(n) < UE(n)$$

$where$: $k\ is\ the\ band\ number$

$n\ is\ the\ sample\ number$

$LE\ (lower\ edge)\ is\ the\ index\ of\ the\ highest\ recent\ updated\ voiced\ band$

$UE\ (upper\ edge)\ is\ the\ index\ of\ the\ lowest\ recent\ updated\ sibilance\ band$

A schematic diagram of the mid-frequency speech loudness interpolator is shown in Figure 8-19.

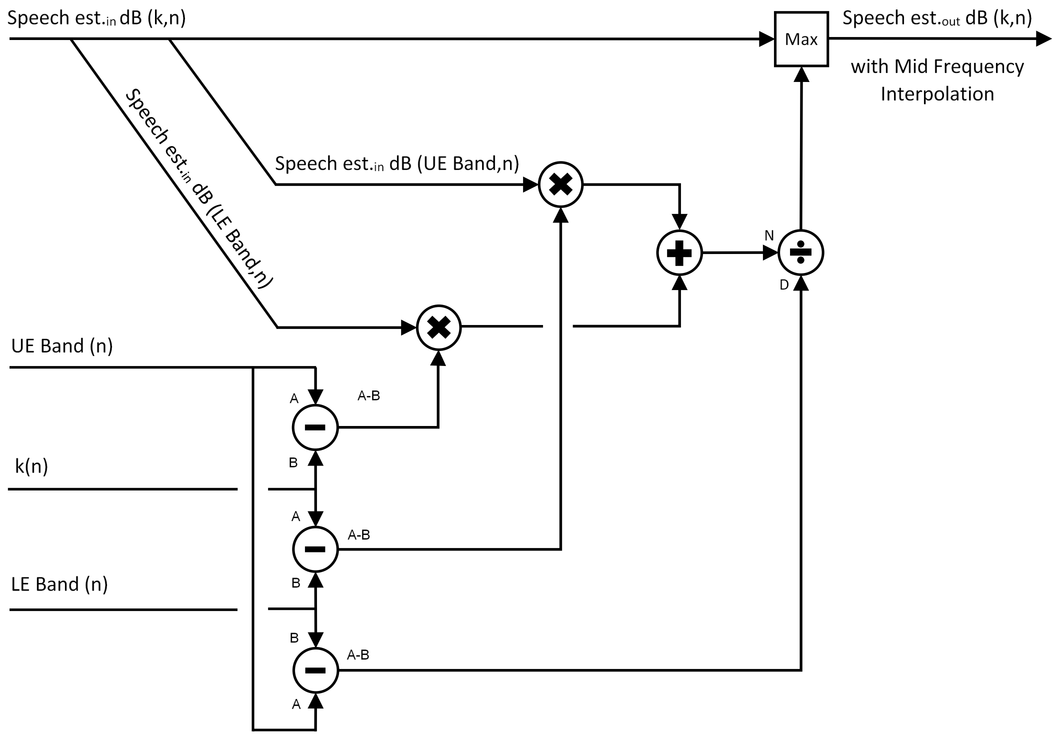


Figure 8-19. *Mid-frequency speech loudness interpolator.*

8.6 Coding, application and code performance

The SRL MKII scheme was developed in MATLAB.[146] The signal processing algorithm was then re-coded in the ANSI C programming language.[147] The novel functions of the gain calculator and the speech dominance detector exist as separate modules, as well as within the combined system, enabling them to be incorporated into other researchers/developers' test platforms. The algorithm is currently being tested by other researchers using their platforms.

The algorithm typically runs 46 times faster than real time, for one channel with a sampling rate of 48 kHz, on a current laptop computer (Intel® Core™ i7-4600M CPU @ 2.89 GHz) using Windows 8.1. About 2% of the computer's processing resources is required to process one channel at 48 kHz. Additional channels require less resources per channel.

A MATLAB mex function incorporating the C-coded algorithm was created to enable fast processing under MATLAB.

A Windows application was created to enable other researchers/developers to use the SRL scheme. This application enables wav files with a range of sampling rates and channels to be processed. The user interface is shown in Figure 8-20.

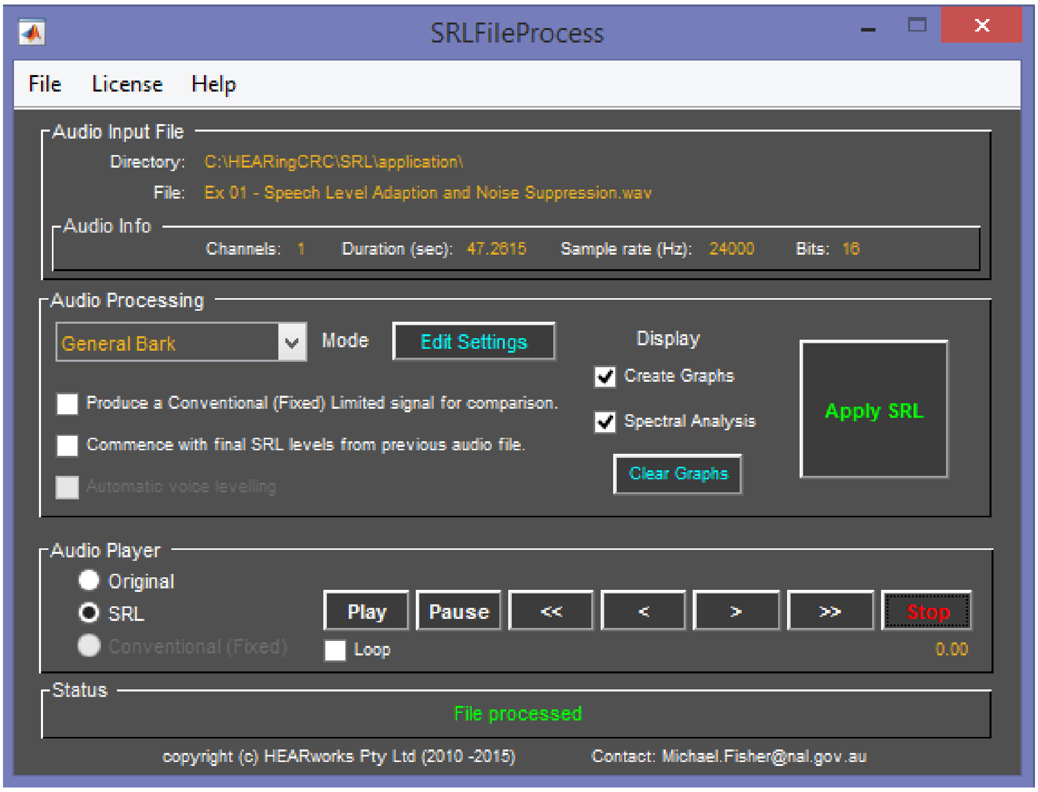


Figure 8-20. *The SRL application.*

An application guide for the SRL MKII program is attached in Appendix E.

The application may be downloaded from the Hearing CRC web site:

http://www.hearingcrc.org/xc/xc5-applications-of-speech-referenced-limiting/

8.7 Summary

This chapter presented the SRL MKII scheme. This scheme was considerably more advanced than the SRL MKI scheme. It contained a frequency-specific speech dominance detector, enabling the scheme to sample speech in frequency regions where speech dominates while simultaneously rejecting other frequency regions in which noise dominates. It contained a gain calculator that had a more advanced method of creating speech reference levels compared to SRL MKI and was considerably more robust in noisy conditions. It performed dual-speed, multi-band limiting, enabling it to better control sound with less delay. It had a maximum delay of 3 ms making it acceptable in a greater range of applications where the user hears their own voice through the processing as well as via natural air and bone conduction. The SRL MKII also compensates for any increase in loudness due to loudness summation in relation to speech loudness. The objective performance of SRL MKII was found to be consistent across different languages.

Chapter 9

Subjective evaluation of SRL MKII

9 Subjective evaluation of SRL MKII

9.1 Introduction

This chapter reports on the subjective evaluation of the SRL MKII scheme.

It was hypothesised that, when listening to a speech signal, listening comfort would be improved through limiting the loudness of non-speech sounds with reference to the loudness of the speech to which the listener was acclimatised. It was also hypothesised that this processing would have minimal effect on the speech quality. These hypotheses rely on there being times when speech is sufficiently dominant within the signal for its loudness to be estimated.

As discussed, in the conclusion to the evaluation of the SRL MKI, in Chapter 7, there were several shortcomings in the method used to evaluate the SRL MKI scheme, these were:

1. The SRL scheme was not compared against a conventional fixed-reference limiting (FRL) scheme and therefore the experiments did not provide evidence of an advantage of the SRL scheme over a conventional FRL scheme despite finding a highly significant reduction in the loudness of noise in relation to speech for the combined speech and noise recordings.
2. The SRL scheme was not assessed using large abrupt changes in speech levels and therefore the experiments did not provide evidence of fast-adaptive speech referenced loudness control nor evidence of the preservation of speech quality under these circumstances.
3. The SRL scheme was not assessed with speech and high-level noise being simultaneously present and therefore the experiments did not provide evidence about the controlling of loudness of noise when it was concurrent with speech nor the effect such control would have on speech quality.

To address these shortcomings, and to assess the SRL MKII scheme in the three potential applications for which the SRL MKI scheme was evaluated, five laboratory experiments were conducted. In all five experiments, the SRL MKII scheme was compared against an equivalent conventional FRL scheme that was identical to the SRL scheme, except that the limits were fixed and there was no correction for loudness summation (see Section 9.2.1 for a detailed description). The first three experiments were targeted at the three applications for which the SRL MKI scheme was previously evaluated, these being:

- hearing aids,
- level-dependent hearing protectors, and
- telephone headsets.

The fourth experiment was aimed at assessing SRL's ability to control the loudness of high-level noise in a situation of large abrupt changes in speech level and its ability to preserve speech quality under those circumstances. The fifth experiment was aimed at assessing the effect of high-level noise simultaneously presented with speech on SRL's ability to control the loudness of high-level noise in relation to the speech loudness and to preserve speech quality.

The aim was to obtain subjective data on both the loudness control and the speech quality provided by the method. It was hypothesised that the SRL MKII scheme would provide the greatest reduction in the excess loudness of an audio signal compared with the loudness of the preceding speech conveyed by the audio signal for the least reduction in the speech loudness and quality.

9.2 Method

The experiments comprised:

1. collecting stimuli that may be encountered when using the SRL method in the three selected applications as well as designing test stimuli;
2. processing these stimuli using the SRL MKII scheme described in Chapter 8 in:
    - fixed-reference limiting (FRL) mode,
    - speech-referenced limiting (SRL) mode;
3. presenting to subjects in the laboratory the:
    - FRL-processed stimuli,
    - SRL-processed stimuli,
    - unprocessed speech and level-adjusted unprocessed noise.
4. collecting the subjects' responses and analysing them.

Unlike the method used in the SRL MKI experiments, the unprocessed noise was not presented to the subjects without alteration. Instead, the unprocessed noise, which was typically recorded along with the speech in a given situation, was adjusted in level so that it had the same RMS level as the speech it was paired with. This was so the subjects would always be presented with a noise accompanying the speech, even in the reference (unprocessed stimuli) condition, but would not be presented with high-level noise.

There were two reasons for this approach. First, I did not want to present the subjects with noise at a level that they would not willingly subject themselves to, which meant allowing the subjects to set the noise limit themselves. Given this requirement, it was not possible to present the unprocessed noises to the subjects at the recorded level.

Second, in order for the subjects to compare the loudness and quality of the FRL- and SRL-processed speech with the unprocessed speech, it was necessary to present the unprocessed speech. However, if the unprocessed speech was presented without accompanying noise in the loudness-rating experiments, the subjects would identify it as being different (i.e. by the absence of the noise) and therefore it would not be a well hidden reference. The unprocessed noise therefore needed to be set to a sound level which would not make it stand out and preferably at a level that would give some additional insight into the subjects' perception of noise loudness. Setting the RMS level of the noise to the RMS level of its associated speech made the reference noise similar in loudness to the SRL-processed noise and therefore less easily identified. Furthermore, setting the RMS level of the noise to that of its associated speech resulted in data being produced that gave insight into perceptual differences of noises and speech of the same RMS level.

In terms of the control of loud sounds, these experiments compared fixed-reference limiting and speech reference limiting. To make this comparison fair and simple, the same processing scheme was used for both limiting methods with only two differences. All the parameters associated with the two limiting schemes were also identical. The two methods differed only in that:

1. the SRL limits varied with the estimated speech loudness within a predefined range as opposed to being fixed, and
2. the SRL processing incorporated a correction to the limiting gains applied to the signal based on an estimate of loudness summation.

The FRL scheme was therefore referred to as the *equivalent conventional FRL scheme.*

9.2.1 Fixed-reference limit calculation

In the experiments, the subjects were able to set their preferred fixed-reference limiting levels using a single limiting level control. This control, however, adjusted by the same amount the limiting levels in all the bands of the multi-band limiter. The adjustable frequency-dependent limiting thresholds were derived from the NAL-SSPL prescription for the real-ear saturation response (RESR).[110,113] The second row in Table 9-1 shows the NAL-SSPL prescribed RESR for a person with a 0 dB hearing threshold level (HTL). Both the FRL and SRL limiting schemes in these experiments used a multi-band structure based on the critical-band scale.[45] The number of bands in the FRL / SRL scheme was more than 16. The NAL-SSPL prescription recommends reducing the RESR values by 8 dB for 16 bands but makes no recommendation for more than 16 bands. I chose a reduction of 10 dB in the RESR values given that there were more than 16 bands. These adjustments are shown in Table 9-1 as 'Multi-band correction factor'. This correction factor was at most 2 dB too conservative compared to the recommendation for 16 bands. As the subjects set their individualized final limits on a broadband basis this adjustment factor was not important with the exception that it affects the range of fixed-limiting levels available to the subjects, however, this range transpired to be more than adequate for the subjects. It did, however, influence the lower bound for the SRL limit range as described in the next section.

For processing the stimuli for these experiments, the absolute level in the FRL and SRL schemes was referenced to sound levels in the diffuse field (DF). The multi-band corrected real-ear (RE) limits were therefore converted to the diffuse field by the 'RE to DF correction' factors,[125] shown in Table 9-1. The RESR is primarily intended for prescribing the saturation response for hearing aids and is only defined for octave frequencies from 250 Hz to 4 kHz. The values at octave frequencies above and below these frequencies were extrapolated to produce the diffuse field (equivalent) saturation response (DFSR) for use in the FRL and SRL schemes. The FRL scheme takes the DFSR values defined at octave frequencies from which it produces limits for all its bands.

Octave frequencies	Hz	63	125	250	500	1000	2000	4000	8000	16000
RESR (HTL = 0 dB)	dB SPL			95	96	95	98	100		
Multi-band correction factor	dB			-10	-10	-10	-10	-10		
	dB SPL			85	86	85	88	90		
RE to DF correction	dB	0	0	-0.5	-1.5	-5	-10.5	-11.5	-6.5	-10.5
DFSR	dB SPL			84.5	84.5	80	77.5	78.5		
Extrapolation	dB SPL	85	85	84.5	84.5	80	77.5	78.5	78.5	78.5
DFSR (Nearest dB)	**dB SPL**	**85**	**85**	**85**	**85**	**80**	**78**	**79**	**79**	**79**

Table 9-1. *NAL-SSPL – conversion from real-ear saturation response (RESR) for 0 dB HL to diffuse-field (equivalent) saturation response (DFSR) for a critical-band-based multi-band limiter.*

9.2.2 SRL & FRL settings and parameters

The SRL (and FRL) application was set to process the stimuli sound files as follows:

- The application's user interface is shown in Figure 9-1. The Mode of operation was set to 'General Bark', where 'General' is the default processing mode with no specific options employed and 'Bark' is the critical-band-based multi-band structure employed. The sampling rate, 'Sampling Rate (Hz):', is derived from the sound file being processed, which for all stimuli in these experiments was 48000 Hz. The number of bits, 'Bits', was derived from the sound file being processed, which for all stimuli in these experiments, was 16 bits.

- The settings, as defined in the SRLsetting.xlsx spreadsheet accessed from the user interface 'Editing Settings' button, are shown in Figure 9-2.

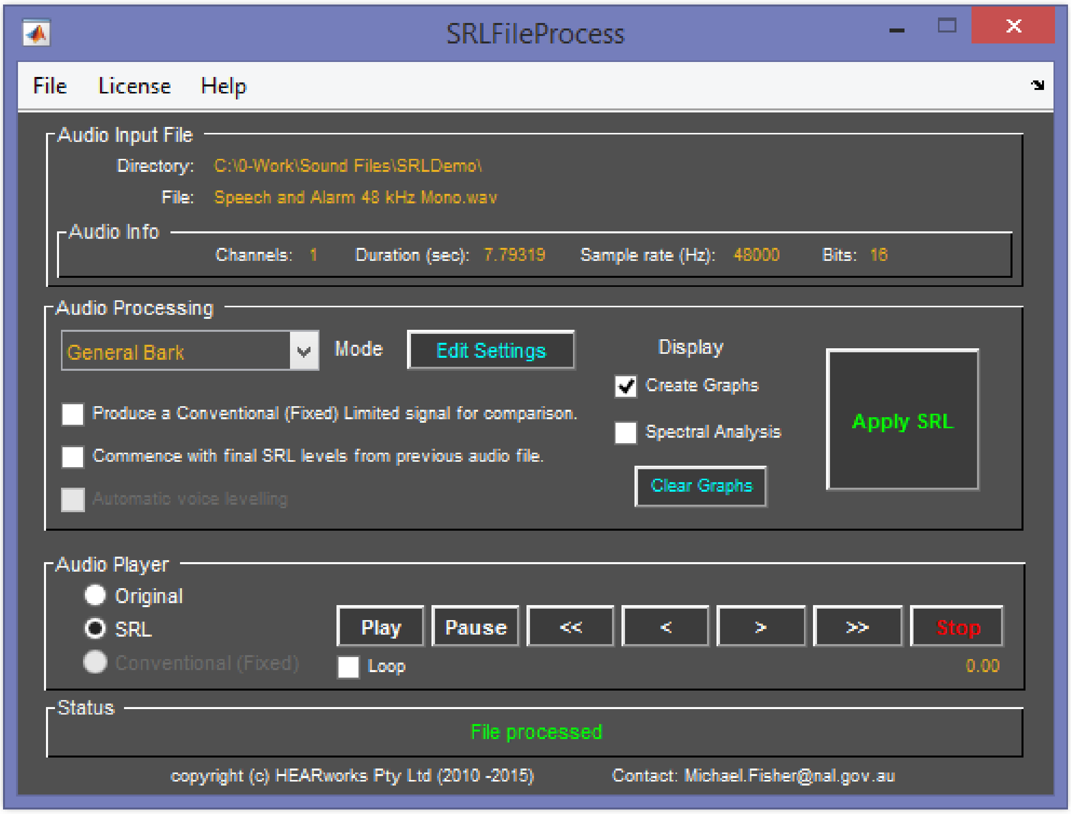
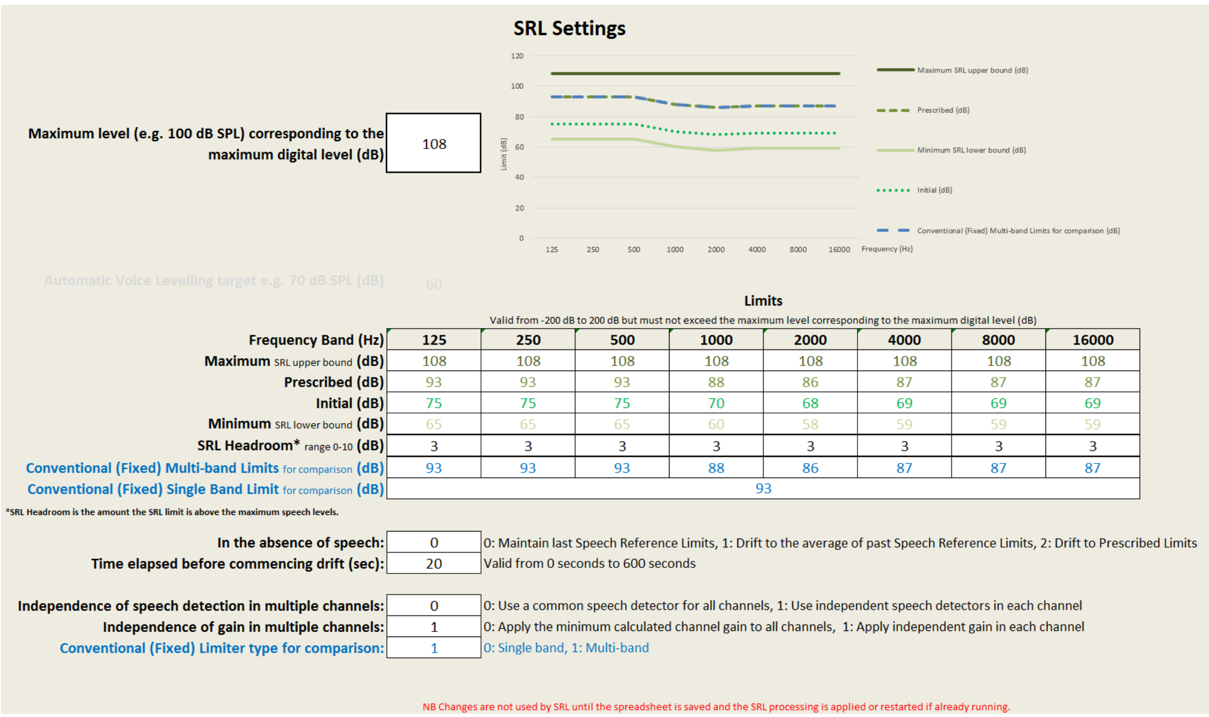


Figure 9-1. *The SRL (and FRL) application's user interface with the Mode set to 'General Bark'.*



Maximum level (e.g. 100 dB SPL) corresponding to the maximum digital level (dB): 108

Automatic Voice Levelling target e.g. 70 dB SPL (dB): 60

Limits

Valid from -200 dB to 200 dB but must not exceed the maximum level corresponding to the maximum digital level (dB)

Frequency Band (Hz)	125	250	500	1000	2000	4000	8000	16000
Maximum SRL upper bound (dB)	108	108	108	108	108	108	108	108
Prescribed (dB)	93	93	93	88	86	87	87	87
Initial (dB)	75	75	75	70	68	69	69	69
Minimum SRL lower bound (dB)	65	65	65	60	58	59	59	59
SRL Headroom* range 0-10 (dB)	3	3	3	3	3	3	3	3
Conventional (Fixed) Multi-band Limits for comparison (dB)	93	93	93	88	86	87	87	87
Conventional (Fixed) Single Band Limit for comparison (dB)	93							

*SRL Headroom is the amount the SRL limit is above the maximum speech levels.

In the absence of speech:	0	0: Maintain last Speech Reference Limits, 1: Drift to the average of past Speech Reference Limits, 2: Drift to Prescribed Limits
Time elapsed before commencing drift (sec):	20	Valid from 0 seconds to 600 seconds
Independence of speech detection in multiple channels:	0	0: Use a common speech detector for all channels, 1: Use independent speech detectors in each channel
Independence of gain in multiple channels:	1	0: Apply the minimum calculated channel gain to all channels, 1: Apply independent gain in each channel
Conventional (Fixed) Limiter type for comparison:	1	0: Single band, 1: Multi-band

NB Changes are not used by SRL until the spreadsheet is saved and the SRL processing is applied or restarted if already running.

Figure 9-2. *SRL settings for the processing of stimuli for the experiments.*

Details of the settings for the processing shown in Figure 9-2 are as follows:

Maximum level (e.g. 100 dB SPL) corresponding to the maximum digital level (dB): This is the maximum SPL corresponding to digital saturation, i.e. 0 dBr (the maximum for the input and output wav files). This was set to 108 dB SPL peak.

Maximum $_{\text{SRL upper bound}}$ (dB): These fast limits would normally be set to around the loudness discomfort level (plus a fast to slow correction factor as explained below). However, in this case they were set to digital saturation, i.e. 108 dB SPL. The reason for this is explained later in the method section.

Prescribed (dB): These are the default limiting levels for when no dominant speech is detected for some time. These values were set to the DFSR values calculated in Table 9.1 (plus a fast-to-slow correction factor of 8 dB). However, while determining other values, they were not used directly in the experiments as the mode of operation for '*In the absence of speech:*' was set to 0, *'Maintain last Speech Reference Limits'.*

Initial (dB): These are the initial limit levels for the fast limiter (the slow limiter's limits are 8 dB below these). These are set so that the limits are initially quite low but not so low that initial speech would be heavily suppressed.

Minimum (dB): This is the lowest limit level for the fast limiter (the slow limiter's limits are 8 dB below these). These are set to prevent the adoption of the loudness of low-level speech as limiting thresholds. For these experiments, the minimum fast limiter limits were set to be 28 dB below the prescribed fast limits.

SRL Headroom* range $_{0\text{-}10}$ (dB): This is the amount by which the limit exceeds the maximum level of dominant speech. It was set to 3 dB at all frequencies. This allowed a small margin for new speech to exceed previous speech levels without any limiting. It also allowed high-level noises to slightly exceed the previous speech levels.

Conventional (Fixed) Multi-band Limits $_{\text{for comparison}}$ (dB): These were set to the DFSR values calculated in Table 9-1 plus the fast-to-slow limit threshold relationship of 8 dB. As will be explained in the method section, in the experiments, these values were overwritten by a script that performed batch processing of the stimuli (sound files) at various fixed-limit levels to produce processed sound files at a range of limit levels.

Conventional (Fixed) Single Band Limit $_{\text{for comparison}}$ (dB): This value was set to 93 dB but was not used in these experiments as the '*Conventional (Fixed) Limiter type for comparison:*' was set to 1, *'Multi-band'*.

In the absence of speech: This was set to 0, *'Maintain last Speech Reference Limits'*. Therefore, only the SRL limits were used when in SRL operational mode. This prevented any prescribed or average speech reference levels being used during processing of the stimuli.

Time elapsed before commencing drift (sec): This was set to 20 seconds. However, it was not used in the experiments as '*In the absence of speech:*' was set to 0, *'Maintain last Speech Reference Limits'.*

Independence of speech detection in multiple channels: This was set to 0, *'Use a common speech detector for all channels'*. Only one speech detector was used. This was of no consequence in processing of the stimuli for these experiments as either the signals were identical in each channel or the alternative channel was silent.

Independence of gain in multiple channels: This was set to 1, *'Apply independent gain in each channel'*. This was also of little consequence in the experiments due to the signals either being identical in both channels or the alternative channel being silent.

Conventional (Fixed) Limiter type for comparison: This was set to 1, '*Multi-band'*. In these experiments, all comparisons were made to a multi-band fixed-reference limiter.

Table 9-2 displays the main SRL MKII processing parameters that were not available via the user interface or the application's spreadsheet but were relevant to the performance of SRL and FRL in processing the stimuli for the experiments.

Parameter	Value
Analysis rate	1 kHz
Analysis window duration	4 ms
Fast-limiter active	Yes
Slow-limiter active	Yes
Fast-limiter attack time	1 ms (2 ms in practice)
Fast-limiter release time	250 ms
Slow-limiter attack time	100 ms
Slow-limiter release time	500 ms
Fast-to-slow limit threshold relationship (init., max & min)	8 dB
Loudness-summation band rate	Critical band rate - Bark scale
Loudness-summation maximum reduction - fast	10 dB
Loudness-summation maximum reduction - slow	8 dB
Speech envelope release rate	12 dB / s
Time period over which a new speech update occurs	200 ms
Rate of convergence to new speech	100 ms
Band linking: maximum inter-band difference	30 dB

Table 9-2. *SRL MKII processing parameters relevant to the pre-processing of stimuli for the experiments.*

9.2.3 Stimuli and their processing

Each of the five experiments presented four sets of stimuli to the subjects. In addition to the five experiments, there was a training session in which two additional sets of stimuli were presented. There were in total 22 sets of stimuli (5 x 4 + 1 x 2). Each set began as a single recording of a phrase of speech and noise, called the unprocessed stimulus. The speech and noise were sequential for the training session and the first four experiments and simultaneous for Experiment 5. All the stimuli were pre-scaled so that digital saturation of the sound files corresponded to 108 dB SPL peak.

9.2.4 SRL processing

The stimuli were processed by the SRL scheme using the parameters and settings as described above. There was no maximum fixed limit other than the digital saturation level. This was because a range of maximum limits was applied subsequently to the SRL-processed stimuli in the same manner as it was to the unprocessed stimuli, as described in the next section. To mimic the pre-establishment of speech reference levels that would have normally occurred if SRL were used outside these experiments, the SRL algorithm was first presented with some additional speech from the speaker whose voice was used in the experiment material. Following this pre-establishment or 'priming' of the speech reference levels, the SRL algorithm was presented with the stimulus and with the processing option '*Commence with final SRL levels from previous audio file*' enabled (see Figure 9-1). This 'priming' process was done for all the stimuli except for those presented in Experiment 4.

For Experiment 4, a concatenated sequence of the four stimuli was processed by the SRL scheme in one go so that the effect of abruptly changing speech levels would be captured. The four SRL-processed stimuli were then separated and saved along with the other SRL-processed stimuli.

9.2.5 FRL, SRL maximum level and reference processing

The unprocessed stimuli and the SRL-processed stimuli were then processed by the scheme, this time in fixed-reference limiting mode, to produce fixed-reference limited versions of the unprocessed stimuli and the SRL-processed stimuli. This was to produce sets of stimuli for the determination of the subject's maximum sound level and stimuli with corresponding maximum limits for the training and the experiments. Applying identical maximum levels to the SRL-processed stimuli ensured that subjects were equally in control of the maximum levels from the SRL-processed stimuli as they were from the FRL-processed stimuli. This maximum limit applied to the SRL-processed stimuli could have been applied by SRL during its processing of the stimuli, however, for logistical reasons it was done together with the FRL processing. This process was performed for 17 fixed limiting levels in 3 dB steps to produce a range of both FRL-processed and SRL-processed stimuli with the same maximum limit levels. See Appendix C for details of the stimulus generation.

Together each stimulus set comprised 70 files as follows:

1. The reference (unprocessed) speech and noise (with noise RMS = speech RMS).
2. The SRL-processed speech and noise, with 17 maximum fixed-limit values.
3. The FRL-processed speech and noise, with 17 maximum fixed-limit values.
4. The reference (unprocessed) speech only.
5. The SRL-processed speech only, with 17 maximum fixed-limit values.
6. The FRL-processed speech only, with 17 maximum fixed-limit values.

There were two exceptions to the above stimulus set composition: 1. there were twice as many speech-only files (i.e. items 4, 5 and 6 above) involved in Experiment 4, as the speech quality was evaluated separately for the initial and final parts of the speech stimuli, and 2. the stimuli set in Experiment 5 comprised only the SRL-processed and FRL-processed speech and noise files (i.e. items 2 and 3 above).

9.2.6 The stimuli

The unprocessed stimuli for Experiments 1, 2 and 3 were identical to the unprocessed stimuli used in Experiments 1, 2 and 3 of the evaluation of the SRL MKI scheme (see Chapter 7). The unprocessed stimuli for these three experiments comprised a speech phrase followed by a noise that might be encountered in the following applications:

1. hearing aids,
2. level-dependent hearing protectors, and
3. telephone headsets.

Details of the unprocessed stimuli for these first three experiments were given in Chapter 7.

The unprocessed stimuli for Experiment 4 comprised speech at various RMS levels followed by a high-level one-second burst of white noise. The speech was spoken by a professional female news reader and recorded at the National Acoustic Laboratories.[181] The speech was the following sentence from the Rainbow Passage: '*When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow*'.[182] It was equalized so that its third-octave, long-term spectrum would be within 6 dB of the International Long-term Average Speech Spectrum.[63] The equalization was minor in its magnitude. The RMS level of the sentence was different for each of the four stimuli. The levels corresponded to the vocal effort categories given in Table 9-3. The three highest of these were the speech levels defined in the

American National Standard, ANSI S3.5-1997.[12] The lowest of these was the typical speech level found by Pearsons et al. (1977) for ambient noise up to a level of 48 dB SPL.[140]

Vocal effort	Speech level dB SPL RMS	Sequence position
Loud	75	4
Raised	68	3
Normal	62	1
Casual	55	2

Table 9-3. *Vocal effort, speech levels and stimuli sequencing for Experiment 4.*

The sequence was chosen to obtain data for the following changes in speech level: -7 dB, +13 dB and +7 dB. The level of the burst of white noise that followed the speech was 90 dB SPL for all four stimuli.

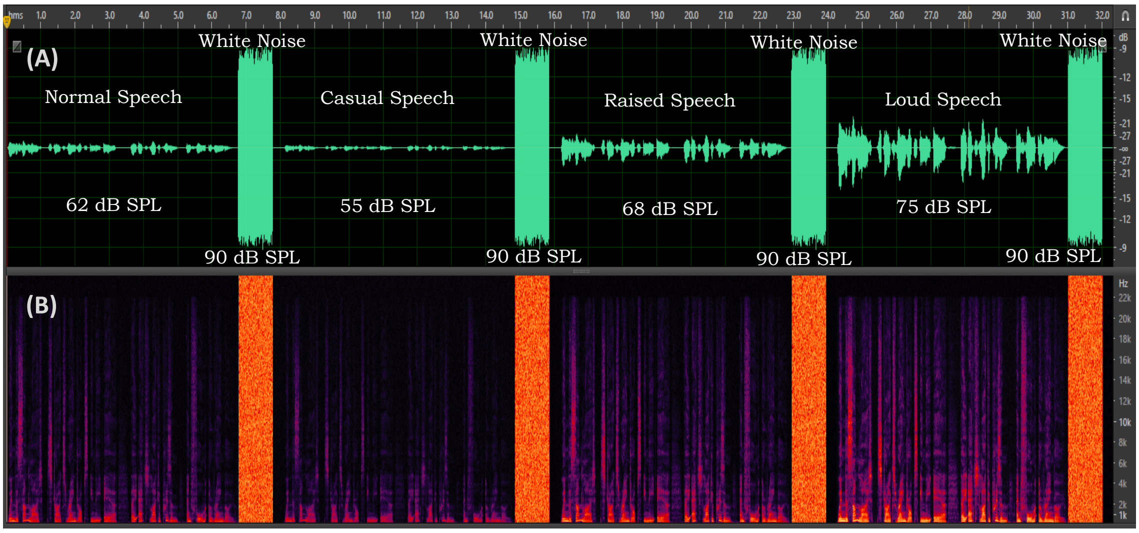


Figure 9-3. (A) *The waveform and* ***(B)*** *the spectrogram of the concatenated unprocessed stimuli sequence to be subjected to the FRL and SRL processing in preparation for Experiment 4.*

Figure 9-3 shows the waveform and spectrogram of the concatenated unprocessed stimuli sequence to be subjected to the FRL and SRL processing in preparation for Experiment 4. The strong contrast between the speech and noise bursts is clearly evident. The magnitude of the contrasts in sequence were 28 dB, 35 dB, 22 dB and 15 dB.

The unprocessed stimuli for Experiment 5 comprised:

1. **Speech in a travelling car:** This was a recording of speech made in a car travelling at a speed of about 50 km/hour. The recording microphone was at the position where an adult passenger's head would have been should they have been seated in the front passenger seat. The speech was that of the male driver of the car who said '*I'm accelerating here, going down the street*' while facing the windscreen. The RMS SpNR was -30 dB.
2. **Speech and hammering:** This was the same stimulus as the unprocessed stimulus 2 in Experiment 2 but with the hammering overlaid on the speech rather than being sequential. The RMS SpNR was -30 dB.
3. **Speech and pressure cleaner:** This was the recording of the speech from stimulus 3 in Experiment 2 and the noise from stimulus 4 in Experiment 2. The speech was increased by 7 dB and the noise reduced by 8 dB to simulate a raised/loud voice and the listener facing away from the noise source and towards the talker. The RMS SpNR was –10 dB.
4. **Speech and alarm:** This was a combination of an evacuation announcement and alarm noise. The speech was taken from an announcement intended to be reproduced over a public address system in which a male speaker says *'Evacuate as directed'* at an RMS level of 80 dB SPL. The recorded alarm sound was a harmonic series comprising a fundamental and harmonics 2, 3 and 4. Its fundamental frequency sweeps cyclically between 500 Hz and 1.5 kHz every 360 ms. The RMS SpNR was –15 dB.

	Sound level (RMS dB SPL $_{\text{diffuse field equivalent}}$)							
	Stimulus 1		Stimulus 2		Stimulus 3		Stimulus 4	
Experiment	Speech	Noise	Speech	Noise	Speech	Noise	Speech	Noise
1	72	99	66	87	72	93	76	81
2	64	90	53	83	65	87	65	90
3	70	90	69	82	68	87	66	87
4	62	90	55	90	68	90	75	90
5	68	98	53	83	72	82	80	95

Table 9-4. *Unprocessed speech and noise levels for all stimuli in the five experiments.*

A summary of the speech and noise levels of all the unprocessed stimuli in the five experiments is given in Table 9-4. Across all the experiments, the RMS level of the speech ranged from 53 to 80 dB SPL and the RMS level of the noise ranged from 81 to 99 dB SPL.

9.2.7 Limit setting stimuli sequences and processing

In order for the subjects to select an appropriate maximum sound level (limiting level) for the stimuli they were going to hear, they were given the opportunity to listen to the high-level noise from the stimuli at a range of maximum sound levels, starting at a low maximum level. The high-level noise stimuli for Experiments 1 and 3 (hearing aids and telephone headsets) were grouped together as they were both band-limited (bandwidth = 4 kHz) and to be presented monaurally. The high-level noise stimuli for Experiments 2, 4 and 5 stimuli (hearing protector, dynamic speech level and simultaneous speech and noise) were grouped together as they were all wideband (bandwidth up to 24 kHz approximately) and to be presented binaurally. Two compilations of the high-level noises were produced, one for the FRL-processed stimuli to be presented monaurally and one for the FRL-processed stimuli to be presented binaurally.

It was unnecessary to include the SRL-processed stimuli as these were limited to a level at least as low as the FRL-processed stimuli. Within each of the compilations, the noises were separated by silence so that the gain reduction for a given noise was not influenced by the gain reduction for the previous noise within the sequence. As a result of this approach, the limited sound levels of the noises in the compilations matched the limited sound levels of the noises in the training session and the experiments.

The subjects could therefore use these compilations to determine the maximum level they were prepared to listen to in advance of the training session and the experiments. The band-limited monaural and wideband binaural compilations were processed by the fixed-reference limiting scheme using the 17 maximum fixed-limit values to produce 17 fixed-reference limited versions of each compilation. The change in the limiting level per step was 3 dB. The waveforms and spectrograms of the compilation of band-limited monaural stimuli for three levels of limiting is shown in Figure 9-4.

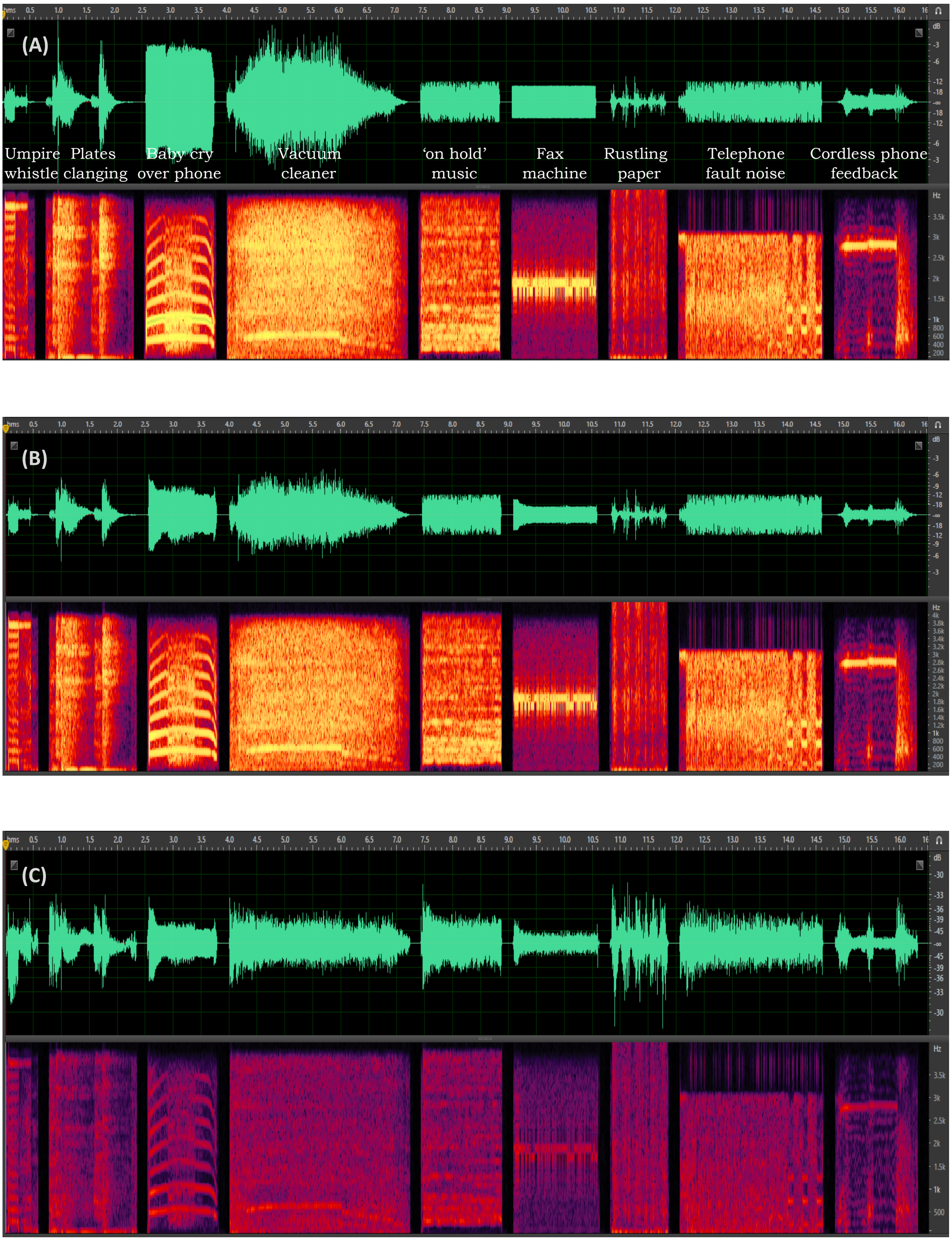


Figure 9-4. *Waveforms and spectrograms (up to 4 kHz) of the compilation of the band-limited monaural noises for three levels of fixed-reference limiting.* ***(A)*** *limiting level 17 (no limiting),* ***(B)*** *limiting level 12 (slight limiting) and* ***(C)*** *limiting level 1 (heavy limiting).*

Note the vertical scale change to the right of the waveform in the bottom figure.

The effect of the maximum limiter setting on the sound level of the band-limited stimuli for monaural presentation (i.e. for Experiments 1 and 3) is shown in Table 9-5. The maximum effect of the limiting was a 39 dB reduction in the 'Baby crying over the telephone' stimulus level. The maximum effect of limiting when averaged across all the band-limited monaural stimuli was 28 dB.

Stimuli	No limiting		Maximum limiting		Limiting	
	Peak dB SPL*	RMS dB SPL*	Peak dB SPL*	RMS dB SPL*	Peak dB	RMS dB
Umpire's whistle	96	81	77	59	19	21
Plates clanging	108	87	75	60	33	27
Baby crying over telephone	106	99	75	60	31	39
Vacuum cleaning	108	93	74	60	33	33
Telephone 'on-hold' music	96	87	77	61	19	26
Fax machine	94	90	69	54	25	36
Rustling paper	98	76	80	61	18	15
Telephone fault noise	96	87	76	60	20	27
Speaker telephone feedback	96	82	75	57	22	26
Maximum	108	99	80	61	33	39
Average	**100**	**87**	**75**	**59**	**25**	**28**

*Diffuse field equivalent

Table 9-5. *The effect of the maximum limiting setting on the sound level of the band-limited stimuli for monaural presentation.*

The waveforms and spectrograms of the compilation of wideband binaural stimuli for three levels of limiting is shown in Figure 9-5.

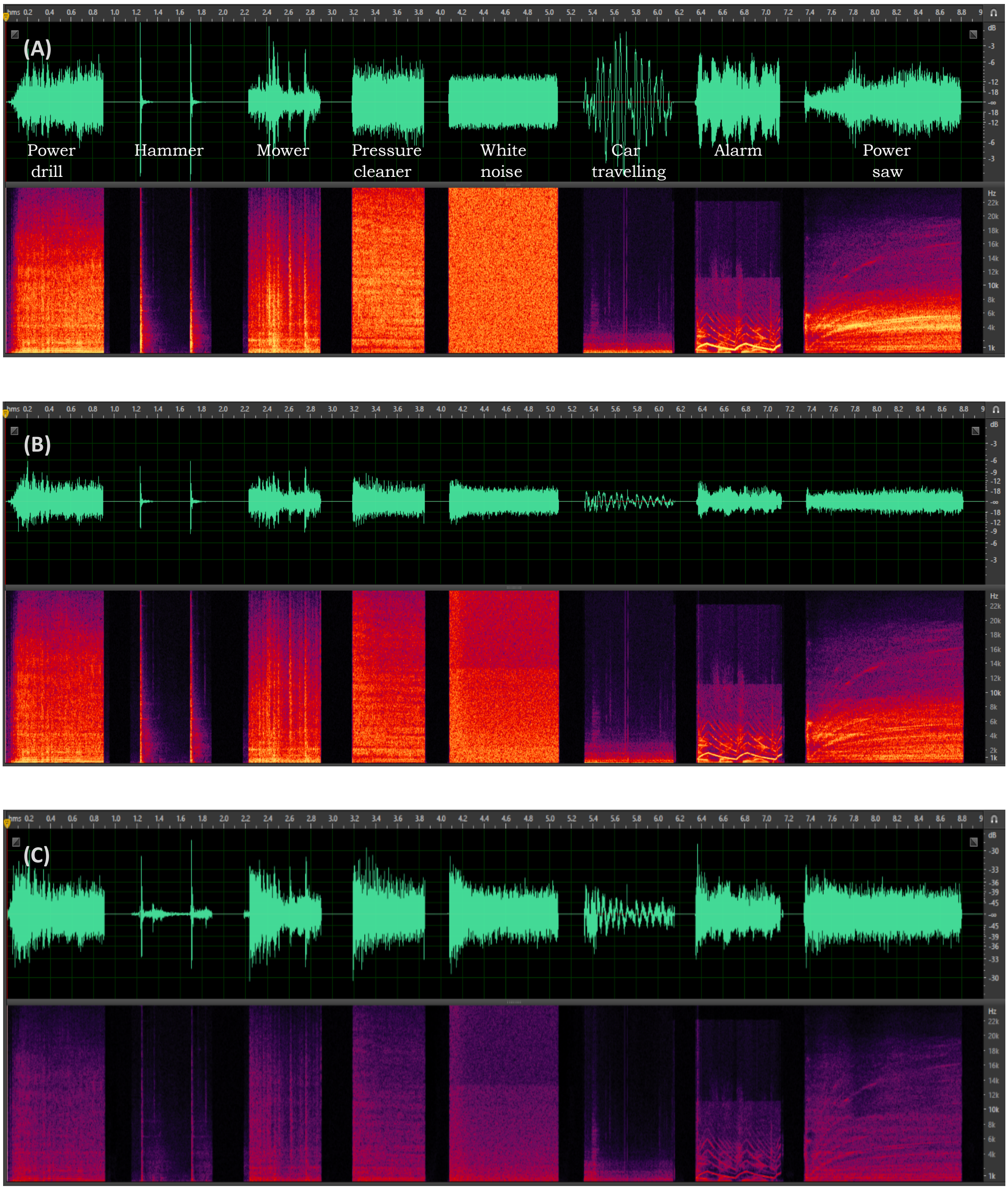


Figure 9-5. *Waveforms and spectrograms (up to 24 kHz) of the compilation of the wideband binaural noise stimuli for three levels of fixed-reference limiting.* ***(A)*** *limiting level 17 (no limiting),* ***(B)*** *limiting level 9 (medium limiting) and* ***(C)*** *limiting level 1 (heavy limiting).*

Note the vertical scale change to the right of the waveform in the bottom figure.

The effect of the maximum limiter setting on the sound level of the wideband stimuli for binaural presentation (i.e. for Experiments 2, 4 and 5) is shown in Table 9-6. The maximum effect of the limiting was a 39 dB reduction in the 'Inside a travelling car' stimulus level. The maximum effect of limiting when averaged across all the wideband binaural stimuli was 30 dB.

Stimuli	No limiting		Maximum limiting		Limiting	
	Peak dB SPL*	RMS dB SPL*	Peak dB SPL*	RMS dB SPL*	Peak dB	RMS dB
Power drill	105	90	78	62	26	27
Hammer	108	83	79	52	29	31
Lawn mower	108	87	79	62	30	25
Pressure cleaner	103	90	79	62	25	27
Burst of white noise	99	90	78	62	22	28
Inside a travelling car	107	98	74	59	33	39
Public address and alarm	104	96	79	62	25	34
Circular power saw	104	89	77	62	27	27
Maximum	108	98	79	62	33	39
Average	**105**	**90**	**78**	**60**	**27**	**30**

*Diffuse field equivalent

Table 9-6. *The effect of the maximum limiting setting on the sound level of the wideband stimuli for binaural presentation.*

More specifically, the effect of the limit levels on the 1-second burst of white noise at a level of 90 dB SPL is shown in Table 9-7. Note that the limit level did not have an effect on the RMS level of this noise above a setting of 13. Its effect increased as the setting was reduced below 13. For a setting of 6 or below, heavy limiting occurred: for each 3 dB step in the limit level, the output level reduced by 3 dB.

The fixed limiter at a setting of 1 reduced the maximum RMS level of the band-limited monaural stimuli to 61 dB SPL and the wideband binaural stimuli to 62 dB SPL and the maximum peaks to 80 dB SPL and 79 dB SPL respectively. This was a more than adequate range to cover the maximum sound level preferences of normal hearers.

Limit setting	dB SPL diffuse field equivalent		Limiting	
#	Peak	RMS	Peak	RMS
17	99	90	0	0
16	99	90	0	0
15	99	90	0	0
14	99	90	0	0
13	99	89	0	1
12	99	88	0	2
11	98	87	1	3
10	98	85	1	5
9	97	84	2	6
8	96	82	4	8
7	94	80	5	10
6	92	77	7	13
5	89	74	10	16
4	87	71	13	19
3	84	68	16	22
2	81	65	19	25
1	78	62	22	28

Table 9-7. *The effect of the fixed limiting value on the sound level of the 1-second burst of white noise at an unprocessed level of 90 dB SPL*

9.2.8 Headphone equalisation

The stimuli were presented to the subjects from Beyerdynamic DT 990 PRO headphones. The equivalent diffuse-field response of these headphones is largely flat. When they were used in the assessment of the SRL MKI scheme (see Chapter 7) no correction for their frequency response was applied as there was no fixed-reference limiting involved and the deviation from the standardised equivalent diffuse-field response, as defined in the ITU-T Rec P.58,[125] was of minor consequence to the spectral balance of the material. However, in the evaluation of SRL MKII, fixed-reference limiting with prescribed frequency-dependent limits referenced to the equivalent diffuse-field was involved. Furthermore, Experiment 4 used precisely recorded stimuli, meaning the results would have greater validity if precise presentation of speech and noise occurred. Hence all the stimuli were equalised to compensate for the deviation in the headphone response from the standardised diffuse-field to real-ear response. The equivalent diffuse-field equalised response of the headphones aligned with the ITU Recommendation ITU-R BS.1284-1 for 'reference monitor headphones'.[183] The frequency response of the DT 990 PRO headphones (average of 4 devices) at the real-ear, the diffuse-field to real-ear response for a Head and Torso Simulator (HATS) as defined in ITU-T Rec P.58,[125] and equalisation to match the headphone response to the diffuse-field to real-ear response is shown in Figure 9-6.

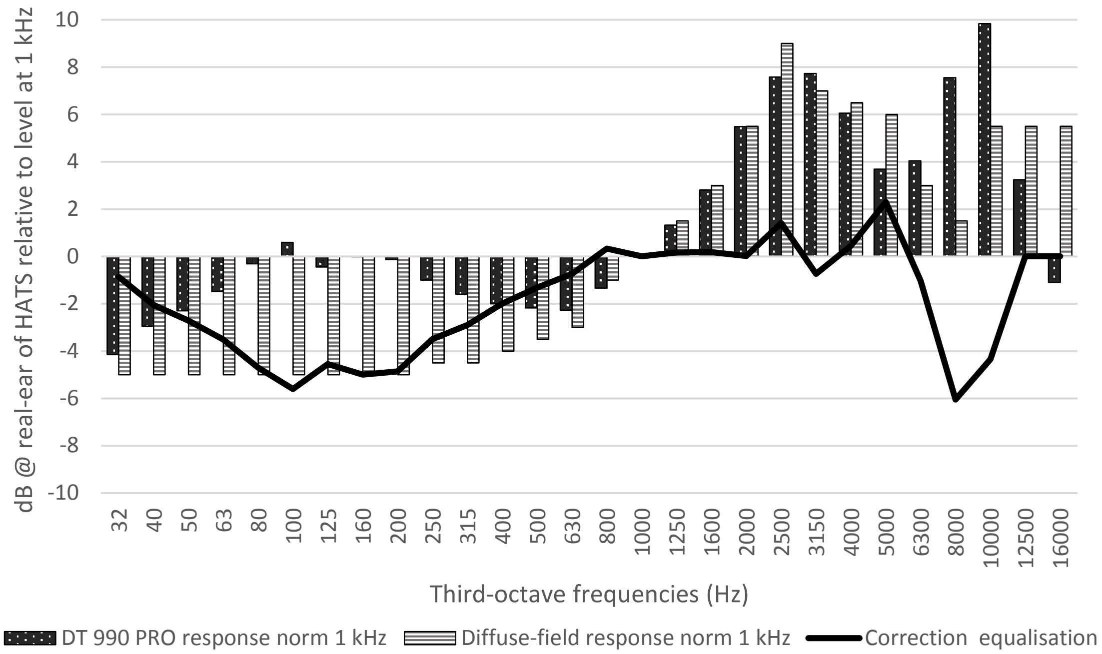


Figure 9-6. *Beyerdynamic DT 990 PRO headphone response, ITU-T Rec P.58 diffuse-field to real-ear response and equalisation to match.*

Due to the unreliability of acoustic measurements of headphones on HATS at frequencies above 10 kHz, and the lack of a specified response for the diffuse-field to real-ear for HATS above 10 kHz, no compensation at these frequencies was made. The equalisation was applied to all the stimuli using a 1,200 tap linear-phase finite-impulse-response (FIR) filter. At the stimuli sampling rate of 48 kHz, this filter had an impulse response duration of 25 ms which enabled it to equalise the response down to a frequency of approximately 40 Hz with good accuracy.

9.2.9 Subjects

There were 16 subjects, 7 females and 9 males. All were normal-hearing adults (normal-hearing criteria: hearing thresholds ≤ 20 dB HL at standard audiometric frequencies) and either worked or studied at the Australian Hearing Hub, Macquarie University, Sydney, Australia. All the subjects were involved in hearing research and many could be considered to be experienced critical listeners. None of the subjects were directly involved in this research.

9.2.10 Experiment overview

All five experiments were conducted in sound-proof booths at the National Acoustic Laboratories located within the Australian Hearing Hub, Macquarie University, Sydney, Australia. The experiments involved the subjects listening to sounds presented by a computer via headphones. The computer was under the control of a program written in MATLAB.[146] The subjects received visual instructions on the computer screen and responded using a computer mouse. The experimenter guided them through the training phase and they were then left to do the experiments alone. The tasks, presentation sequences and on-screen information for triggering and responding to the sounds was pseudo-randomised using counter-balancing based on Graeco-Latin squares. The experiments were double-blind and included a hidden reference, defined by the International Telecommunications Union as a 'reference not identified to the test subject'.[159]

9.2.11 Experiment setups and calibration

There were two testing setups, each in a separate sound-proof booth. Both setups used Beyerdynamic DT 990 PRO headphones to present the sounds to the subjects. The hardware in the two setups differed in that one setup used a personal computer and RME Fireface UC sound device to drive the headphones while the other used a Hewlett Packard HP Elitebook 8540p laptop computer with an in-built IDT High Definition CODEC sound device to drive the headphones.

Calibration of the acoustic level was performed by measuring the acoustic level produced by the headphones using a Brüel & Kjær Type 4128 Head and Torso Simulator (HATS) connected to a pair of Brüel & Kjær Type 2610 Measurement Amplifiers driving a Stanford Research Systems SR785 Dynamic Signal Analyser. The test system was calibrated using a Brüel & Kjær Type 4231 calibrator with a current calibration certificate. Transformation of the signal measured at the real-ear to an equivalent diffuse-field measurement was performed by a filter (Behringer Ultra-Curve DSP 8000 24-bit Dual-DSP Mainframe Digital EQ) inserted into the external filter option of the Brüel & Kjær Type 2610 Measurement Amplifiers.

Sound files containing a 1 kHz tone (at -18 dBr) and the speech (stimulus 4 from Experiment 4 at -33 dBr prior to equalisation) were replayed. The level of the tone was expected to be 90 dB SPL $_{\text{diffuse-field equivalent}}$. The long-term level of the speech was expected to be 75 dB SPL $_{\text{diffuse-field equivalent}}$. These levels were confirmed for both setups.

9.2.12 Subject treatment

All the subjects were volunteers. After reading an information sheet about the experiment and agreeing to participate, they signed a consent form. A copy of this information and consent form is shown in Appendix A. The subjects sat on a chair in front of a desk on which there was a computer screen with a keyboard and mouse or laptop and mouse. They responded using the computer's mouse.

The experimenter was present through the first two stages of the experiment session. This involved the subjects setting their maximum sound levels for band-limited monaural and wideband binaural presentation of sound and a training session in which the subjects performed tasks identical to those required of them in Experiments 1 to 4. The difference in what was required from them for Experiment 5 was verbally explained to them at the end of the training session. After commencing the formal experiments, and being observed to be confident in what they were doing, they were left to complete all five experiments alone. They could take as much time as they liked to make their assessment of the material presented to them by the computer and could stop and rest at any time. The time to set the maximum sound levels, perform the training and the five experiments varied between subjects from about 35 minutes to about 70 minutes.

9.2.13 Presentation sequence and main screen

The order of the five experiments, the order of the four stimuli within each experiment and the allocation of the processed versions of the stimuli to presentation buttons / rating controls were all counter-balanced based on Graeco-Latin squares. All subjects received all the stimuli in all the conditions but all with different counter-balanced sequencing and allocation to presentation buttons / rating controls. The only noncounter-balanced aspect to the presentation sequences was that the speech quality rating section in each experiment always followed the loudness rating section. Furthermore, all odd-numbered subjects received the monaural presented stimuli in their left ear, while all even-numbered subjects received the monaural presented stimuli in their right ear. This was achieved by asking the subjects to put the side of the headphones with the cable attached to a particular ear at the beginning of the experiment session.

The computer program dictated exactly what the subjects must do in order to move on to the next stage. The subjects had no choice but to follow the pre-defined sequence the computer presented. The subjects had to listen to and respond to all the stimuli presented. If any stimulus was not listened to or responded to the program would not advance. The main experiment screen with the initial instructions for the subjects is shown in Figure 9-7.

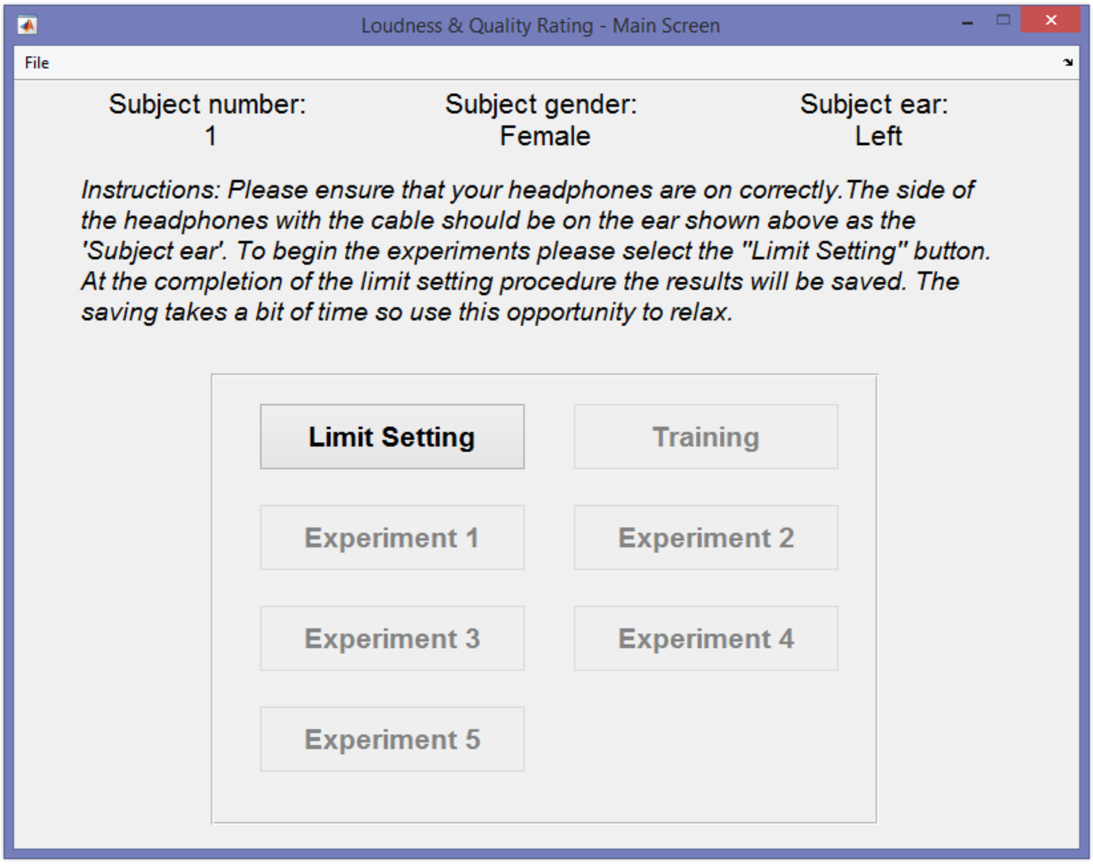


Figure 9-7. *Main screen with initial instructions.*

After reading the instructions and fitting the headphones accordingly, the subjects selected the 'Limit Setting' button. After the subjects performed the limiting setting task, the screen shown in Figure 9-8 was displayed, providing them with general instructions for the experiments. After completing the training, subjects performed each of the five experiments shown.

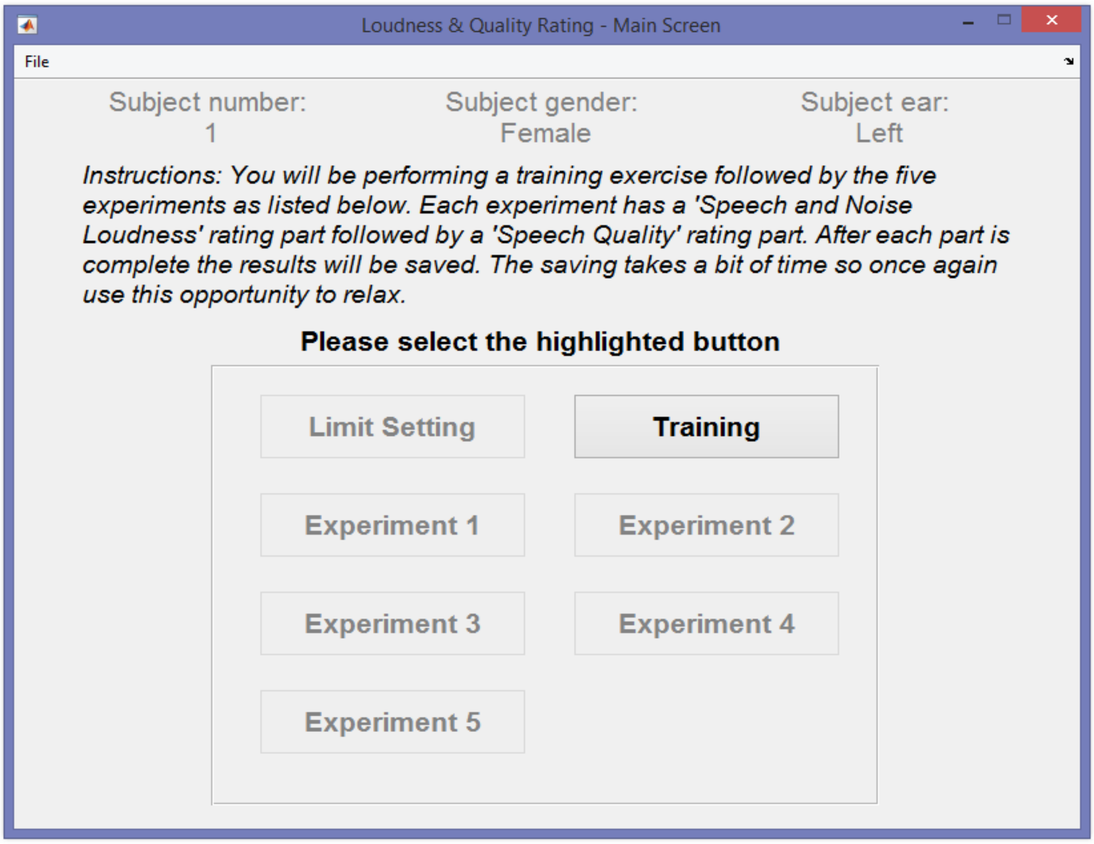


Figure 9-8. *Main screen with general instructions.*

9.2.14 Maximum sound level setting

The subjects performed two maximum loudness setting tasks, one for the band-limited monaurally-presented noises that they would be presented with in Experiments 1 and 3, and one for the wideband binaurally-presented noises that they would be presented with in Experiments 2, 4 and 5. The monaural maximum loudness setting screen is shown in Figure 9-9. The binaural maximum loudness setting screen was identical to this, except for its title, and the procedure was also identical. The subjects were initially presented with the noise compilations (as described in Section 9.2.7) at the softest level, i.e. maximum limiting. The initial maximum sound levels at this setting were:

- Monaural: 61 dB SPL RMS and 80 dB SPL peak
- Binaural: 62 dB SPL RMS and 79 dB SPL peak

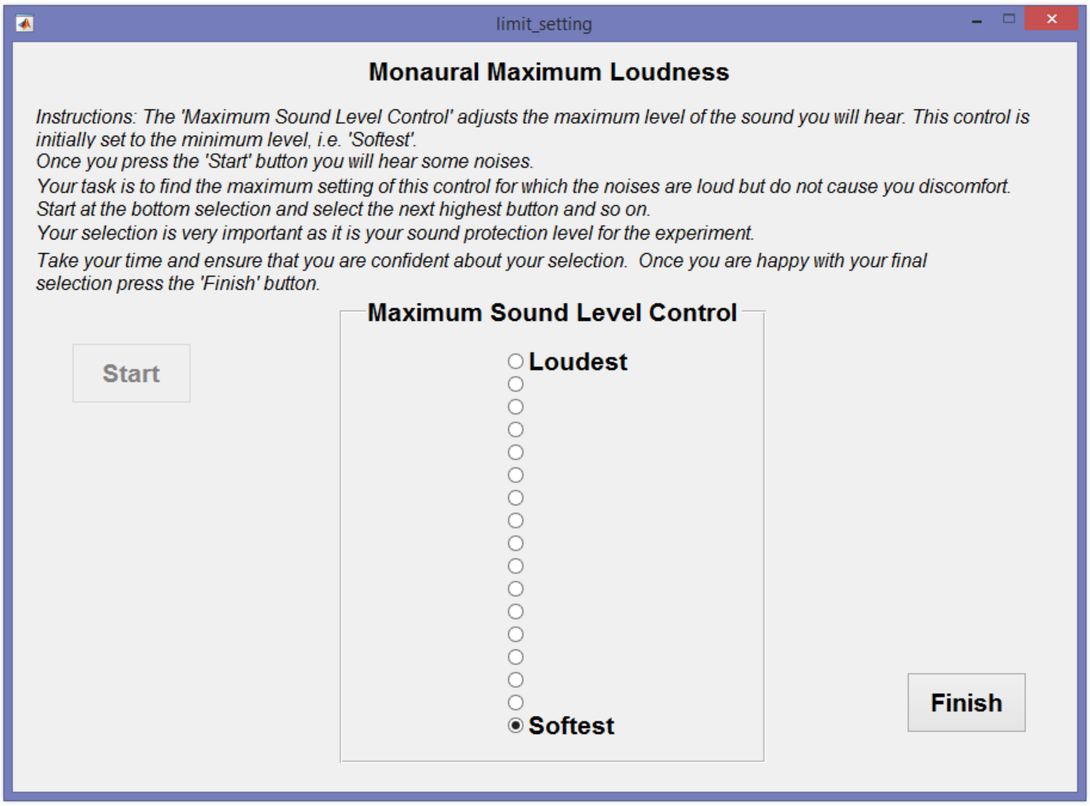


Figure 9-9. *Monaural maximum loudness setting screen.*

The subjects increased the maximum loudness by selecting the next highest button until they found *the maximum setting of this control for which the noises are loud but do not cause you discomfort*. Their selection was recorded for both the monaural and binaural presentations. For the training and the experiments, only pre-processed stimuli with the subjects' selected maximum loudness were presented to them.

9.2.15 Speech and noise loudness rating

In all the experiments, the subjects were required to rate the loudness of the speech and noise. The interactive speech and noise loudness rating screen displayed to the subjects is shown in Figure 9-10. It used the same seven-level loudness-rating scale used in the evaluation of SRL MKI with the loudness category labels defined by Cox et al.[157] The instructions and control were largely the same as those provided for the evaluation of the SRL MKI scheme. The notable exception was that there were three stimuli to be rated rather than two (i.e. FRL-processed, SRL-processed and the reference stimuli), except for Experiment 5 which only had two.

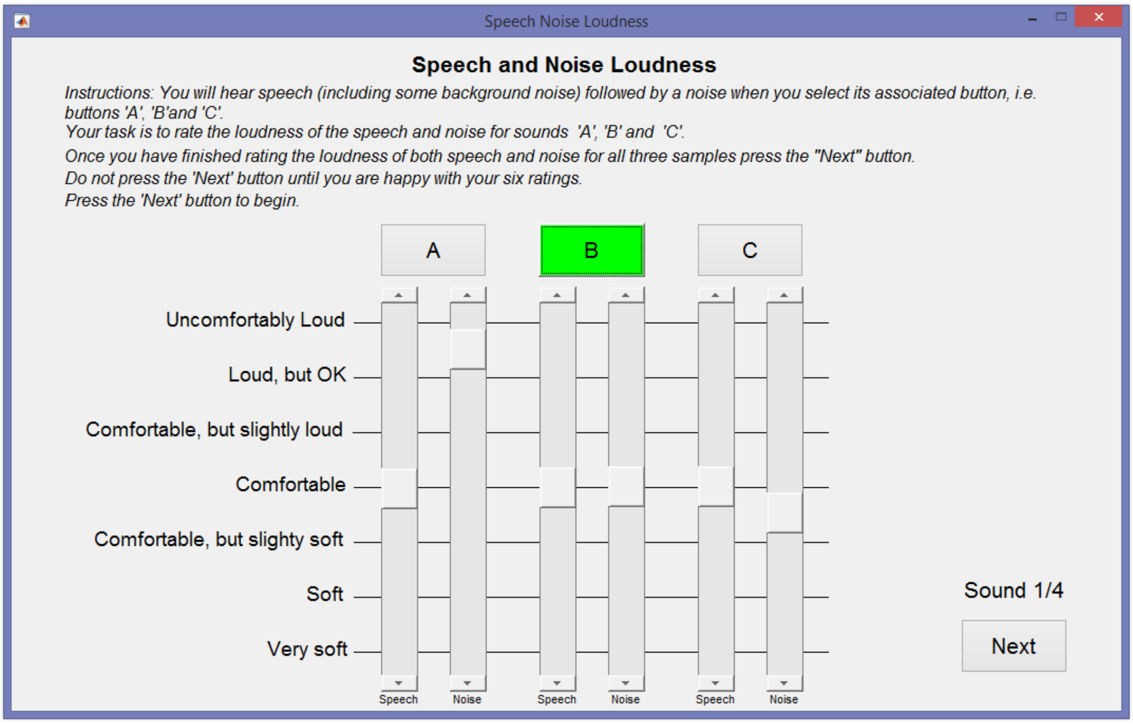


Figure 9-10. *Speech and noise loudness rating screen.*

The subjects were provided with three buttons on the screen that played three stimuli, labelled 'A', 'B' and 'C'. Below the respective buttons was a pair of slider controls, one for speech loudness rating and the other for noise loudness rating, which the subjects adjusted using a computer mouse. They could listen to the stimuli as many times as they needed to in order to be satisfied with their ratings before moving on to the next stimulus or test. The program would not allow them to advance until all the stimuli had been played and rated. The FRL-processed, SRL-processed and reference stimuli were pseudo-randomly assigned to the three buttons.

9.2.16 Speech quality rating – relative

In Experiments 1 to 4, the subjects were required to rate the quality of the speech relative to the reference speech.

The speech quality rating relative to the reference had a range from -3 to +3 using the category labels recommended by the International Telecommunications Union, Recommendation ITU-R BS.1284-1, 'General methods for the subjective assessment of sound quality',[183] as follows:

3	Much better
2	Better
1	Slightly better
0	The same
-1	Slightly worse
-2	Worse
-3	Much worse

Table 9-8. *Comparative speech quality rating labels.*

The interactive speech quality rating screen displayed to the subjects is shown in Figure 9-11.

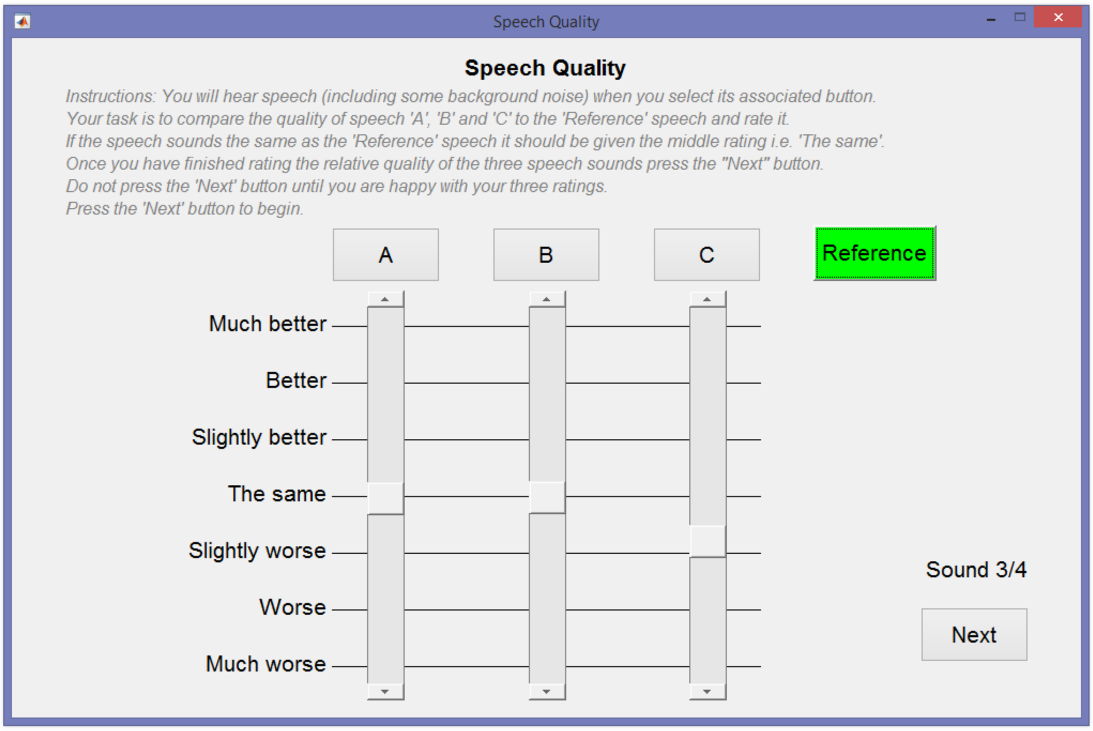


Figure 9-11. *Relative speech quality rating screen.*

The instructions and controls were similar to those provided for the evaluation of the SRL MKI scheme. The notable exceptions were the labels and the number of stimuli to be rated, three rather than two (i.e. FRL-processed, SRL-processed and the reference stimuli). The subjects were provided with four buttons on the screen to play the reference (unprocessed) stimulus and stimuli A, B and C. Below each A, B and C button was a slider control to rate the speech quality of stimuli which the subjects adjusted using a computer mouse. The subjects were asked to compare each of the A, B and C stimuli with the reference and rate the comparative speech quality. Only the speech portion of the recording was played. They could listen to the stimuli as many times as they needed in order to be satisfied with their ratings before moving on to the next stimulus or test. The FRL-processed, SRL-processed and reference stimuli were pseudo-randomly assigned to the three buttons.

Experiment 4 differed slightly from Experiments 1, 2 and 3 in that the subjects rated the speech quality in two sections separately. The first section was the first half of the sentence, '*When the sunlight strikes raindrops in the air*' and the second section was the second half of the sentence, '*they act as a prism and form a rainbow*'. The reason for this separation was to investigate the perceived quality of the level transition portion of the sentence separately from the latter part of the sentence when the level had been stable for some time.

9.2.17 Speech quality rating – absolute

The stimuli for Experiment 5 comprised noise simultaneously presented with speech. As a result, the speech did not exist in an isolated form where its quality alone could be rated in a comparative test. Presenting the unprocessed noise along with the unprocessed speech was not an option as this would mean presenting the subjects with high-level sounds (i.e. the noises) at levels likely to be above the maximum sound level they had set at the beginning of the testing. As the reference stimuli could not be presented, absolute quality rating rather than relative quality rating was employed for Experiment 5.

The interactive absolute speech quality rating screen displayed to the subjects is shown in Figure 9-12. The instructions and controls were similar to those provided for the speech quality evaluation of the SRL MKI scheme. The notable exceptions were that the subjects were instructed to make absolute rather than relative quality ratings and there were no reference stimuli for comparison and hence no hidden reference. Only the FRL-processed and SRL-processed stimuli were presented.

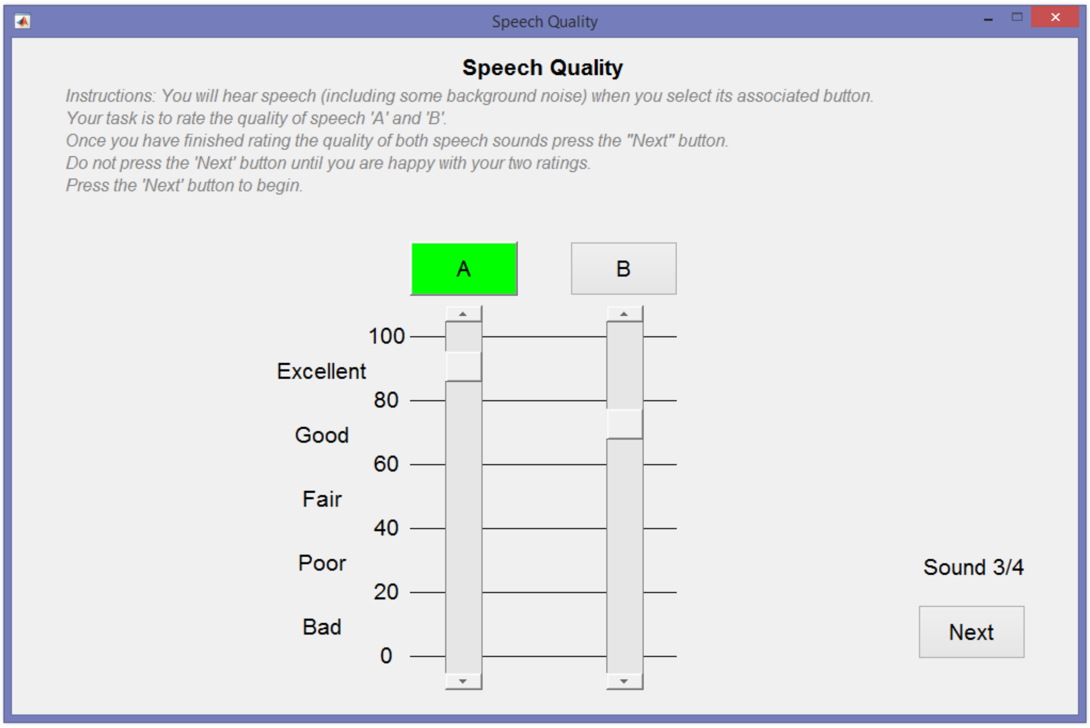


Figure 9-12. *Absolute speech quality rating screen.*

9.3 Results and discussion

The results from the limiting setting procedure and the five experiments were statistically analysed using the Statistica™ software package from Dell™ and the Excel® software package from Microsoft®. Mean ratings across subjects and 95% confidence intervals of the means were produced for the results of all five experiments. Further statistical analysis was performed using the analysis of variance (ANOVA) method.[161] The raw data from all the experiments including basic statistical analysis is given in Appendix D.

9.3.1 Maximum level

The fixed-reference limit settings of the sixteen subjects for band-limited (BL) monaural presentation and wideband (WB) binaural presentation are shown in Table 9-9 and a statistical summary of the data is shown in Table 9-10.

Gender	Subject	Fixed-reference limit setting selected for: BL monaural presentation	WB binaural presentation
F	1	14	7
F	2	6	6
F	4	14	9
F	5	13	9
F	6	14	5
F	7	12	6
F	8	12	13
M	3	11	8
M	9	11	13
M	10	8	8
M	11	12	12
M	12	13	12
M	13	6	4
M	14	17	12
M	15	14	13
M	16	13	12

Table 9-9. *The fixed-reference limit settings selected by the 16 subjects for band-limited (BL) monaural and wideband (WB) binaural presentations.*

Statistic	Fixed-reference limit setting selected for: BL monaural presentation	WB binaural presentation
Maximum	17	13
Mean	12	9
Median	13	9
Minimum	6	4

Table 9-10. *Summary of the fixed-reference limit settings selected by the 16 subjects for band-limited (BL) monaural and wideband (WB) binaural presentations.*

Figure 9-13 shows the distribution of the subjects' selected fixed-reference limit settings. For the band-limited monaural presentation, 13 of the 16 subjects selected limit settings of 11 or higher.

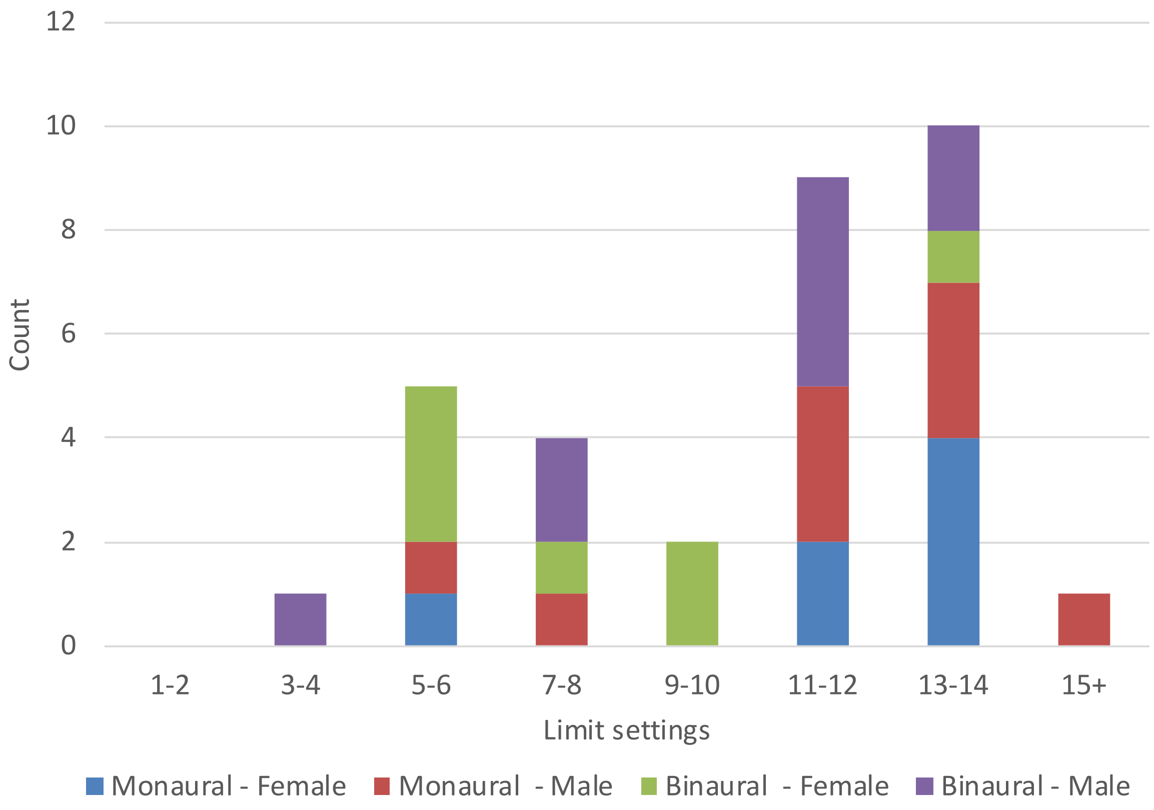


Figure 9-13. *Distribution of subject-selected fixed-reference limit settings.*

However, for the wideband binaural presentation only 7 of the 16 subjects selected limit settings of 11 or higher. Several factors may explain this difference. First, binaural loudness summation would have occurred due to the sound being presented to both ears. Second, spectral loudness summation would have occurred due to the stimuli being of considerably wider bandwidth (24 kHz instead of 4 kHz). Third, some of the binaurally presented sounds, such as the power drill had a modulated but unrelenting temporal intensity leading to greater loudness integration. The peaks of the modulation signal, such as those of the drill, are known to drive the loudness sensation to higher levels.[48] These differences are discussed further in Section 9.3.1.3 and the effects were discussed more generally in Chapter 2.

There was a difference in the limit setting selection of the genders for the wideband binaural presentation, 6 out of the 7 female subjects selected a limit setting of 10 or less while only 3 out of the 9 males did the same. It would appear that the band-limited monaurally-presented sound produced a loudness sensation of a magnitude that required only a small amount of limiting to avoid loudness discomfort in the vast majority of subjects with only 2 females and 1 male selecting a limit setting of 10 or less. However, when the sounds were wideband noises presented binaurally, the increase in loudness sensation caused discomfort to those with lower loudness discomfort levels, and more of these subjects were female in this experiment. This was not unexpected given the findings of Thomas and Jones (1982), who reported that their female subjects had on average a 13 dB lower uncomfortable loudness level than their male subjects.[67]

There may also be an order effect leading to a greater perception of loudness for the wideband binaurally presented sounds. All the subjects were presented with sounds monaurally then binaurally, and it could be that some subjects had simply had enough of the loud noises by the time these were presented binaurally. In retrospect, it might have been better to have randomised the order of the monaural and binaural presentations to avoid this potential effect.

Two of the sixteen subjects, however, selected a slightly higher limiting level for the wideband binaurally-presented sounds than the band-limited monaurally-presented sounds. One of these subjects was female (subject 8) and her selection of limiting levels only differed by one limiting level step, i.e. from 12 to 13. The other was male (subject 9) and his selection of limiting level differed by two limiting level steps, i.e. from 11 to 13. It is not surprising to find variation in loudness perception between individuals and within an individual as discussed in Chapter 2. It would appear that these two subjects, did not experience the spectral and binaural loudness summation effects experienced by the average subject for the noises presented. Subject 9 never gave a rating of uncomfortably loud in any of the testing that followed using these settings and subject 8 only once gave a rating of uncomfortably loud in tests that followed using these settings. This would indicate that, despite selecting higher limits for wideband binaurally-presented stimuli, they had selected limits they were not uncomfortable with.

In order for these limit settings to have a physical meaning the sound levels that resulted from their application were assessed. The results are presented and discussed in the following subsections 9.3.1.1. to 9.3.1.5.

9.3.1.1 Analysis of the band-limited monaural-compilation maximum level

Table 9-11 shows the resulting peak and RMS sound levels in dB SPL $_{\text{diffuse-field equivalent}}$ for the band-limited monaurally-presented stimuli after being limited using the average fixed-reference limit setting (i.e.12) of the subjects for this stimulus compilation. Also shown are the maximum and average values of these peak and RMS sound levels.

	Sound level (dB SPL $_{\text{diffuse field equivalent}}$)	
Stimuli	**Peak**	**RMS**
Umpire's whistle	96	80
Plates clanging	103	84
Baby crying over telephone	102	91
Vacuum cleaning	103	89
Telephone 'on-hold' music	96	87
Fax machine	94	85
Rustling paper	98	76
Telephone fault noise	96	87
Speaker telephone feedback	96	80
Maximum	**103**	**91**
Average	**98**	**84**

Table 9-11. *The sound levels resulting from the application of the fixed-reference limiter, FRL to the band-limited monaurally-presented stimuli with the average limit setting (i.e.12) of the subjects for these stimuli.*

For band-limited monaural presentation, the maximum peak sound level the average subject was exposed to was 103 dB SPL $_{\text{diffuse-field equivalent}}$ and the maximum RMS sound level they were exposed to was 91 dB SPL $_{\text{diffuse-field equivalent}}$. The corresponding average peak and RMS sound levels were 98 dB SPL $_{\text{diffuse-field equivalent}}$ and 84 dB SPL $_{\text{diffuse-field equivalent}}$. In contrast to these limited values were the non-limited values, and these are compared in Table 9-12.

Sound level (dB SPL diffuse field equivalent)			
Statistic	**No limiting**	**Limit setting = 12**	**Limiting (dB)**
Maximum peak	108	103	5
Average peak	100	98	2
Maximum RMS	99	91	8
Average RMS	87	84	3

Table 9-12. *Change in maximum sound levels of the band-limited monaurally-presented stimuli for the subjects with the average limit setting (i.e. 12).*

The differences between the non-limited values and those with a limit setting of 12 are shown in the right-hand column of Table 9-12 and labelled 'Limiting'. The most perceptually relevant of these is arguably the change in the maximum RMS value as the subjects were most likely to set the limit to protect themselves in the worst case rather than the average case and the RMS level correlates more closely with the perception of loudness than does the peak level. This is due to loudness integration over time,[25] as discussed in Chapter 2. The change in the maximum RMS is therefore the preferred measure, and for this band-limited monaural case it was 8 dB. The effect of the subjects' average limit setting (i.e.12) on the waveform and spectrogram of the band-limited monaurally-presented noises is shown in the middle plot, (B) of Figure 9-4 in Section 9.2.7. The limiting effect on the higher level sounds, such as the baby crying and the vacuum cleaning, is evident as a reduction in the temporal envelope and the intensity of the spectra (third and fourth stimuli from the left).

9.3.1.2 Analysis of the wideband binaural-compilation maximum level

Table 9-13 shows the resulting peak and RMS sound levels in dB SPL $_{\text{diffuse-field equivalent}}$ for the wideband binaurally-presented stimuli with the average fixed-reference limit setting (i.e. 9) of the subjects for this stimulus compilation. Also shown are the maximum and average values of these peak and RMS sound levels.

Stimuli	Sound level (dB SPL $_{\text{diffuse field equivalent}}$) Peak	RMS
Power drill	102	86
Hammer	101	74
Lawn mower	101	84
Pressure cleaner	99	83
Burst of white noise	97	84
Inside a travelling car	91	82
PA announcement and alarm	96	84
Circular power saw	94	81
Maximum	**102**	**86**
Average	**98**	**82**

Table 9-13. *The sound levels resulting from the application of the fixed-reference limiter, FRL to the wideband binaurally-presented stimuli with the average limit setting of the subjects for these stimuli (i.e. 9).*

For wideband binaural presentation, the maximum peak sound level the average subject was exposed to was 102 dB SPL $_{\text{diffuse-field equivalent}}$ and the maximum RMS sound level they were exposed to was 86 dB SPL $_{\text{diffuse-field equivalent}}$. The corresponding average peak and RMS sound levels were 98 dB SPL $_{\text{diffuse-field equivalent}}$ and 82 dB SPL $_{\text{diffuse-field equivalent}}$. In contrast to these limited values were the non-limited values, and these are compared in Table 9-14. The difference between the non-limited values and those with a limit setting of 9 are shown in the right hand column of Table 9-14. As previously discussed, the most perceptually relevant of these is arguably the change in the maximum RMS value, which for these stimuli was 13 dB. The effect of the subjects' average limit setting (i.e. 9) on the waveform and spectrogram of the wideband binaurally presented noises is shown in the middle plot, (B) of Figure 9-5 in Section 9.2.7.

	Sound level (dB SPL $_{\text{diffuse field equivalent}}$)		
Statistic	**No limiting**	**Limit setting = 9**	**Limiting (dB)**
Maximum peak	108	102	6
Average peak	105	98	7
Maximum RMS	99	86	13
Average RMS	90	82	8

Table 9-14. *Change in maximum sound levels of the wideband binaurally-presented stimuli for the subjects' average limit setting (9).*

9.3.1.3 Comparison of the band-limited monaural and wideband binaural maximum RMS levels after fixed-reference limiting

A comparison of the average subjects' band-limited monaural and wideband binaural maximum-RMS noise levels, after user selected fixed-reference limiting, revealed a 5 dB difference (91 dB SPL compared to 86 dB SPL). The two main factors that are likely to have contributed to this perceptual difference are binaural and spectral loudness summation. Interaction between these factors may also have occurred. As discussed in Chapter 2, the degree of binaural and spectral loudness summation is not fully understood, nor is the interaction between them, nor the effect of temporal integration on, at least, spectral loudness summation if not both. The 5 dB difference is below the 10 dB effect that alone would result from a complete doubling of loudness as predicted by the binaural summation model of Moore et al. (1997).[24] The 5 dB difference is more aligned with the 5 dB effect predicted to result from binaural loudness summation at high sound levels by Fastl and Zwicker (2006).[25] However, an increase in spectral summation due to the increased stimulus bandwidth also needs to be considered which indicates that the binaural effect alone must be less than 5 dB. The spectral summation, in turn, depends on the stimulus, its level, and temporal characteristic as well as its spectral characteristic as discussed in Chapter 2. It is likely that all these characteristics as well as the binaural characteristics affected the difference in the perception found. An investigation into interactions among these effects, including the binaural effects, and the degree of these interactions, is beyond the scope of this thesis but would make an interesting study.

9.3.1.4 Analysis of the selected wideband binaural maximum levels

As discussed in Chapter 2, there can be a wide range in selected maximum sound levels between individuals. Table 9-15 shows the range of peak and RMS sound levels that resulted from 1 second bursts of white noise over the range of fixed-limit settings available to the subjects. This was data was previously shown in Table 9-7 but added to it is the subjects' selected maximum, average and minimum limit settings. The subjects in this evaluation displayed a considerable range in their selection, with an 18 dB difference in the RMS level of this limited stimulus.

Limit setting	SPL dB diffuse field equivalent		Limiting		Subject
#	peak	RMS	peak	RMS	selection
17	99	90	0	0	
16	99	90	0	0	
15	99	90	0	0	
14	99	90	0	0	
13	99	89	0	**1**	**Maximum**
12	99	88	0	2	
11	98	87	1	3	
10	98	85	1	5	
9	97	84	2	**6**	**Average**
8	96	82	4	8	
7	94	80	5	10	
6	92	77	7	13	
5	89	74	10	16	
4	87	71	13	**19**	**Minimum**
3	84	68	16	22	
2	81	65	19	25	
1	78	62	22	28	

Table 9-15. *Wideband binaural maximum sound levels for white noise over the range of fixed-reference limit setting with subject selection data.*

9.3.1.5 Analysis of the noise levels resulting from FRL and SRL processing

Table 9-16 shows the peak and RMS sound levels in dB SPL $_{\text{diffuse-field equivalent}}$ for the noises in Experiments 1 to 5 resulting from FRL and SRL processing with the average fixed-reference limits set by the subjects (i.e. limiter setting 12 for band-limited monaurally-presented stimuli and limiter setting 9 for the wideband binaurally-presented stimuli).

		Sound level (dB SPL $_{\text{diffuse field equivalent}}$)			
		Peak		RMS	
Experiment	Stimuli	FRL	SRL	FRL	SRL
1	Baby crying over telephone	102	91	91	79
	Plates clanging	103	96	84	75
	Vacuum cleaning	103	90	89	78
	Umpire's whistle	96	92	80	78
2	Power drill	102	82	86	69
	Hammer	101	73	74	46
	Lawn mower	101	83	84	70
	Pressure cleaner	100	82	83	67
3	Fax machine	95	88	85	67
	Speaker telephone feedback	96	82	80	65
	Telephone 'on-hold' music	96	82	87	73
	Telephone fault noise	96	82	87	70
4	White noise – speech @ 62 dB SPL	97	80	84	62
	White noise – speech @ 55 dB SPL	97	75	84	59
	White noise – speech @ 68 dB SPL	97	80	84	66
	White noise – speech @ 75 dB SPL	97	86	84	73
5	Inside a travelling car	93	88	82	71
	Hammering (with speech)	101	88	71	57
	Pressure cleaning (with speech)	98	88	80	71
	PA announcement and alarm	96	95	83	81
1 and 3	**Average**	**98**	**88**	**86**	**73**
2, 4 and 5	**Average**	**98**	**83**	**81**	**66**
All	**Average**	**98**	**85**	**83**	**69**

Table 9-16. *Levels of the FRL and SRL processed noises in Experiments 1 to 5 for the average fixed-reference limits set by the subjects.*

The band-limited monaural SRL-processed noises (Experiments 1 and 3) had an average RMS sound level of 73 dB SPL $_{\text{diffuse-field equivalent}}$; this was 3 dB higher than the associated speech level. The wideband binaural SRL-processed noises (Experiments 2, 4 and 5) had an average RMS sound level of 66 dB SPL $_{\text{diffuse-field equivalent}}$; this was 1 dB higher than the associated speech level. In contrast, the corresponding FRL-processed noise levels for the average subject were 86 dB SPL $_{\text{diffuse-field equivalent}}$ and 81 dB SPL $_{\text{diffuse-field equivalent}}$; these exceeded the SRL-processed noise levels by 13 dB and 15 dB respectively and exceeded the associated speech levels by 16 dB for both stimuli types. For the average subject, over all the experiments, the FRL-processed noises were, on average, 14 dB greater in level than the SRL-processed noises.

9.3.2 Experiment 1: Hearing aid application

9.3.2.1 Loudness

The results of the loudness ratings of speech and noise, encountered in hearing aid applications, are shown in Figure 9-14. The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the speech and noise in the conventional FRL-processed, the SRL-processed and the reference conditions for each noise type. The raw data and statistics underlying this plot are documented in Tables D-1 and D-2, in Appendix D.

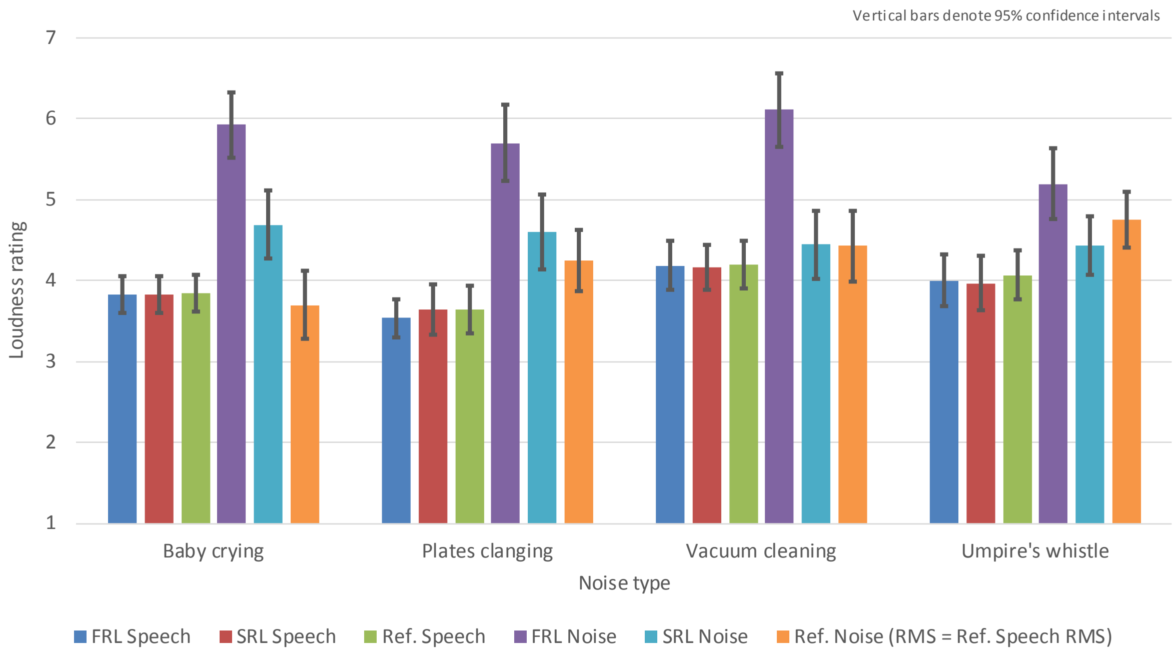


Figure 9-14. *Loudness of speech and noise encountered in hearing aid applications.*

9.3.2.2 Speech loudness

All the speech was rated on average as being within the range of 'comfortable' (category 4). The differences between the loudness ratings of the FRL-processed speech, the SRL-processed speech, and the reference speech, averaged across the four noise types, was less than 0.1 loudness categories (LCs) (all three average speech loudness ratings were 3.9).

A two-way repeated-measures ANOVA was performed on the speech loudness ratings with processing condition and noise type as factors. The main effect of processing condition was found to be statistically significant ($p=0.02$) although the estimated mean differences were less than 0.1 LCs, which is too small to be of any practical significance. Post hoc analysis was performed using the Bonferroni method[162] to compare the speech loudness ratings. This test revealed that the effect of processing condition was significant only for FRL-processed speech compared to the reference speech ($p=0.021$). As mentioned, the mean loudness difference was less than 0.1 LCs and hence was of no practical significance. The SRL-processed speech was within 0.1 LCs of the reference speech and the difference was neither practically nor statistically significant using a Bonferroni test ($p=1$).

9.3.2.3 Noise loudness

In contrast to the speech loudness results, there were highly significant differences in the noise loudness ratings as a result of processing condition. All the FRL-processed noises, with the exception of the umpire's whistle, were rated as being 'loud, but O.K.' (category 6). The FRL-processed umpire's whistle was rated as being 'comfortable, but slightly loud' (loudness rating = 5.2). All the SRL-processed noises were rated as being between 'comfortable' to 'comfortable, but slightly loud' (categories 4 and 5). The difference between the loudness ratings of the FRL-processed noise and SRL-processed noise, averaged across the four noise types, was 1.2 LCs (i.e. 5.7 for FRL-processed noise, compared to 4.5 for SRL-processed noise). A two-way repeated-measures ANOVA was performed on the loudness ratings of the noises with processing condition (FRL and SRL) and noise type as factors. The results showed a highly significant effect of the processing condition ($p<0.001$). The main effect of noise type was also statistically significant ($p=0.021$), as was the interaction between the processing condition and noise type ($p<0.001$). Post hoc Bonferroni tests revealed that the effect of processing condition was highly significant for SRL-processing compared with FRL-processing for every noise ($p<0.001$).

9.3.2.4 SRL-processed speech and noise loudness

The principal aim of the SRL method was to reduce the loudness of non-speech sounds with reference to the loudness of speech sounds. An analysis of the loudness of the SRL-processed noise in relation to the SRL-processed speech was therefore of interest (i.e. how close was the loudness of the SRL-processed noise to that of the SRL-processed speech?). The average loudness rating of the SRL-processed noise was 4.5 and of the SRL-processed speech was 3.9, a difference of 0.6 LCs. For comparison, the average loudness rating of the FRL-processed noise was 5.7 and the FRL-processed speech was 3.9, a difference of 1.8 LCs. The SRL processing had reduced the exceedance by 1.2 LCs.

A two-way repeated-measures ANOVA was performed on the loudness ratings of SRL-processed stimuli with stimulus type (speech or noise) and noise type as factors. The results showed a highly significant effect of the stimulus type ($p<0.001$). The noise type was not significant ($p=0.74$). However, the interaction between the stimulus type and noise type was significant ($p=0.011$). Post hoc Bonferroni tests revealed that the loudness of the SRL-processed noise was not significantly different to the loudness of the SRL-processed speech for the vacuum cleaner ($p=1$) and the umpire's whistle ($p=0.13$). However, there were highly significant differences for the baby crying ($p<0.001$) and plates clanging ($p<0.001$) although the loudness of these sounds were within one LC range of the speech.

9.3.2.5 Speech quality

The results of the quality ratings of speech, encountered in hearing-aid applications, are shown in Figure 9-15. The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the FRL-processed, the SRL-processed and the reference speech for each noise type. The raw data and the basic statistics underlying this plot are documented in Table D-3, in Appendix D.

All the speech was rated as being 'the same' as the reference speech. A two-way, repeated-measures ANOVA was performed on the speech quality ratings with processing condition and noise type as factors. The results showed no significant effect of the processing condition ($p=0.97$), of the noise type ($p=0.83$) or of the interaction between the processing condition and noise type ($p=0.17$).

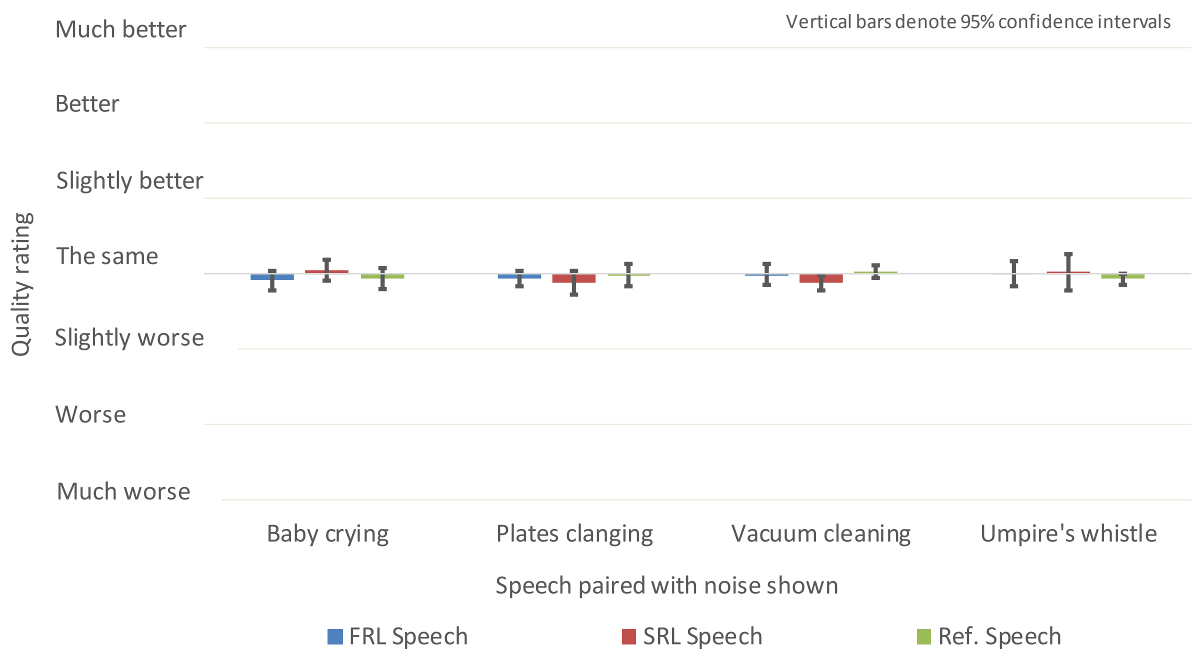


Figure 9-15. *Quality of speech encountered in hearing aid applications.*

9.3.2.6 Summary

Overall, the most striking feature of the loudness ratings was the combination of a 1.2 loudness-category reduction in noise loudness as a result of SRL processing, compared to conventional FRL processing, with no change in speech loudness. This reduction in loudness resulted in no practical or statistically significant change in the speech quality in all cases. Furthermore, the loudness of the noises was only 0.6 LCs greater than the speech, compared to 1.8 LCs for the FRL processing, and comparable to the 0.5 LCs found with the SRL MKI scheme. Based on this data SRL MKII was performing slightly better for hearing-aid application than SRL MKI which produced a small reduction in both speech loudness and quality.

9.3.3 Experiment 2: Hearing protector application

9.3.3.1 Loudness

The results of the loudness ratings of the speech and noise, in the hearing protector application, are shown in Figure 9-16. The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the speech and noise in the conventional FRL-processed, the SRL-processed and the reference conditions for each noise type. The raw data and statistics underlying this plot are documented in Tables D-4 and D-5, in Appendix D.

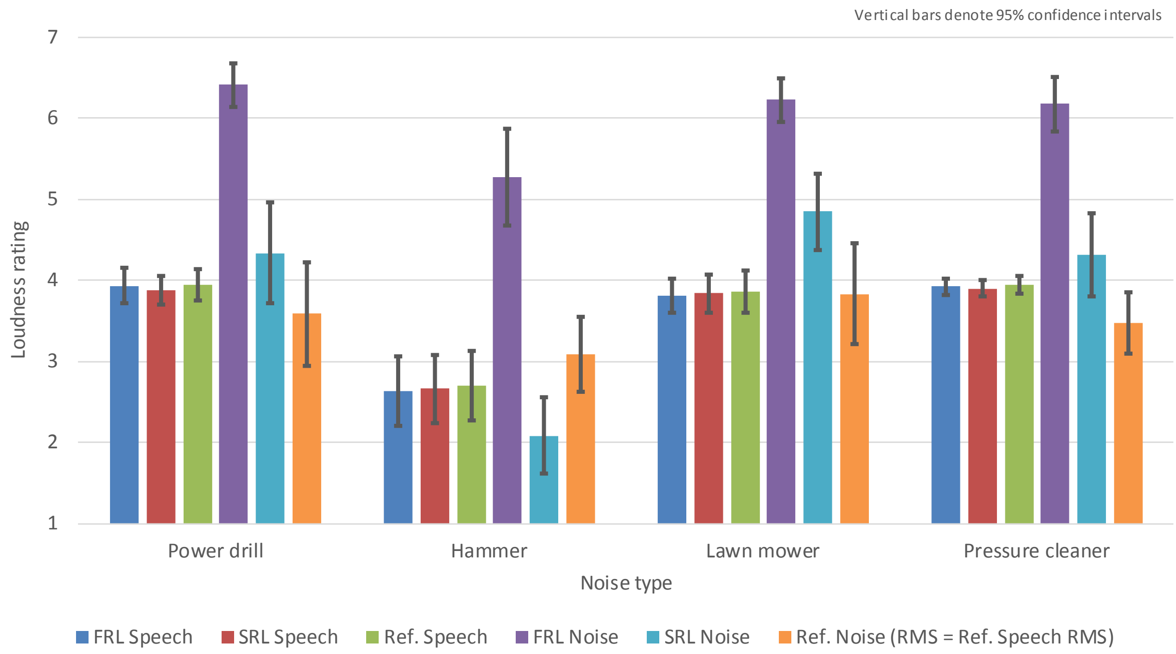


Figure 9-16. *Loudness of speech and noise encountered in hearing protector applications.*

9.3.3.2 Speech loudness

All the speech was rated as being ‘comfortable, but slightly soft’ to ‘comfortable’ (categories 3 and 4). The differences between the loudness ratings of the FRL-processed speech, the SRL-processed speech, and the reference speech, averaged across the four noise types, was less than 0.1 LCs (all three average speech loudness ratings were 3.6).

A two-way repeated-measures ANOVA was performed on the speech loudness ratings with processing condition and noise type as factors. The main effect of processing condition was found to be statistically significant ($p=0.024$) although the estimated mean differences were less than 0.1 LCs, which was too small to be of any practical significance. The effect of noise type was highly significant ($p<0.001$) as the speech associated with the hammering noise was considerably lower in level. The interaction between the processing condition and noise type was not significant ($p=0.38$). The SRL-processed speech was within 0.1 LCs of the reference speech for all noises and the difference was neither practically or statistically significant using a Bonferroni test ($p=1$).

9.3.3.3 Noise loudness

In contrast to the speech loudness results, there were highly significant differences in the noise loudness ratings as a result of processing condition. All the FRL-processed noises, with the exception of the hammer, were rated as being as 'loud, but O.K.' (category 6). The FRL-processed hammer noise was rated as being 'comfortable, but slightly loud' (loudness rating = 5.3). The SRL-processed noises were rated as being between 'comfortable' to 'comfortable, but slightly loud' (loudness categories 4 and 5) with the exception of the hammer noise which was rated as 'soft' (loudness rating = 2.1).

The difference between the loudness ratings of the FRL-processed noise and SRL-processed noise, averaged across the four noise types, was 2.1 LCs (i.e. 6.0 for FRL-processed noise compared to 3.9 for SRL-processed noise). A two-way, repeated-measures ANOVA was performed on the loudness ratings of the noises with processing condition (FRL and SRL) and noise type as factors. The results showed the main effects of processing condition and noise type and the interaction between the processing condition and the noise type all to be highly significant ($p<0.001$). Post hoc Bonferroni tests revealed that the effect of processing condition was highly significant for FRL processing compared with SRL processing for all noise types ($p<0.001$).

Post hoc Bonferroni tests also revealed that the loudness of the hammer noise was significantly different to the other noises ($p<0.001$), which was not unexpected given that the hammer was recorded at a lower level, as previously mentioned. This effect of noise type can be seen in Figure 9-16 where the loudness ratings of the FRL-processed and SRL-processed hammer noise are lower than corresponding loudness ratings for the other three noises. The interaction of the processing condition is also evident in Figure 9-16. The additional reduction in the loudness rating of the hammer due to SRL processing compared to FRL processing was 3.2 LCs (i.e. 5.3 for FRL-processed hammer noise compared to 2.1 for SRL-processed hammer noise). The loudness-summation control within the gain calculator was very effective at controlling this hammer noise. Like the SRL MKI's broadband limiter, it may have been too effective as it reduced the hammer's loudness to below speech loudness, making it unnaturally soft. This suggests that it might be advisable to reduce the 'Loudness-summation maximum reduction – fast' parameter from the 10 dB value it was set to for this experiment.

9.3.3.4 SRL-processed speech and noise loudness

The average loudness rating of the SRL-processed noise was 3.9 and the SRL-processed speech was 3.6, a difference of 0.3 LCs. For comparison, the average loudness rating of the FRL-processed noise was 6.0 and the FRL-processed speech was 3.6, a difference of 2.4 LCs. The SRL processing reduced the exceedance by 2.1 LCs.

A two-way, repeated-measures ANOVA was performed on the loudness ratings of the SRL-processed stimuli with stimulus type (speech and noise) and noise type as factors. The loudness of the SRL-processed noise was not significantly different to the loudness of the SRL-processed speech for all the noises ($p=0.066$). The noise type, however, had a highly significant effect on loudness ($p<0.001$) and the interaction between the stimulus type and noise type was highly significant ($p<0.001$). Post hoc Bonferroni tests revealed that the difference in loudness due to noise type was due to the lower level of the hammer noise and its associated speech. The source of the significant interaction between noise type and processing type can be seen in Figure 9-16 where there is a much greater effect of SRL (relative to FRL) for the hammer noise than for the lawn mower noise.

9.3.3.5 Speech quality

The results of the quality ratings of the speech, encountered in the hearing protector application, are shown in Figure 9-17. The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the FRL-processed, the SRL-processed and the reference speech for each noise type. The raw data and statistics underlying this plot are documented in Table D-6, in Appendix D.

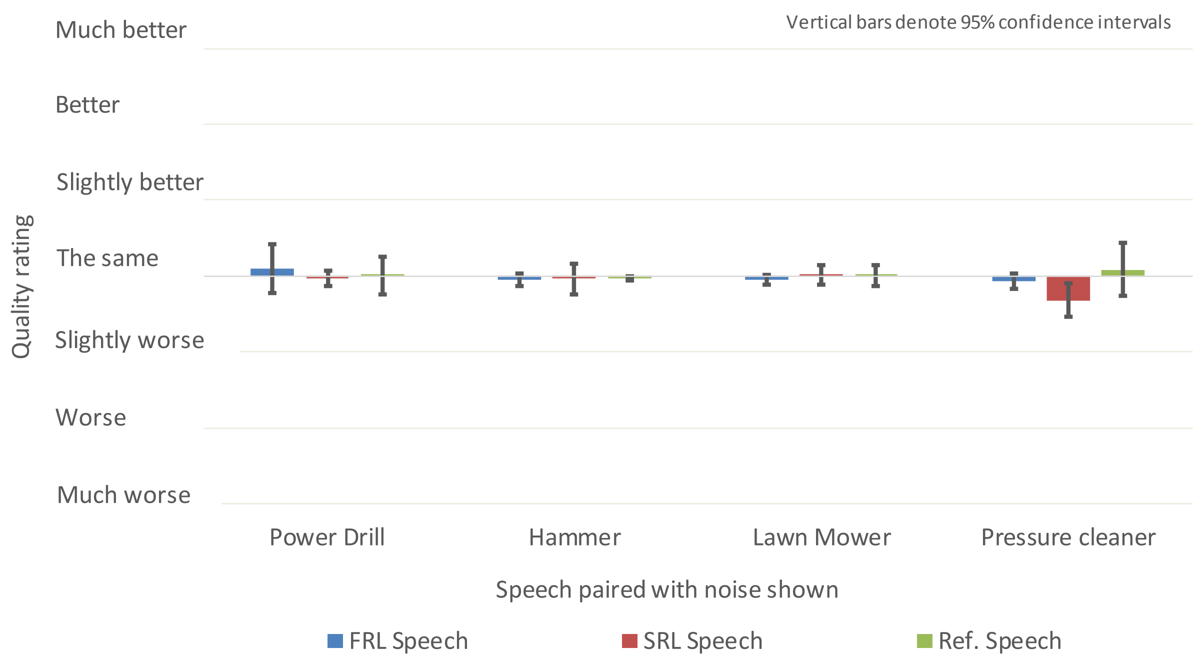


Figure 9-17. *Quality of speech encountered in hearing protector applications.*

All the speech was rated as being 'the same' as the reference speech. A two-way, repeated-measures ANOVA was performed on the speech quality ratings with processing condition and noise type as factors. The results showed no significant effect of the processing condition ($p=0.31$), of the noise type ($p=0.12$) or of the interaction between the processing condition and noise type ($p=0.076$).

9.3.3.6 Summary

Overall, the strongest feature of the loudness ratings was the combination of a 2.1 loudness-category reduction in noise loudness as a result of SRL processing, compared to conventional FRL processing, with no change in speech loudness. This reduction in loudness resulted in no practical or statistically significant change in the speech quality in all cases. Furthermore, the loudness of the SRL-processed noises was only 0.3 LCs greater than the SRL-processed speech, compared to 2.4 LCs for the FRL processing and comparable to the equal loudness found with the SRL MKI scheme. Based on this data, SRL MKII was performing as well as SRL MKI for application in level-dependent hearing protectors.

9.3.4 Experiment 3: Telephone headset application

9.3.4.1 Loudness

The results of the loudness ratings of the speech and noise, in the telephone headset application, are shown in Figure 9-18. The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the speech and noise in the conventional FRL-processed, the SRL-processed and the reference conditions for each noise type. The raw data and statistics underlying this plot are documented in Tables D-7 and D-8, in Appendix D.

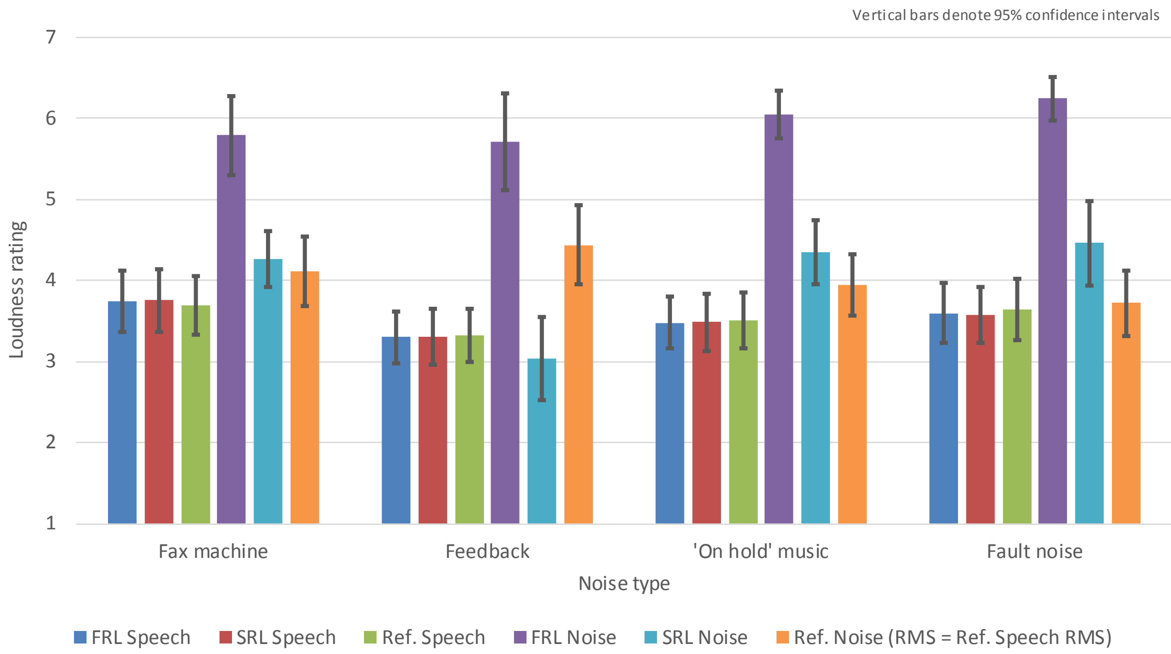


Figure 9-18. *Loudness of speech and noise encountered in telephone headset applications.*

9.3.4.2 Speech loudness

All speech was rated as being between ‘comfortable, but slightly soft’ to ‘comfortable’ (categories 3 and 4). The differences between the loudness ratings of the FRL-processed speech, the SRL-processed speech, and the reference speech, averaged across the four noise types, was less than 0.1 LCs (all three average speech loudness ratings were 3.5).

A two-way, repeated-measures ANOVA was performed on the speech loudness ratings with processing condition and noise type as factors. The main effect of processing condition was not statistically significant (p=0.78). The effect of noise type also was not significant (p=0.067), nor was the interaction between the processing condition and noise type significant (p=0.28).

9.3.4.3 Noise loudness

In contrast to the speech loudness results, there were significant differences in the noise loudness as a result of processing condition. Averaged across participants, all FRL-processed noises were rated as being 'loud, but OK' (category 6). The SRL-processed noises were rated as being 'comfortable, but slightly soft' or 'comfortable' (categories 3 and 4).

The difference between the loudness ratings of the FRL-processed noise and SRL-processed noise, averaged across the four noise types, was 1.9 LCs (i.e. 5.9 for FRL-processed noise compared to 4.0 for SRL-processed noise). A two-way, repeated-measures ANOVA was performed on the loudness ratings of the noises with processing condition (FRL and SRL) and noise type as factors. As expected, the results showed a highly significant main effect of the processing condition (p<0.001). The noise type was also highly significant (p<0.001) and the interaction between the processing condition and noise type was highly significant (p<0.001). Post hoc Bonferroni tests revealed that the effect of processing condition was highly significant for all noises types (p<0.001). The tests also revealed that SRL processing had a significantly greater effect on the feedback noise than on the other noises (p<0.001). This effect of noise type can be seen in Figure 9-18 where the feedback noise is suppressed below the speech level while the other sounds were not.

9.3.4.4 SRL-processed speech and noise loudness

The average loudness rating of the SRL-processed noises was 4.0, while for SRL-processed speech it was 3.5. A two-way, repeated-measures ANOVA was performed on the loudness ratings of SRL-processed stimuli with stimulus type (speech or noise) and noise type as factors. The loudness of the SRL-processed noises was significantly different to the loudness of the SRL-processed speech (p=0.015). The effect of the noise type on the loudness rating was highly significant (p<0.001), as was the interaction of the stimulus type and the noise (p<0.001). The effect of noise type was discussed in the preceding section on noise loudness. Post hoc Bonferroni tests revealed that both the 'on hold' music and the fault noise were significantly louder (p=0.002) than the speech (by 0.9 LCs) after SRL processing.

9.3.4.5 Speech quality

The results of the speech quality ratings are shown in Figure 9-19.

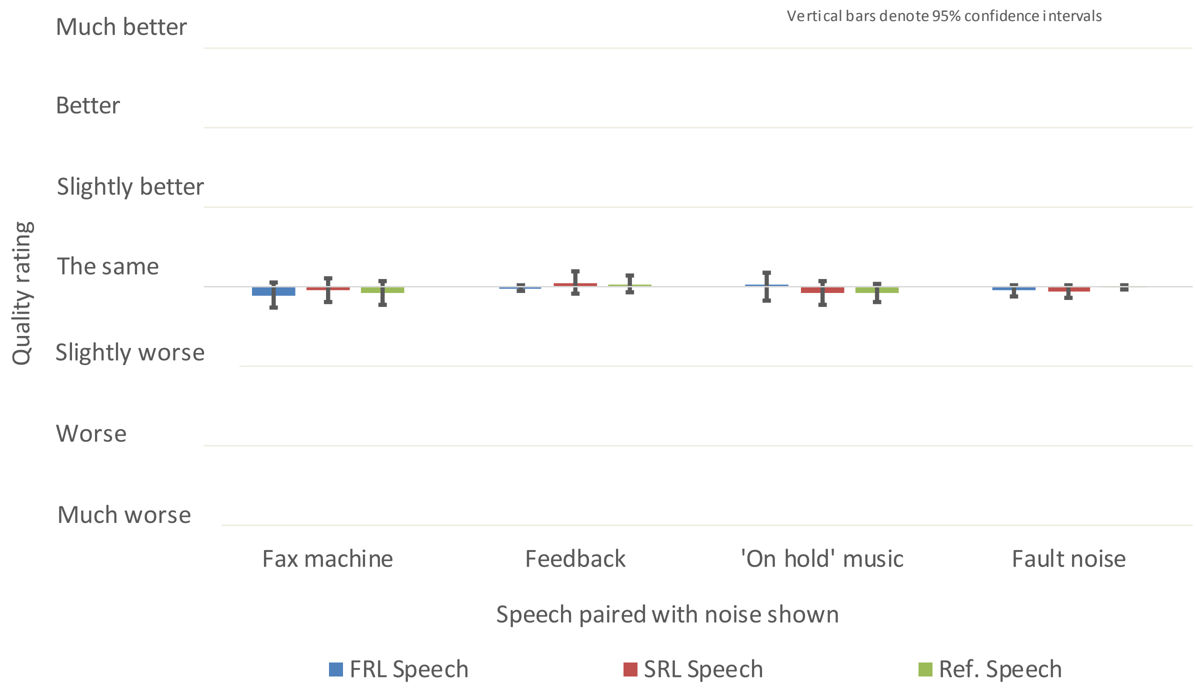


Figure 9-19. *Quality of speech encountered in telephone headset applications.*

The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the FRL-processed, the SRL-processed and the reference speech for each noise type. The raw data and statistics underlying this plot are documented in Table D-9, in Appendix D.

All the speech was rated as being 'the same' as the reference speech. A two-way, repeated-measures ANOVA was performed on the speech quality ratings with processing condition and noise type as factors. The results showed no significant effect of the processing condition (p=0.94), of the noise type (p=0.29) or of the interaction between the processing condition and noise type (p=0.44).

9.3.4.6 Summary

Like the hearing aid application and the hearing protector application, the most striking difference was the large reduction, in this case by 1.9 LCs, in the loudness of the noises as a result of SRL processing, compared to conventional FRL processing, with no reduction in the speech loudness. This reduction in the loudness resulted in no practical or statistically significant change in the speech quality in all cases.

Furthermore, the loudness of the noises was only 0.5 LCs greater than the speech, compared to 2.4 LCs for the FRL processing. Based on this data, SRL MKII was performing on par with SRL MKI for application in telephone headsets.

9.3.5 Experiment 4: Dynamic speech level

Experiment 4 was aimed at assessing the effect of large, abrupt variations in speech level on SRL's ability to control the loudness of high-level noise and preserve speech quality under those circumstances. The main results of the experiment are shown in Figure 9-20 comprising three time-aligned sub-figures: (A) waveforms, (B) loudness ratings and (C) speech quality ratings. The raw data and statistics underlying the loudness and speech quality rating plots are documented in Tables D-10, D-11, D-12 and D-13, in Appendix D.

Figure 9-20 (A) shows the waveform of two signals: one is the unprocessed signal (coloured grey) and the other is the SRL-processed signal (coloured green). The unprocessed signal is described in Section 9.2.6. The waveforms span approximately 32 seconds and comprise four sentences and noise bursts. These sentences were all identical in their unprocessed form except for their levels which sequentially were 62, 55, 68 and 75 dB SPL. Each sentence was approximately 7 seconds in duration, and was followed by a burst of white noise with a level of 90 dB SPL and a duration of 1 second.

The unprocessed speech signal is obscured by the SRL-processed speech signal for most of the waveform due to the SRL-processed speech signal being identical to the unprocessed speech signal. A difference in the waveforms, however, is evident just after the second and third blue vertical bars, which have been overlaid on the figure to demarcate the sections containing different speech levels. These two regions of discrepancy, in which the unprocessed speech signal's waveform is partially revealed, are both approximately 1 second in duration. The unprocessed signal is very evident in the large bursts of noise that follow each sentence due to the SRL processing reducing the noise to approximately the same level as its preceding speech.

(A)

(B)

(C)

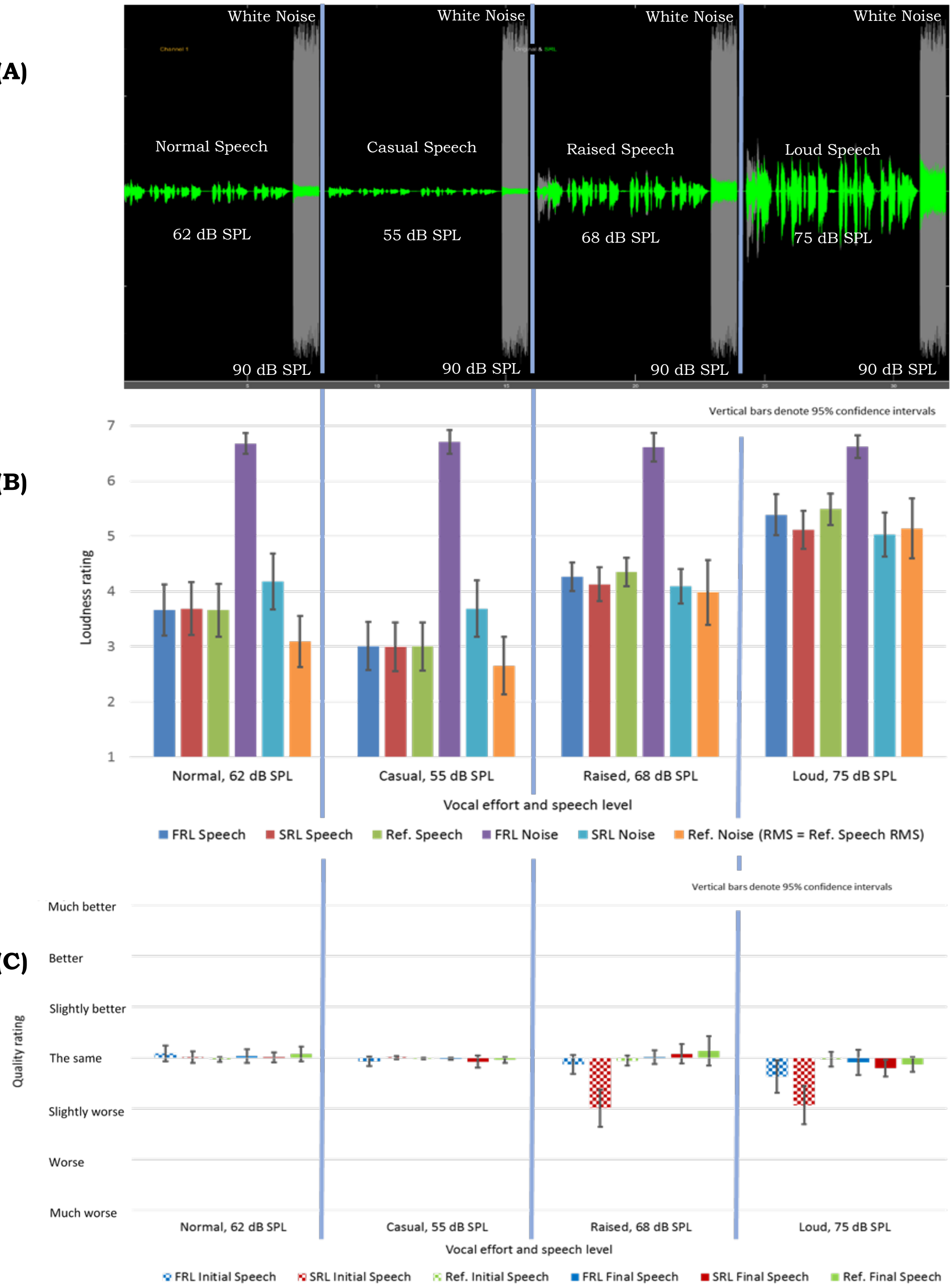


Figure 9-20. ***(A)*** *Stimulus waveforms - unprocessed (grey) and SRL-processed (green). Both waveforms comprise 4 identical sentences at 4 levels: 62, 55, 68 and 75 dB SPL interlaced with one second bursts of white noise at 90 dB SPL.* ***(B)*** *Loudness ratings.* ***(C)*** *Speech quality ratings.*

Figure 9-20 (B) shows the mean speech and noise loudness ratings, of the 16 subjects, for each of the four speech and noise pairs, under the three processing conditions (FRL-processed, SRL-processed and Reference), and the 95% confidence intervals for these means.

Figure 9-20 (C) shows the mean speech quality ratings, of the 16 subjects, separately for both the initial part and the final part of each of the four sentences, under the three processing conditions, and the 95% confidence intervals for these means.

In addition to Figure 9-20, a more detailed acoustic analysis of the unprocessed and the SRL-processed signal is shown in Figure 9-21. For orientation purposes Figure 9-21 (A) replicates the waveforms of the unprocessed (original) and the SRL-processed signals shown in Figure 9-20 (A). Figure 9-21 (B) shows a spectrogram of the unprocessed (original) signal. Figure 9-21 (C) shows a spectrogram of the SRL-processed signal and Figure 9-21 (D) shows a spectrogram of the attenuation produced by SRL (i.e. the difference between the spectrograms of the unprocessed and the SRL-processed signals).

Statistical analysis techniques were employed using two-way, repeated-measures ANOVA with post hoc testing using the Bonferroni test where applicable.

A two-way, repeated-measures ANOVA with processing condition (FRL-processed, SRL-processed, reference (unprocessed)) and speech level (62, 55, 68 and 75 dB SPL) as factors was performed on the speech loudness ratings. The main effect of processing condition was significant ($p<0.002$), as was the speech level ($p<0.001$) and the interaction between the two ($p<0.01$).

A two-way, repeated-measures ANOVA with processing condition (FRL-processed, SRL-processed) and speech level (62, 55, 68 and 75 dB SPL) as factors was performed on the noise loudness ratings. The main effect of processing condition was highly significant ($p<0.001$), as was the speech level ($p<0.001$) and the interaction between the two ($p<0.001$).

A two-way, repeated-measures ANOVA with stimuli type (SRL-processed speech and SRL-processed noise) and speech level (62, 55, 68 and 75 dB SPL) as factors was performed on the loudness ratings for SRL-processed stimuli. The main effect of the stimuli was not significant ($p=0.116$), however, the speech level was highly significant ($p<0.001$) and the interaction between the two was significant ($p=0.016$).

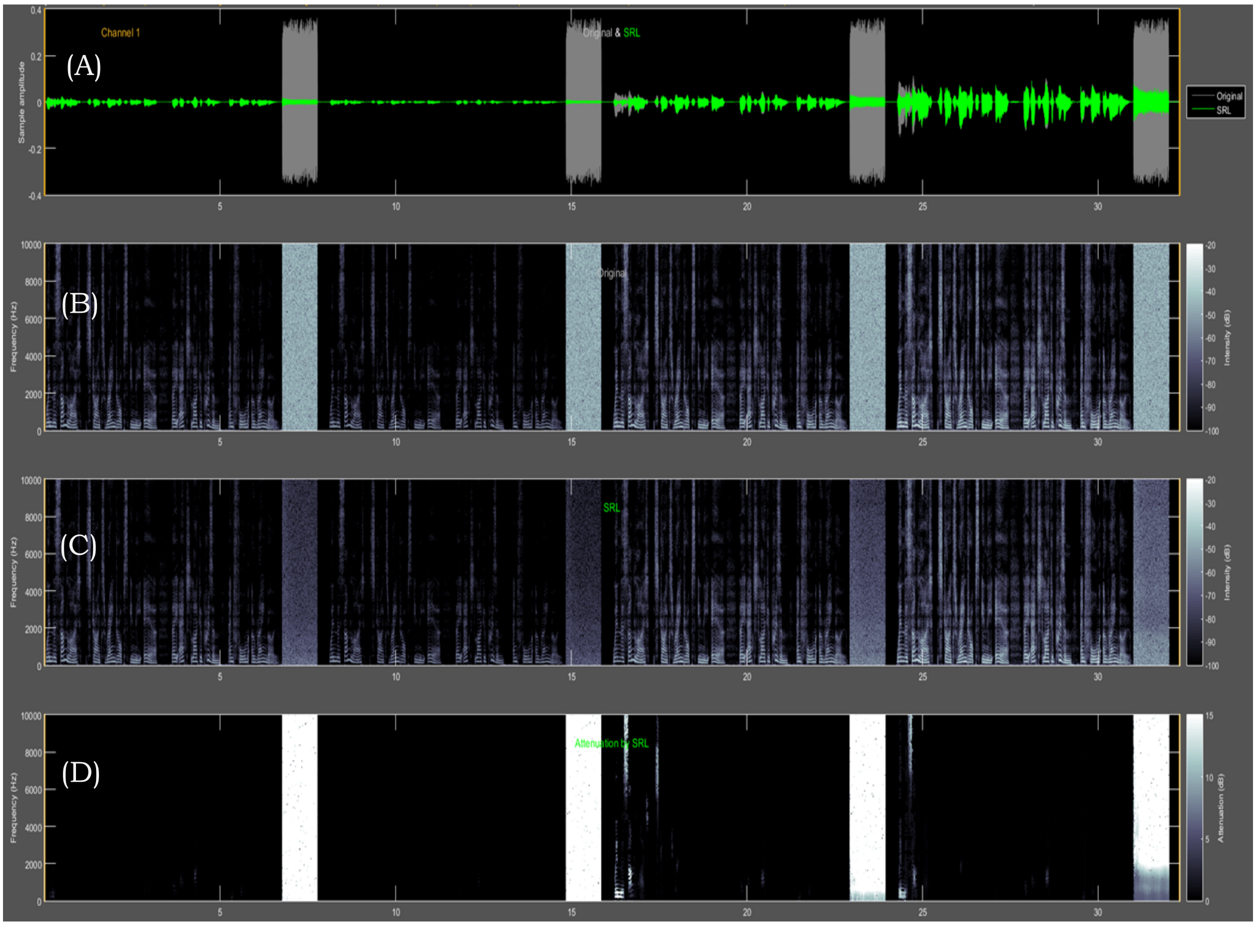


Figure 9-21. ***(A)*** *Stimulus waveforms – original unprocessed (grey) and SRL-processed (green). Both waveforms comprise 4 identical sentences, at 4 levels: 62, 55, 68 and 75 dB SPL interlaced with one second bursts of white noise at 90 dB SPL.* ***(B)*** *Spectrogram of the original unprocessed signal.* ***(C)*** *Spectrogram of the SRL-processed signal.* ***(D)*** *Spectrogram of the attenuation produced by SRL.*

A two-way, repeated-measures ANOVA with processing condition (FRL-processed, SRL-processed, reference (unprocessed)) and speech level (62, 55, 68 and 75 dB SPL) as factors was performed on the speech quality ratings of the initial speech (i.e. the first half of the sentence). The main effect of processing was highly significant ($p<0.001$), as was the speech level ($p<0.001$) and the interaction between the two ($p<0.001$).

A two-way, repeated-measures ANOVA with processing condition (FRL-processed, SRL-processed, reference (unprocessed)) and speech level (62, 55, 68 and 75 dB SPL) as factors was performed on the speech quality ratings of the final speech (i.e. the second half of the sentence). The main effect of processing was not significant

($p<0.386$), nor was the speech level ($p=0.09$) nor the interaction between the two ($p=0.67$).

Observations and analysis of the results for each sentence are set out below.

9.3.5.1 Sentence 1

The waveform of SRL-processed first sentence in Figure 9-20 (A) appears identical to the reference (unprocessed) sentence which was at a level of 62 dB SPL. Inspection of the attenuation plot in Figure 9-21 (D) confirms no attenuation occurred during the first sentence (i.e. the plot is all black in this region). The speech loudness for all conditions (FRL-processed, SRL-processed and reference (unprocessed)) was 3.7 LCs. Post hoc testing using the Bonferroni test, following a significant two-way, repeated-measures ANOVA result, showed no significant difference between any of the conditions ($p=1$) at this speech level. There was no significant difference in speech quality as a result of condition, for either the initial or final parts of the sentence: all were rated as 'the same' as the Reference, i.e. the unprocessed speech, (initial: $p=0.15$, final: $p=0.66$). The 90 dB SPL noise following the first sentence was substantially reduced by the SRL processing but not entirely reduced to the speech level due largely to the minimum speech reference levels set within the SRL processing. This was reflected in the loudness rating for the noise which were slightly higher than the preceding speech by 0.5 LCs (noise: 4.2 LCs, speech: 3.7 LCs, a highly significant difference statistically, $p<0.001$, but not very significant from a practical perspective). In contrast to this, the FRL-processed noise had a loudness rating of 6.7 LCs. This was 2.5 LCs greater than the SRL-processed noise and the difference was highly significant ($p<0.001$).

9.3.5.2 Sentence 2

The substantial attenuation applied by the SRL processing to the first noise burst had no effect on the speech that followed it, nor did the 7 dB fall in the unprocessed speech level from 62 dB to 55 dB SPL. The waveform of the SRL-processed second sentence was identical to the unprocessed speech, as shown in the detailed attenuation plot of the second sentence in Figure 9-21 (D) which is all black in this region. The speech loudness for all conditions (FRL-processed, SRL-processed and reference (unprocessed)) was 3.0 LCs. Post hoc testing using the Bonferroni test, following a significant interaction between processing condition and speech level in the corresponding two-way ANOVA, showed no significant difference between any of the processing conditions (p=1) at this speech level. There was no significant difference in speech quality as a result of condition, for either the initial or final parts of the sentence: all were rated as 'the same' as the reference (initial: p=0.093, final: p=0.28). The attenuation applied by the SRL processing to the noise burst that followed this sentence at a level of 55 dB SPL was greater than was applied to the noise burst that followed the previous sentence, which had a level of 62 dB SPL. The increased attenuation was not quite as great as the drop in the speech level. This was due to it being restricted by the minimum speech reference levels set within the SRL program. This was reflected in the loudness rating for the noise being higher than the preceding speech by 0.7 LCs (noise: 3.7 LCs, speech: 3.0 LCs, a highly significant difference statistically, p<0.001, but not very significant from a practical perspective given the lower loudness). In contrast to this, the FRL-processed noise had a loudness rating of 6.7 LCs. This was 3.0 LCs greater than the SRL-processed noise and the difference was highly significant (p<0.001).

9.3.5.3 Sentence 3

The unprocessed third sentence increased by 13 dB relative to the previous sentence (i.e. from 55 to 68 dB SPL). Initially, the SRL-processed speech was attenuated to a level slightly above that of the previous sentence. This attenuation diminished with time and after about 1 second was no longer apparent. The effect can be seen in the detailed attenuation plot in Figure 9-21 (D), where both the voicing, largely below 2 kHz, followed by two bursts of sibilance, largely above 5 kHz, were attenuated. This reduction in the initial level of the speech is reflected in the slightly lower loudness ratings for the SRL-processed sentence compared to the reference (unprocessed) speech (SRL-processed: 4.1 LCs, reference: 4.4 LCs), the FRL-processed speech was also slightly lower at 4.3 LCs. Post hoc testing using the Bonferroni test, following a significant two-way, repeated-measures ANOVA result, showed no significance of the 0.3 LC difference between the SRL-processed and reference condition (p=0.42) nor between the other conditions (p=1) at this speech level. The reduction in the initial speech level of the SRL-processed speech was also reflected in the reduced speech quality rating for the initial speech (-1.0, 'slightly worse', p<0.001). There was, however, no effect on the speech quality rating of the final SRL-processed speech (+0.1, p=0.38). The third noise burst was attenuated by the SRL processing to about the same level as the preceding speech. This is reflected in the loudness rating of the SRL-processed noise being the same as the preceding SRL-processed speech (both 4.1 LCs, no significant difference, p=0.86). In contrast to this, the loudness rating of the FRL-processed noise was 6.6 LCs. This was 2.5 LCs greater than the SRL-processed noise and the difference was highly significant ($p<0.001$).

9.3.5.4 Sentence 4

The unprocessed fourth sentence increased in level by 7 dB relative to the previous sentence (i.e. from 68 to 75 dB SPL). The waveforms show that initially the SRL-processed speech was attenuated to a level slightly above that of the preceding sentence. Being less of a step in level than in the previous case, the attenuation was less and was no longer apparent after about 1 second. The attenuation can be seen in the detailed attenuation plot in Figure 9-21 (D), where the voicing, largely below 2 kHz, followed by one burst of sibilance, largely above 5 kHz, were attenuated. The reduction in the initial speech level was reflected in the lower loudness ratings for this SRL-processed sentence compared to the unprocessed, reference sentence (SRL-processed: 5.1 LCs, Reference: 5.5 LCs). Post hoc testing using the Bonferroni test, following a significant two-way, repeated-measures ANOVA result in the corresponding AVOVA, showed the 0.4 LC difference between the SRL-processed and reference condition to be highly significant ($p<0.001$). None of the other differences in condition were significant ($p>0.05$) at this speech level. The reduction in the initial speech level was also reflected in the reduced speech quality rating for the initial part of the SRL-processed sentence (-0.9, 'slightly worse', $p<0.001$). There was, however, little to no effect on the speech quality rating of the final part of the SRL-processed sentence (-0.2, $p=0.50$). The initial part of the fourth sentence also appeared to be affected by the fixed-reference limiting thresholds set by the subjects as their rating of the quality of the FRL-processed speech during its initial period was also reduced (-0.4, $p=0.08$), although this was not statistically significant. This would, however, have contributed to the reduction in quality of the SRL-processed speech during its initial period, as it was processed by the FRL process in addition to its SRL processing (see Section 9.2.5 for details). The final noise burst was attenuated by the SRL processing to approximately the level of the preceding speech. This was reflected in the loudness rating of the SRL-processed noise being about the same as the preceding speech (noise: 5.0 LCs, speech: 5.1 LCs, no significant difference, $p=0.74$). In contrast to this, the loudness rating of the FRL-processed noise was 6.6 LCs. This was 1.6 LCs greater than the SRL-processed noise and the difference was highly significant ($p<0.001$).

9.3.5.5 Summary

This experiment evaluated the ability of the SRL MKII processing scheme to control the loudness of noise relative to speech for large abrupt changes in speech level. Over a 20 dB range of speech levels (55, 62, 68 and 75 dB SPL), corresponding to casual, normal, raised and loud vocal efforts, the SRL processing was shown to track the speech loudness and control the noise loudness in relation to it. On average, it reduced the loudness of high-level noise by 2.4 LCs compared to the fixed-reference limiting set by the subjects (i.e. from 6.7 LCs for the FRL-processed stimuli compared to 4.2 LC's for the SRL-processed stimuli). The loudness of the SRL-processed noise was on average 0.3 LCs above the loudness of the SRL-processed speech and this difference was not significant ($p=0.17$). The SRL processing was very effective as it almost matched the loudness of the noise to that of the speech that immediately preceded it, despite the level of that speech varying over a 20 dB range across the four sentences. This was very apparent at the higher speech levels where it was more closely matched than was expected for a headroom setting of 3 dB. It would appear that the loudness-summation reduction maximum values, which were set to 10 dB and 8 dB for the fast and slow limiters respectively, enabled the loudness of this wideband noise to be controlled very tightly in relation to the speech loudness.

The loudness of the SRL-processed speech was the same as the unprocessed reference speech for speech at 55 and 62 dB SPL ($p=0.28$ and $p=0.59$ respectively). A small reduction of 0.3 LCs ($p=0.42$) in the speech loudness occurred for speech at 68 dB SPL and 0.4 LCs ($p<0.001$) for speech at 75 dB SPL. Inspection of the attenuation revealed these reductions were in the initial period, of up to approximately 1 second, following a large increase in the speech level. These reductions were not as a result of the noise attenuation, if this had not been the case these reductions would have occurred equally for the lower-level speech.

The quality of the settled SRL-processed speech was the same as the unprocessed, reference speech for all sound levels ($p= 0.28$, $p=0.66$, $p=0.38$ and $p=0.50$ for speech levels of 55, 62, 68 and 75 dB SPL respectively). Only in the period immediately following an increase in the speech level was the SRL-processed speech different from the unprocessed, reference speech. In these two cases, it was rated as 'slightly worse' ($p<0.001$) for speech levels of 68 and 75 dB after level rises of 13 dB and 7 dB respectively. The period of attenuation, which caused this reduction in perceived quality, was up to approximately 1 second in both cases.

9.3.6 Experiment 5: Simultaneous noise

This fifth experiment was aimed at assessing the effect of high-level noise simultaneously presented with speech on the SRL's ability to control the loudness of high-level noise in relation to the speech loudness and preserve speech quality.

The results of the loudness ratings of the speech and noise are shown in Figure 9-22. The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the speech and noise in the conventional FRL-processed and the SRL-processed conditions for each noise type. The raw data and statistics underlying this plot are documented in Tables D-14 and D-15, in Appendix D.

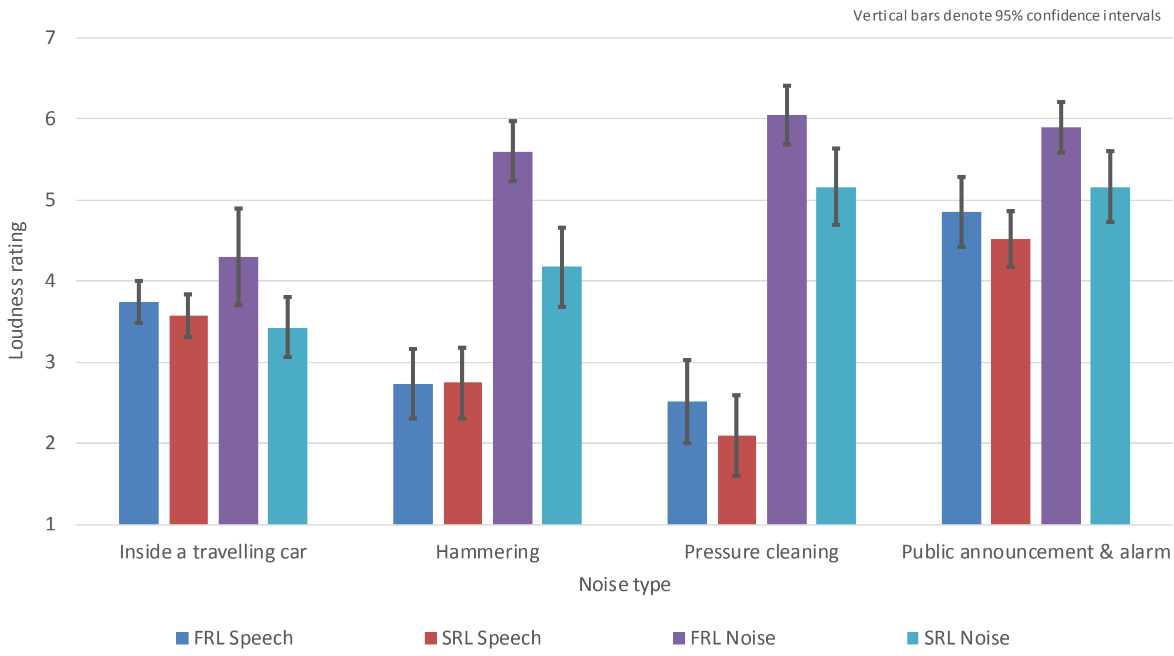


Figure 9-22. *Loudness ratings of speech and noise for four stimuli comprising simultaneously present speech and noise.*

(The unprocessed speech levels for the four stimuli in order were: 68, 53, 65 and 80 dB SPL.)

The results of the speech quality ratings are shown in Figure 9-23. The figure shows the mean ratings of the 16 subjects and the 95% confidence intervals of the means for the conventional FRL-processed and the SRL-processed speech for each noise type. The raw data and statistics underlying this plot are documented in Table D-16, in Appendix D.

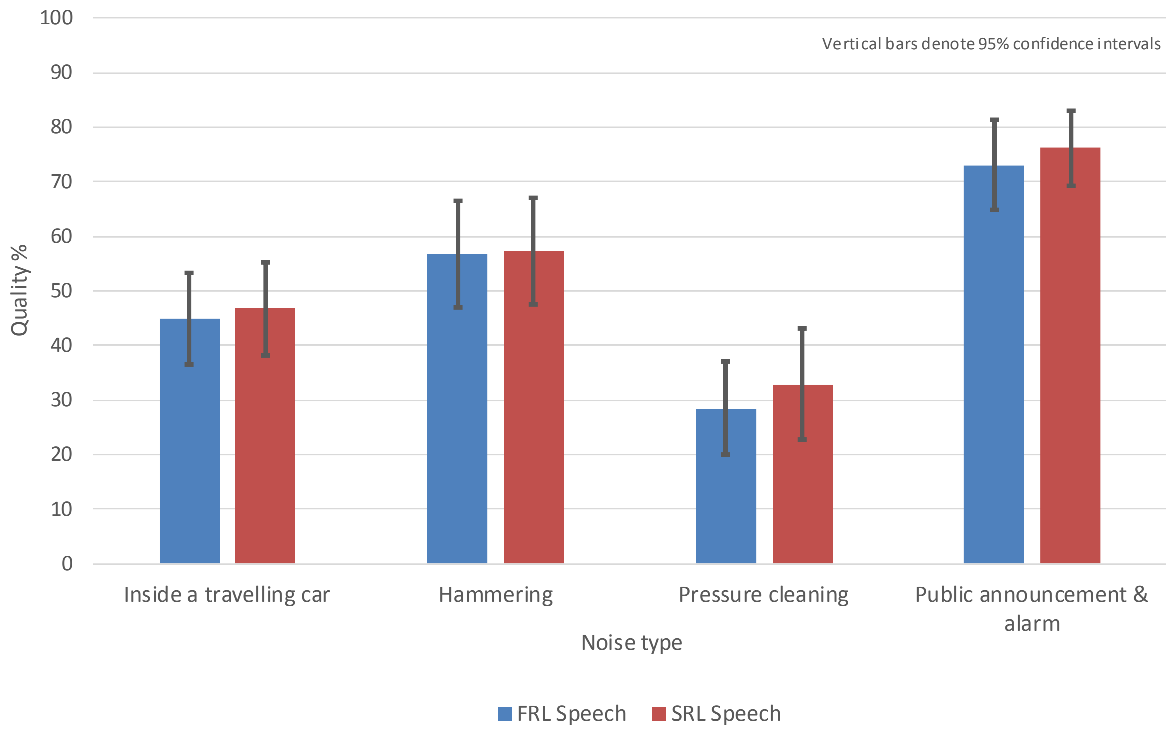


Figure 9-23. *Speech quality ratings for four stimuli comprising simultaneously presented speech and noise.*

Statistical analysis techniques were employed using two-way repeated-measures ANOVA followed by post hoc testing using the Bonferroni test.

Speech loudness: When averaged across the four stimuli, the SRL-processed speech had a mean loudness rating of 3.2 LCs, this was 0.3 LCs below the loudness of the FRL-processed speech (3.5 LCs). A two-way, repeated-measures ANOVA was performed on the speech loudness ratings with processing condition (FRL and SRL) and stimuli type as factors. The analysis revealed the main effect of processing condition to be significant ($p=0.003$). The effect of stimuli and the interaction between processing condition and stimulus type were also found to be significant ($p<0.001$ and $p=0.008$ respectively).

Noise loudness: When averaged across the four stimuli the SRL-processed noise had a loudness of 4.5 LCs, this was 1.0 LC below the loudness of the FRL-processed noise (5.5 LCs). A two-way, repeated-measures ANOVA was performed on the noise loudness ratings with processing condition (FRL and SRL) and stimuli type as factors. The analysis revealed the main effect of processing condition to be highly significant ($p<0.001$). The effect of stimuli and the interaction between processing condition and stimulus type were also found to be highly significant (p<0.001) and significant (p=0.004) respectively.

SRL speech to noise loudness difference: When averaged across the four stimuli the SRL-processed noise had a loudness of 4.5 LCs, this was 1.2 LCs above the loudness of the SRL-processed speech (3.2 LCs) after rounding. A two-way, repeated-measures ANOVA was performed on the noise loudness versus speech loudness ratings with speech-versus-noise and stimuli type as factors. The analysis revealed the main effect of speech-versus-noise to be highly significant (p<0.001). The effect of stimuli and the interaction between speech-versus-noise and stimuli were also found to be highly significant (p<0.001) in both cases.

Speech quality: The scale for the absolute quality ratings was described in Section 7.2 and the computer screen absolute quality rating scale used by the subjects was shown in Figure 9-12. When averaged across the four stimuli the SRL-processed speech had a mean quality rating of 53.2%, this was 2.4% above the quality rating of the FRL-processed speech (50.8%). These mean ratings both fall within the range of the 'fair' quality category. A two-way, repeated-measures ANOVA was performed on the speech quality ratings with processing condition (FRL and SRL) and stimuli type as factors. The analysis revealed the main effect of processing to be non-significant (p=0.093). The effect of stimuli was found to be highly significant (p<0.001). The interaction between process and stimuli, however, was non-significant (p=0.63).

Observations and analysis of the results for each of the four stimuli are set out below.

9.3.6.1 Stimulus 1 - Speech in a travelling car

The loudness rating of the SRL-processed speech (3.6 LCs) was 0.1 LCs below that of FRL-processed speech (3.7 LCs). A difference of 0.1 LCs is of no practical significance and post hoc testing using the Bonferroni test revealed this difference to be non-significant (p=1). The speech quality rating of the SRL-processed speech (46.7%) was 1.9% above that of the FRL-processed speech (44.8%). Post hoc testing using the Bonferroni test revealed this difference to be non-significant (p=1). The loudness rating of the SRL-processed noise (3.4 LCs) was 0.9 LCs below that of the FRL-processed noise (4.3 LCs). Post hoc testing using the Bonferroni test revealed this difference to be highly significant (p<0.001). Furthermore, the SRL-processed noise was 0.2 LCs softer than the SRL-processed speech. Post hoc testing using the Bonferroni test revealed this difference to be non-significant (p=1). In summary, the

SRL processing significantly reduced the perceived loudness of the noise to such a degree that it was slightly below that of the speech; it improved the perceived speech quality by a small but statistically non-significant degree compared with FRL processing; and produced a non-significant 0.1 LC reduction in the speech loudness compared with FRL processing.

9.3.6.2 Stimulus 2 - Speech and hammering

The loudness rating of the SRL-processed speech (2.7 LCs) was the same as the FRL-processed speech (2.7 LCs). Post hoc testing using the Bonferroni test revealed this difference to be non-significant ($p=1$). The speech quality rating of the SRL-processed speech (57.2%) was 0.4% above that of the FRL-processed speech (56.8%). Post hoc testing using the Bonferroni test revealed this difference to be non-significant ($p=1$). The loudness rating of the SRL-processed noise (4.2 LCs) was reduced to 1.4 LCs below that of the FRL-processed noise (5.6 LCs). Post hoc testing using the Bonferroni test revealed this difference to be highly significant ($p<0.001$). The SRL-processed noise was reduced to a loudness 1.5 LCs above that of the SRL-processed speech. Post hoc testing using the Bonferroni test revealed this difference to be highly significant ($p<0.001$). In summary, the SRL processing reduced the noise to a comfortable loudness (4.2 LCs) which was 1.4 LCs below that of the FRL-processed noise with no change in speech quality or loudness.

9.3.6.3 Stimulus 3 - Speech and pressure cleaner

The loudness rating of the SRL-processed speech (2.1 LCs) was 0.4 LCs below that of the FRL-processed speech (2.5 LCs). Post hoc testing using the Bonferroni test revealed this difference to be highly significant ($p=0.001$) although the difference was small. Both loudness ratings were substantially lower than the expected loudness if the speech was not accompanied by noise (which was estimated to be around 4.5 LCs). The speech quality rating of the SRL-processed speech (32.8%) was 4.3% above that of the FRL-processed speech (28.5%). Post hoc testing using the Bonferroni test revealed this difference to be non-significant ($p=1$). The loudness rating of the SRL-processed noise (5.2 LCs) was 0.8 LCs below that of FRL-processed noise (6.0 LCs). Post hoc testing using the Bonferroni test revealed this difference to be highly significant ($p<0.001$). The SRL-processed noise, however, was 3.1 LCs louder than the SRL-processed speech while the FRL-processed noise was 3.5 LCs louder than the FRL-processed speech, both differences were highly significant using a Bonferroni test ($p<0.001$). The reduced speech loudness in both the FRL and the SRL conditions can be attributed primarily to partial masking, i.e. a reduction in the loudness of a target sound due to being partially masked by a competing sound (see Moore et al. 1997 for a description). [24] In addition to partial masking, some of the loudness reduction, for the average subject, was due to a small change in the sound level of 2 dB in the FRL condition, and a larger change in the sound level of 11 dB in the SRL condition. This additional loudness reduction in the SRL condition can be attributed to the ratio of the noise to the preceding speech level (10 dB) which was across many frequencies and was particularly large at high frequencies. This extra attenuation, however, did not degrade the rated speech quality. In summary, the SRL-processing reduced the amount that the noise loudness exceeded the speech loudness by 0.4 LCs with no statistically significant change in the speech quality.

9.3.6.4 Stimulus 4 - Speech and alarm

The loudness rating of the SRL-processed speech (4.5 LCs) was 0.4 LCs below that of the FRL-processed speech (4.9 LCs). Post hoc testing using the Bonferroni test revealed this difference to be significant ($p=0.013$). The speech quality rating of the SRL-processed speech (76.1%) was 3.1% above that of the FRL-processed speech (73.0%). Post hoc testing using the Bonferroni test revealed this difference to be non-significant ($p=1$). The loudness rating of the SRL-processed noise (5.2 LCs) was 0.7 LCs below that of the FRL-processed noise (5.9 LCs). Post hoc testing using the Bonferroni test revealed this difference to be highly significant ($p<0.001$). Furthermore, the SRL-processed noise was reduced to being within 0.7 LCs of the SRL-processed speech. Post hoc testing using the Bonferroni test revealed this difference to be non-significant ($p=0.15$). In summary, the SRL processing reduced the perceived loudness of the alarm noise to a loud but more comfortable level. The loudness of the alarm noise was not significantly different from the perceived speech loudness. The SRL-processing resulted in a slight, but non-significant, improvement in the speech quality.

9.3.6.5 Summary with further comment

This experiment evaluated the ability of the SRL MKII processing scheme to control the loudness of noise relative to speech for noises simultaneously presented with the speech. The stimuli represented a range of conditions.

- **Low-frequency noise (stimulus 1):** The noise in the travelling car was largely low-frequency. In this frequency region, the speech level was low and hence limiting the car noise to the speech level resulted in a positive improvement in loudness comfort of 0.9 LCs for SRL over FRL with no significant change in either the speech level or quality. This was achieved because most of the noise was in a frequency region which was of minor importance to the overall perception of speech quality and loudness and therefore reduction in this frequency region had no effect on the speech. The fact that neither loudness nor quality was adversely affected through controlling the noise is evidence of the effectiveness of the low-frequency resolution and the adaptive control of the SRL MKII scheme.

- **Time-separated impact noise (stimulus 2):** The hammering comprised of large brief periods of energy separated in time. The noise had a far higher short-term level than the speech during these periods so limiting it with reference to the speech resulted in a positive improvement in loudness comfort of 1.4 LCs for SRL over FRL with no significant change in either the speech level or quality. This occurred because the noise only overlapped with the speech during a limited number of brief periods. The fact that neither the loudness nor the quality of the speech was adversely affected through controlling the noise is testament to the fast recovery of the SRL MKII scheme to brief noise bursts. The fact that the noise was so heavily suppressed is evidence of an effective attack time within the scheme. However, the resulting noise loudness was higher than the speech loudness, this would indicate that some of the simultaneously hammering noise may have corrupted the speech reference levels, i.e. the tails of the reverberant hammering energy may have been sampled along with the speech. For comparison the hammering was far more suppressed when not simultaneously presented with the speech in Experiment 2.
- **Wideband continuous turbulent noise (stimulus 3):** The pressure cleaner noise was turbulent and wideband. The spectral peaks of speech, i.e. formants, being stronger in the low to mid-frequencies, regularly exceeded the noise in this frequency region, but rarely so in the high-frequency region. As a result, limiting the noise in relation to the speech decreased the level of high-frequency noise (and the high-frequency speech) resulting in a positive improvement in loudness comfort of 0.8 LCs for SRL over FRL with only a 0.4 LC reduction in the speech level and no significant change in the speech quality. This occurred because of the inherent low-frequency emphasis of speech[63] compared to the flatter spectrum of the pressure cleaner.

- **Mid-frequency modulated harmonic series (stimulus 4):** The alarm and the speech overlapped considerably in their frequency content, their level and in time, as the alarm was constant. The spectro-temporal differences between the speech and the noise enabled the SRL-processing to produce a 0.7 LC improvement in the loudness comfort of the noise for a 0.4 LC reduction in the speech loudness with no significant change in the speech quality. This occurred because of the spectro-temporal differences between the two sounds and is a reflection of the multi-band resolution of the SRL MKII scheme and the speed at which it adapts to the changes in the noise.

In each case, the SRL-processed noise was significantly reduced in loudness, and for half the stimuli, the noise loudness was not significantly different to the speech loudness. In half the cases, the change in speech loudness was not significant and when it was significant, it was reduced by only 0.4 LCs relative to the FRL condition. However, the effect of partial masking of the speech by the pressure cleaner noise was highly significant for both FRL and SRL conditions. In each case, the speech quality rating showed a small, but non-significant, improvement as a result of SRL processing when compared to FRL processing. The effect, when averaged over the four stimuli, and analysed using a two-factor repeated-measures ANOVA, was not significant ($p=0.093$).

9.3.7 Loudness discomfort

Prior to commencing these five experiments the subjects set their individual limiting level according to the instruction *find the maximum setting of this control for which the noises are loud but do not cause you discomfort.* They did this first while listening to all the band-limited monaurally-presented noises they were going to hear later in the training session and the experiments. It was emphasised to the subjects to be sure they were confident in their selection before committing to it. The subjects repeated this process for the wideband binaurally-presented noises. Their choices were recorded and only preprocessed stimuli with the same limiting level for that case were presented to them. However, during the five experiments there were 47 loudness ratings in which the subjects rated what they heard as being uncomfortably loud (i.e. a loudness rating of 7, the maximum loudness rating). One would presume that if ratings higher than 7 were available to the subjects many would have chosen them (i.e. the sound being perceived as being louder than just uncomfortably loud). Except for two ratings, all these uncomfortably loud ratings occurred with the FRL-processed noises they had listened to during the limit setting process. The subjects who gave these responses had clearly changed their opinion about the loudness of these noises from when they were presented during the limit setting stage to the loudness rating stage in the experiments. The context in which they heard the sounds had, however, changed, rather than being a noise in a sequence of noises, as they were in the limit setting procedures, they were in a speech context.

Table 9-17 shows the limit settings and the number of maximum loudness ratings (i.e. 7 or uncomfortably loud) that occurred during the experiments for each subject, grouped by the subject gender, presentation method and experiment (i.e. BL monaural experiments 1 & 3 and WB binaural experiments 2, 4 and 5).

The codes for the stimuli for which there was a loudness discomfort rating are as follows:

N#:	FRL-processed noise
S#:	FRL-processed speech
Ref N#:	Reference noise

where the # is the number of the stimulus in the given experiment: 1 to 4, where 1 corresponds to the far left-hand stimulus in the loudness rating figures through to 4 which corresponds to the far right-hand stimulus in the loudness rating figures.

None of the loudness discomfort ratings occurred when the subjects were listening to the SRL-processed sounds. With the exception of one rating of speech in the FRL-processed condition and one rating of noise in the reference condition, all the loudness discomfort ratings occurred when the subjects were listening to the FRL-processed noises. The two mentioned exceptions were ratings by subject 6, who rated the fourth FRL-processed speech and the fourth reference noise (which had the same RMS level as the reference speech) in Experiment 4 as being uncomfortably loud (i.e. a rating of 7). These stimuli had, at most, a sound level of 75 dB $SPL_{\text{diffuse-field equivalent}}$, a speech level that the average person considers to be loud but not uncomfortably loud.[140]

Female subjects				Loudness rating of 7 (uncomfortably loud)					
	Limiter settings			BL Monaural Experiments		WB Binaural Experiments			
Subject #	BL Monaural	WB Binaural		1	3	2	4	5	Tally
1	14	7			N4		N1,N2,N3,N4		5
2	6	6							0
4	14	9					N1,N2,N3		3
5	13	9							0
6	14	5			N1,N3,N4	N1,N3,N4	S4, N1,N2,N4 & REF N4		11
7	12	6							0
8	12	13					N2		1
Average	**12**	**8**	**Total**		**4**	**3**	**13**		**20**

Male subjects				Loudness rating of 7 (uncomfortably loud)					
	Limiter settings			BL Monaural Experiments		WB Binaural Experiments			
Subject #	BL Monaural	WB Binaural		1	3	2	4	5	Tally
3	11	8					N1,N2,N3		3
9	11	13							0
10	8	8							0
11	12	12		N3		N1,N4	N1,N2,N3,N4	N3	8
12	13	12				N4		N3	2
13	6	4							0
14	17	12					N2		1
15	14	13		N3			N3,N4		3
16	13	12		N1,N2,N3	N1,N2		N1,N2,N3,N4	N4	10
Average	**12**	**10**	**Total**	**5**	**2**	**3**	**14**	**3**	**27**

Total number of loudness ratings of 7 (uncomfortably loud) for females and males **47**

Table 9-17. *Limit settings and number of maximum loudness ratings (i.e. 7 or uncomfortably loud) that occurred during the experiments for each subject, grouped by the subject gender, presentation method and experiment (i.e. BL Monaural experiments 1 & 3 and WB Binaural experiments 2, 4 and 5).*

Subject 14, a male with the highest overall limit setting, recorded one loudness discomfort rating. Interestingly, this was in the binaural condition for which he did not have the highest limit setting. Subject 13, a male with the lowest overall limit setting, recorded no loudness discomfort rating. Subject 6, a female, and subjects 11 and 16, two males, recorded the majority of the loudness discomfort ratings with 11, 8 and 10 discomfort ratings respectively out of the total of 47 discomfort ratings, these are highlighted with a grey background in the table. Of the 47 loudness discomfort ratings 37 occurred with the wideband binaurally presented stimuli. There were 320 ratings of the FRL-processed noise (16 subjects x 5 experiments x 4 noises). Of these presentations of FRL-processed noise, 45 were rated as uncomfortably loud. This corresponds to approximately 14% of these FRL-processed noise ratings. Given that all these sounds were previously presented and considered by the subjects not to cause discomfort within, at most, the previous 70 minutes, one could be excused for thinking this figure should have been 0%. This effect of loudness discomfort for stimuli previously considered to not cause loudness discomfort does not correlate with the limit setting and was distributed evenly across the genders. The effect has a greater prevalence in certain subjects.

The FRL-processed burst of white noise in Experiment 4 caused the greatest number of loudness discomfort ratings of all the stimuli presented as shown in Table 9-17. A more insightful view of this data is shown in columns 3 to 6 of Table D-11, in Appendix D. In this table, the cells highlighted in red indicate the maximum loudness rating. In these columns there were multiple maximum loudness ratings of 7 given. The stimuli were the same as the FRL-processed burst of white noise that was presented to the subjects in the limit selection procedure. This was confirmed through re-analysing the noise levels post the experiment. The only difference between the noises was the sequence of sounds in which they were presented. In the limit setting procedure, the noise was part of a compilation of noises, while in Experiment 4 it was preceded by 7 seconds of speech at RMS levels of 55, 62, 68 and 75 dB SPL $_{\text{diffuse-field equivalent}}$. The number of incidents of the noise being given a loudness rating of 7 (uncomfortably loud) was negatively correlated with the preceding speech level. This is shown in Table 9-18.

Speech level RMS dB SPL $_{\text{diffuse-field equivalent}}$	Number of subjects with a loudness rating of 7 (uncomfortably loud) for noise that followed the speech	Percentage of subjects with a loudness rating of 7 (uncomfortably loud) for noise that followed the speech
75	4	25%
68	6	38%
62	6	38%
55	9	56%

Table 9-18. *The number and percentage of loudness ratings of 7 (uncomfortably loud, the maximum), given in Experiment 4, for the FRL-processed burst of white noise as a function of the preceding speech level.*

This table shows an inverse relationship between the number of maximum loudness ratings (i.e. a rating of 7 or uncomfortably loud) caused by noise of fixed characteristics and the preceding speech which varied only in level. This data gives additional weight to the argument that controlling noise levels in relation to speech levels will reduce the occurrence of loudness discomfort. Although, this is a minor observation in comparison to the compelling data on the substantially reduced loudness of SRL-controlled noises compared to FRL-controlled noises, it is a very interesting effect, particularly considering that the acclimatisation period to the speech, to create such an effect, was only 7 seconds.

In summary, loudness discomfort levels depend very much on the individual. A substantial variation in the self-selected, maximum sound levels occurred across the subjects. This resulted in an 18 dB range in the selected maximum sound levels for bursts of white noise. It was shown that loudness discomfort occurs despite the listener previously setting the exact sound that caused the discomfort to a level that would not cause discomfort. Only the context changed between earlier and later presentations. The occurrence of loudness discomfort caused by noise was not only shown to be context sensitive but was also shown to be negatively correlated with the preceding speech level. Controlling the loudness of noise relative to the loudness of speech using the SRL MKII scheme prevented this loudness discomfort from occurring.

9.4 Conclusion and recommendations

This chapter reported on the subjective evaluation of the SRL MKII scheme. From the experiment results, we can derive the following conclusions beginning with the confirming of the hypothesis.

It was hypothesised that the SRL MKII scheme would provide the greatest reduction in the excess loudness of an audio signal compared with the loudness of the preceding speech conveyed by the audio signal for the least reduction in the speech loudness and quality.

First, the aspect of the hypothesis relating to the reduction in the excess loudness of an audio signal compared with the loudness of the preceding speech conveyed by the audio signal is true for higher-level non-speech sounds. On average, for the SRL MKII scheme compared to a conventional FRL scheme, using the same underlying processing, there was a 1.9 LC reduction in the loudness of the higher-level non-speech sounds presented sequentially with the speech and a 1.0 LC reduction in the loudness of the higher-level non-speech sounds presented simultaneously with the speech. In all five experiments, the loudness of higher-level non-speech sounds was brought closer to the loudness of the preceding speech and in the majority of cases (11 out of 20) the loudness of the higher-level non-speech sounds was not significantly different to the preceding speech loudness.

Second, the aspect of the hypothesis relating to the least reduction in the speech loudness was found to be true. In all of the first four experiments, the rated loudness of the SRL-processed speech was not statistically different to the rated loudness of the reference speech, with the one exception in Experiment 4. In this case, when the unprocessed speech abruptly increased from a level of 68 dB SPL to a level of 75 dB SPL the SRL-processed speech was rated as 0.4 LCs below the reference speech. The period of reduction in loudness was up to about 1 second immediately following the level increase. In the simultaneous speech and noise cases (i.e. Experiment 5) the speech loudness resulting from the SRL processing was on average 0.3 LCs below the fixed-reference limited stimuli; a small amount compared with the reduction in the noise loudness.

Third, the aspect of the hypothesis relating to least reduction in the speech quality was found to be true for 22 out of the 24 speech samples assessed. In all of the first four experiments, the rated quality of the SRL-processed speech was not statistically different to the rated quality of the reference speech, with the exception of the transitional parts of the speech following two abrupt increases in the speech level in Experiment 4. The increases were of one and two vocal effort categories with level increases of 7 dB and 13 dB respectively. The periods of reduction in quality were up to about 1 second each immediately following the level increase. The rated quality in relation to the reference was 'slightly worse' and was statistically significant. This hypothesis also held true for the four cases of speech presented with simultaneous noise. In all these cases, the mean speech quality rating for SRL processing was slightly higher (by 2.4%) than with the FRL processing, although the difference was not statistically significant.

Overall, the SRL processing reduced the loudness of the noise by nearly two loudness categories in relation to the loudness resulting from conventional FRL processing, with user selected limits, while having only a small effect on speech loudness and quality.

Previous concerns about the over suppression of sounds such as feedback and alarms by SRL were examined. The telephone feedback in Experiment 3 was suppressed to 0.3 LCs below the speech although this difference was not statistically significant. The alarm in Experiment 5 was reduced in loudness by the SRL processing by 0.7 LCs relative to the conventional FRL case. This brought it to be within 0.7 LCs of the speech, a statistically non-significant difference. While this was at a more comfortable level it remained well above the loudness that would be necessary to get a person's attention.

Previous concerns about the over suppression of impulse sounds such as hammering by SRL were examined with mixed results. The hammer sound in Experiment 2, which was presented sequentially after the speech, remained over suppressed. Its loudness relative to the speech was -0.6 LCs, this difference was statistically significant. However, the reduction in the loudness of hammering simultaneous present with the speech in Experiment 5 was considerably less. The hammering was reduced by 1.4 LCs more with the SRL processing than with the conventional FRL processing. However, its reduced loudness was still 1.4 LCs above the speech loudness. This discrepancy in the performance of the SRL MKII scheme in these two situations requires further investigation post this doctoral thesis.

Through these experiments the limitations of fixed-reference limiting have been brought forth. Despite the subjects having full control of setting their limit with exactly the same sounds they were to hear later, some still experienced loudness discomfort when presented with the same sounds within a short time later. The occurrence of loudness discomfort was shown to be context sensitive and negatively correlated with the preceding speech level. It was found that controlling the loudness of noise relative to the loudness of speech using the SRL MKII scheme prevented this loudness discomfort from occurring.

The data also revealed an interesting result on the perception of the reference noises. These had an RMS level equal to the speech (i.e. these being the unprocessed noises which for the reference case were scaled to have an RMS level equal to the preceding speech). The theory on loudness summation, as discussed in Chapter 1, predicts that sounds with a bandwidth greater than the speech but with the same RMS level would be perceived as being louder than the speech, and those with a narrower bandwidth than speech but also with the same RMS level would be perceived as being softer, although, temporal integration effects may alter this. A comparison of the loudness data from Experiments 1 to 4 for continuous sounds with the same RMS level as speech shows that the sounds with a narrower bandwidth than speech, e.g. umpire's whistle, fax machine, feedback were statistically greater in loudness than the speech but many sounds with an equal or greater bandwidth than speech, e.g. baby crying, power drill, lawn mower, pressure cleaner and telephone fault noise were not. This data is at odds with the effect predicted by the spectral loudness summation theory. Furthermore, impact/impulsive sounds such as plates clanging and hammering were perceived as loud in relation to speech of the same RMS level. The higher perceived loudness for narrow-band sounds and narrow-time sounds (i.e. impact/impulsive) correlates with the reports of acoustic shock summarised in Chapter 3. This finding further supports the SRL approach of suppressing narrow-band and narrow-time signals in order to improve loudness comfort and acoustic safety.

The SRL MKII scheme has been shown to perform well in the laboratory using real-life recordings and specifically designed test material. Its performance against an equivalent conventional FRL scheme has been assessed using a range of stimuli with noise being both sequentially and simultaneously present with the speech. In all experiments the data has shown the SRL MKII scheme to be superior to an equivalent conventional FRL scheme in terms of noise loudness control and equal in speech quality with the exception of transitory reduction in speech quality following an abrupt increase in speech level. Further assessment of its performance in the field compared to conventional FRL schemes would be desirable.

Chapter 10

Conclusion

10 Conclusion

This thesis has presented the research, development and evaluation associated with a novel method of improving listening comfort and acoustic safety for people listening primarily to speech produced by electronic devices, such as headsets, telephones, headphones, hearing aids, cochlear implants, level-dependent hearing protectors and public address systems.

The method controlled an audio signal in a manner that preserved the quality and the intelligibility of speech conveyed by the audio signal, provided other conveyed signals were estimated to be softer and less powerful than it. If the other signals were estimated to be louder and/or more powerful than the preceding speech, then the method minimised the excess. The method regularly updated estimates of the loudness and power of the speech to use as its reference for limiting the estimated excess loudness and power of the signal, and hence was termed speech referenced limiting. It performed this processing on both a frequency-specific and a total/broadband basis.

The method addressed the trade-off between providing comfort and protection and providing good speech quality and intelligibility through adaptively setting the limiting levels based on estimates of the loudness and power of the speech. It is arguably the optimum limiting strategy for protecting the listener from high-level sounds while preserving speech, providing *the greatest limiting of noise for the least limiting of speech.*

The method directly addressed the issue of acoustic shock at a high neurophysiological level, through control of the estimated loudness, and at a low neurophysiological level, through controlling abrupt changes in the acoustic power and therefore minimising the triggering of new somatic responses.

Two schemes based on the method were developed, the SRL MKI and the SRL MKII scheme. The latter scheme was far superior with the ability to estimate the speech loudness and power from frequency regions where speech was dominant, while ignoring frequency regions where it was not. It also contained a novel method of determining the amount of additional control needed to correct for the loudness summation of noises that have a bandwidth exceeding that of speech based on the count of exceedances within the specific-frequency bands. It also applied this technique to provide fast control over the signal's peak power using a shorter delay. The signal delay was 3 ms in the evaluated SRL MKII scheme. This short delay meant the scheme could be used in applications where the user hears their own voice through the processing. Although complex, the SRL MKII scheme was efficient. A single channel with a sampling rate of 48 kHz used only about 2% of the computing power of a current laptop computer.

Subjective evaluation of the SRL MKI and SRL MKII schemes conducted in the laboratory confirmed that the performance for both schemes was as hypothesised. Using stimuli typical of those experienced in the three main intended applications (hearing aids, level-dependent hearing protectors and telephone headsets) the following two hypothesis set for SRL MKI were confirmed (with three minor exceptions):

- that the method controls the loudness of non-speech sounds relative to speech sounds for higher level non-speech sounds, and
- that the method does not degrade speech quality,

and the following hypothesis set for SRL MKII was confirmed:

> the method provides the greatest reduction in the excess loudness of an audio signal compared with the loudness of the preceding speech conveyed by the audio signal for the least reduction in the speech loudness and quality.

The SRL MKII scheme was also evaluated using abruptly changing speech levels. With the exception of a slight decrease in speech quality and loudness immediately following an abrupt increase in speech level, the above hypothesis was confirmed under this condition. Further evaluation, using noise concurrent with speech, revealed that the noise had a significantly lower loudness and speech quality was preserved for the SRL MKII scheme when compared to an equivalent conventional FRL scheme (i.e. one identical to the SRL system in regards to all processing but without the variable (i.e. SRL) limits or the SRL loudness summation method) which used fixed-reference limiting levels set by each individual subject.

Data collected during the evaluation of SRL MKII also revealed that when using fixed limits alone (i.e. FRL processing) many subjects changed their minds about the sound level that caused them loudness discomfort; the ability of some subjects to select a limiting level that would ensure that they did not experience loudness discomfort in the future was not reliable. This was true even though the noises used to set the limits were the same as those that later caused discomfort, they only differed in the context in which they were presented. Loudness discomfort resulting from noise was also shown to be related to the loudness of the speech that preceded it, further supporting the approach of using speech as a limiting reference, as the SRL scheme did. The evaluation also revealed an 18 dB range between the sound levels, resulting from fixed limits, set by the subjects, for white noise stimuli, confirming data found by others that showed a significant inter-individual variation in loudness discomfort levels.[67] Overall the data provided strong evidence to support the use of adaptive limits and provided evidence that one fixed limit level did not suit all users and that an approach such as SRL was preferable.

In summary, the hypotheses have not been disproven, with the exception of a slight, brief reduction in speech quality and loudness following an abrupt increase in speech level.

10.1 Applications

As mentioned throughout this thesis, the SRL scheme could be used to process signals in many applications where speech is the primary signal of interest. These include headsets, telephones, headphones, hearing aids, cochlear implants, level-dependent hearing protectors, public address systems and a growing number of computer-based applications. SRL is currently being assessed in a level-dependent earmuff in combination with a binaural, beam-forming, directional-microphone strategy and preliminary results are encouraging. The area of computer-delivered audio is rapidly growing and with it come concerns about acoustic safety and comfort. A computer application that operates in the background to intercept the audio and protect the user is a further potential application for SRL.

To experience the SRL MKII scheme in operation, the reader can download the SRL sound file processing application from the HEARing Co-operative Research Centre (CRC) web site. The application can process wave files of any number of channels at all the standard sampling rates up to and including 48 kHz. The link to the web site is:

http://www.hearingcrc.org/xc/xc5-applications-of-speech-referenced-limiting/

10.2 Further research

During the undertaking of this research, it has become apparent that the potential of SRL is for far more than simply a method for improving listening comfort and safety. The research goes to the fundamental question of what amplification and compression of speech should do for the listener. If it is to improve speech intelligibility or reduce the intrusiveness of noise, then it should improve the signal-to-noise ratio in at least one frequency region, or reduce either temporal or off-frequency spectral masking of the speech. It therefore should not reduce the level of speech, temporally or spectrally, with respect to the noise, and it should not increase the level of noise, either temporally or spectrally, with respect to the speech. In particular, it should reduce noise that masks the speech either temporally or spectrally. Put simply, noise that exceeds speech should be reduced but speech that exceeds noise should not be. This would imply that fast-acting multiband compression of speech that exceeds noise is detrimental to speech intelligibility and the intrusiveness of noise but that fast-acting multiband compression of noise that exceeds speech is, in theory, beneficial, ignoring co-modulation effects. This is what SRL does, based on the loudness of the speech and noise. Further research on this topic, particularly in relation to hearing aid amplification, would be worthwhile.

The potential coupling of an SRL-based system with a speech-detection controlled volume adjustment for ensuring speech is mapped across frequencies to the appropriate level for the individual would also appear worthy of further investigation.

It is the author's hope that other researchers will take inspiration from this research and this will lead to further research, development and evaluation in this area.

Michael Fisher

Michael.Fisher@nal.gov.au

Appendix A: Ethics approvals and forms

Application Form for AHHEC Class 1 Projects
(Projects with Negligible Risk)

Project number:

1. **Project title: Speech Referenced Limiting in Telecommunications**

2. **Investigators: Michael Fisher**

3. **Objective of study: To assess the efficacy of the Speech Referenced Limiting processing**

4. **Target group: Normal hearing subjects**

5. **Data collection method:**

   - ❒ questionnaire – mailout
   - ❒ questionnaire – in laboratory
   - ❒ telephone survey
   - ❒ retrospective study of existing data
   - ❒ standard clinical assessment techniques in laboratory
   - ☑ other (*please specify*):

   Presentation on acoustic stimuli to subjects via headphones by a computer in the laboratory.

6. **Brief description of protocol:**

The experiment involves presenting sounds representative of those found in these potential applications in both processed and unprocessed forms to subjects in the laboratory via headphones. The subjects perform the following tasks using the computer: loudness balancing and rating of loudness, quality and comfort.

7. **Proposed commencement date: 4th August 2010**

8. **Projected completion date: 8 September 2010**

Recommended for approval as a Class I project:

M J Fisher

Signature of Principal Investigator

Date 4/8/10.

Approved as a Class I Project:

Signature of (Research Director)

Date 4/8/10.

Australian Hearing Hub, Level 5
16 University Avenue, Macquarie University,
NSW 2109, Australia
T +61 2 9412 6872
F +61 2 9412 6769
www.nal.gov.au

NAL National Acoustic Laboratories

Australian Hearing Human Research Ethics Committee *APPROVAL FOR RESEARCH INVOLVING HUMAN SUBJECTS*	
APPROVAL NUMBER: AHHREC2015-2	
Project Number	XC5a
Project Title	SRL applications - Communications and hearing protection headset experiments
Classification	*Class 2: Project with low risk*
Principal Investigators authorized to conduct research	Michael Fisher
Date Approved	5/3/2015
Approval Method	Approved at the meeting of the full Committee.

This approval is based on the information contained in the ethics application that was presented to the Committee on 24/2/2015 and is conditional upon your continuing compliance with the National Statement on Ethical Conduct in Human Research (2007) available at:

https://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research .

A duplicate set of the documents is enclosed for your records.

Annual reporting to the Committee on progress of the project is required including a final report when the work is completed or discontinued for any reason. Reminders will be sent when progress reports are due.

The Committee expects to be notified of any changes to the approved protocol or other issues that may have an impact on the ethics of the project either by means of the annual progress reports (checklists) or as an application for variation. Adverse or unforeseen events that affect the continued ethical acceptability of the project should be reported to the Chairman immediately.

All future correspondence relating to the ethical aspects of this project must quote the above Approval Number.

TGainsford

Dr Tim Gainsford
Operations & Finance Manager, NAL
and AHHREC Secretary

Australian Hearing Hub, Level 5
16 University Avenue, Macquarie University,
NSW 2109, Australia
T +61 2 9412 6872
F +61 2 9412 6769
www.nal.gov.au

National Acoustic Laboratories

The National Acoustic Laboratories (NAL) is the research arm of Australian Hearing, a statutory government authority. NAL is funded from the government, from sponsors, grants and from commercialization of its research. The results of this research will generally be reported on in scientific publications and on the NAL website which can be accessed from www.nal.gov.au. Occasionally, a research project might additionally form part of a researcher's Master's or PhD thesis.

Information and Consent Form

Information for Participants

Date: 15/06/2015

Project title: SRL Applications – Experiments

Project number: XC5a

Background information

This study is being conducted by the National Acoustic Laboratories in conjunction with the HEARing CRC. The experiments aim to assess the effectiveness of a new method for controlling amplified sound.

The study will examine the effect of the new method on the loudness and quality of speech and noise.

This research will lead to a better understanding of how to control sound for communications and hearing protection headset wearers and hearing-aid wearers. The outcomes of the experiments will be published on the NAL website and submitted to international journals for publication. The experiments and their outcomes will also be included in the PhD thesis of Michael Fisher

What is involved?

You will participate in the experiments in one session at NAL. The session should not take more than one hour. We will present you with sounds from headphones. You will use a computer mouse to adjust the sounds you hear so they correspond to given criteria. You will use the computer mouse to rate the loudness and quality of the speech and noise that you hear. You may find some of the sounds quite loud and of variable quality. If you are not an Australian Hearing employee you will receive $30 in appreciation of your participation at the end of the session.

Australian Hearing Hub, Level 5
16 University Avenue, Macquarie University,
NSW 2109, Australia
T +61 2 9412 6872
F +61 2 9412 6769
www.nal.gov.au

National Acoustic Laboratories

Privacy of information and research outcomes

Australian Hearing is committed to protecting the privacy of our clients and research participants. You can access our Privacy Policy from http://www.hearing.com.au/privacy-policy/ or by contacting us. It outlines how we manage personal information we hold about you and how you may access or correct personal information or how you can make a complaint if you are not satisfied with the way we have handled your personal information. The personal information and data Australian Hearing collects about you as part of a research project will be treated as strictly confidential. The data aggregated by NAL may be used in research publications, however, we will not disclose your individual personal information which will be de-identified prior to inclusion in any reports of results so you will not be identified in any way.

Your rights

Participation is entirely voluntary. Although we value your participation, you are free to decide whether you will participate, and to withdraw from the research at any time. Refusal to participate or withdrawal will not affect your right to receive services from Australian Hearing and should you be an Australian Hearing employee will not affect your employment with Australian Hearing.

Contact details of researcher

If you have any concerns about this project at any stage, you are welcome to contact:

Michael Fisher
Senior Research Engineer
National Acoustic Laboratories
Tel 02 9412 6834

e-mail: michael.fisher@nal.gov.au

Complaint mechanism

The ethical aspects of this research have been approved by the Australian Hearing Human Research Ethics Committee. If you have any complaints or reservations about any ethical aspect of this research, you may contact the Committee through the Secretary (Tim Gainsford) on 02 9412 6862 or tim.gainsford@nal.gov.au. Any complaints will be treated in confidence and investigated, and you will be informed of the outcome.

Name and signature of project leader

MJ Fisher

Michael Fisher

Australian Hearing Hub, Level 5
16 University Avenue, Macquarie University,
NSW 2109, Australia
T +61 2 9412 6872
F +61 2 9412 6769
www.nal.gov.au

NAL National Acoustic Laboratories

Consent from Participant

Project title: SRL Applications Project number: XC5a
Date: _______________
I, ______________________________, have read and understood this Information and Consent Form. I freely choose to participate in this research and understand I can withdraw from participation at any time. I consent to the use of my personal and sensitive information for the purpose of my participation in this research project.
Signed: ________________________ (Participant)
Signed: ________________________ (Researcher)

Appendix B: SRL MKI subjective assessment data

Experiment 1 - Hearing Aid Application

Subject	Processing Condition & Noise Type							
	U 1	U 2	U 3	U 4	SRL 1	SRL 2	SRL 3	SRL 4
1	4.0	4.0	5.1	5.5	4.0	4.0	5.0	5.1
2	4.0	4.0	4.0	4.4	4.0	4.0	4.0	4.2
3	5.4	4.8	6.0	6.3	5.2	4.0	4.0	5.1
4	4.0	4.0	5.0	5.0	4.0	3.6	4.9	4.0
5	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
6	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
7	4.0	4.2	4.0	4.9	4.0	4.0	4.0	4.0
8	5.0	4.0	4.8	5.1	4.0	4.0	4.7	5.5
9	4.0	4.0	4.0	4.9	4.0	4.0	4.0	4.3
10	4.0	4.0	4.0	3.9	4.0	4.0	4.0	4.0
11	4.0	4.0	4.0	5.1	4.0	4.0	4.0	4.0
12	3.9	3.9	4.4	4.4	3.5	4.2	3.9	3.9
13	3.2	3.0	3.3	4.0	3.0	3.0	3.3	3.6
14	4.2	4.0	4.0	5.0	4.0	4.0	4.0	5.1
15	3.5	4.0	4.3	4.3	3.7	4.0	4.0	4.1
16	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
Average	4.1	4.0	4.3	4.7	4.0	3.9	4.1	4.3

Code	Description
U	Unprocessed
SRL	Speech Referenced Limiting

Stimulus number	Noise description
1	Baby crying
2	Plates clanging
3	Vacuum cleaning
4	Umpire's whistle

Table B-1. *SRL MKI: Exp. 1 – Speech loudness ratings.*

Experiment 1 - Hearing Aid Application

Subject	Processing Condition & Noise Type							
	U 1	U 2	U 3	U 4	SRL 1	SRL 2	SRL 3	SRL 4
1	7.0	7.0	7.0	7.0	5.8	4.8	5.1	5.7
2	6.0	5.6	6.4	5.4	4.1	4.0	4.0	4.1
3	7.0	7.0	7.0	6.7	5.9	4.7	4.4	5.3
4	6.0	6.0	5.8	5.9	4.0	4.8	4.8	4.9
5	5.1	6.5	5.6	4.9	3.8	5.1	4.0	4.0
6	6.3	6.4	6.5	6.1	4.9	4.6	4.5	4.5
7	6.8	6.5	6.7	6.6	3.6	4.7	4.5	3.9
8	7.0	7.0	7.0	6.6	6.1	6.3	4.7	5.4
9	6.4	6.6	6.6	4.9	4.5	4.0	4.0	4.0
10	6.0	6.1	7.0	6.0	5.0	4.7	5.0	5.0
11	6.0	6.1	6.1	6.0	4.0	4.0	4.0	4.0
12	6.3	6.4	6.0	6.1	4.5	4.8	4.5	5.5
13	6.0	6.2	6.7	5.6	4.3	3.5	4.0	3.0
14	7.0	6.3	7.0	7.0	6.3	4.7	6.0	5.1
15	7.0	7.0	7.0	6.6	6.6	4.6	4.0	4.0
16	6.2	6.1	6.0	6.5	5.1	5.0	4.0	4.0
Average	6.4	6.4	6.5	6.1	4.9	4.6	4.5	4.5

Code	Description
U	Unprocessed
SRL	Speech Referenced Limiting

Stimulus number	Noise description
1	Baby crying
2	Plates clanging
3	Vacuum cleaning
4	Umpire's whistle

Table B-2. *SRL MKI: Exp. 1 – Noise loudness ratings.*

Experiment 1 - Hearing Aid Application

	Processing Condition & Noise Type							
Subject	**U 1**	**U 2**	**U 3**	**U 4**	**SRL 1**	**SRL 2**	**SRL 3**	**SRL 4**
1	100	100	100	100	100	100	100	100
2	100	100	100	94	100	98	100	70
3	100	96	74	100	99	97	99	70
4	100	100	100	100	89	100	92	93
5	100	100	100	100	100	100	100	95
6	100	100	100	100	100	100	100	79
7	99	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100	89
9	100	100	100	100	100	100	100	100
10	100	100	100	100	100	100	100	100
11	100	100	100	100	87	95	100	71
12	100	98	100	98	96	82	100	90
13	100	100	100	100	100	100	100	90
14	97	100	98	100	94	45	100	81
15	100	100	100	100	100	100	100	98
16	100	100	100	100	100	100	100	100
Average	100	100	98	99	98	95	99	89

Code	Description
U	Unprocessed
SRL	Speech Referenced Limiting

Stimulus number	Noise description
1	Baby crying
2	Plates clanging
3	Vacuum cleaning
4	Umpire's whistle

Table B-3. *SRL MKI: Exp. 1 – Speech quality ratings.*

Experiment 2 - Hearing Protector Application

	Processing Condition & Noise Type							
Subject	**U 1**	**U 2**	**U 3**	**U 4**	**SRL 1**	**SRL 2**	**SRL 3**	**SRL 4**
1	4.0	3.3	4.0	4.0	4.0	3.0	4.0	4.0
2	3.6	2.4	3.0	3.5	4.2	1.9	3.2	3.5
3	3.9	3.7	4.0	4.1	4.0	3.8	4.0	4.0
4	4.8	4.0	4.0	4.0	4.8	4.0	4.0	4.0
5	4.0	2.9	4.0	4.0	4.0	3.0	4.0	4.0
6	4.0	3.0	4.0	4.0	3.9	2.9	4.0	4.0
7	4.0	2.9	4.0	4.0	4.0	3.0	4.0	4.0
8	4.0	2.8	3.8	4.0	4.0	2.2	4.0	3.6
9	3.1	2.0	3.1	3.9	3.1	2.0	3.0	4.0
10	4.0	2.7	2.9	3.9	4.0	2.8	3.0	4.0
11	4.0	4.0	4.0	4.2	4.5	4.0	4.0	4.2
12	4.0	1.8	2.9	4.1	4.0	1.8	2.9	3.0
13	4.1	2.7	4.2	4.2	4.3	2.7	4.1	3.4
14	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
15	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
16	4.0	3.0	4.0	3.6	4.0	3.1	4.0	3.5
Average	4.0	3.1	3.7	4.0	4.0	3.0	3.8	3.8

Code	Description
U	Unprocessed
SRL	Speech Referenced Limiting

Stimulus number	Noise description
1	Power drill
2	Hammer
3	Lawn mower
4	Pressure cleaner

Table B-4. *SRL MKI: Exp. 2 – Speech loudness ratings.*

Experiment 2 - Hearing Protector Application

Subject	Processing Condition & Noise Type							
	U 1	**U 2**	**U 3**	**U 4**	**SRL 1**	**SRL 2**	**SRL 3**	**SRL 4**
1	7.0	7.0	6.1	6.7	4.0	3.4	4.0	4.2
2	6.5	6.5	6.4	6.4	3.4	2.3	3.5	3.0
3	7.0	7.0	7.0	7.0	6.3	2.1	3.3	5.5
4	7.0	7.0	6.6	7.0	6.3	1.8	2.5	4.9
5	6.9	6.6	6.0	6.8	4.0	2.9	4.0	4.0
6	7.0	6.2	6.4	7.0	2.0	1.0	2.0	3.2
7	7.0	6.0	7.0	7.0	6.0	4.0	4.6	4.0
8	7.0	5.6	6.4	7.0	3.9	2.4	4.0	4.0
9	7.0	6.2	7.0	7.0	1.9	2.0	5.0	2.9
10	7.0	6.1	6.0	6.0	4.9	1.8	3.0	3.1
11	6.7	6.1	6.7	6.3	4.0	2.6	4.3	4.2
12	7.0	7.0	6.0	6.2	3.9	1.8	2.9	3.0
13	7.0	6.3	7.0	7.0	6.3	2.8	6.3	5.0
14	7.0	7.0	7.0	7.0	3.0	3.1	4.0	4.0
15	6.6	6.7	6.1	6.6	4.9	3.3	4.0	4.0
16	7.0	6.6	6.8	7.0	5.5	2.5	5.0	4.1
Average	6.9	6.5	6.5	6.7	4.4	2.5	3.9	3.9

Code	Description
U	Unprocessed
SRL	Speech Referenced Limiting

Stimulus number	Noise description
1	Power drill
2	Hammer
3	Lawn ower
4	Pressure cleaner

Table B-5. *SRL MKI: Exp. 2 – Noise loudness ratings.*

Experiment 2 - Hearing Protector Application

Subject	Processing Condition & Noise Type							
	U 1	**U 2**	**U 3**	**U 4**	**SRL 1**	**SRL 2**	**SRL 3**	**SRL 4**
1	100	100	100	100	100	100	100	100
2	98	100	97	96	99	98	97	98
3	100	100	100	100	100	100	100	100
4	100	100	100	99	100	97	100	95
5	100	99	99	100	100	99	99	96
6	96	95	95	88	95	87	94	80
7	50	100	100	100	100	100	100	100
8	100	100	100	100	100	100	100	100
9	100	100	100	100	100	100	100	100
10	100	100	100	100	100	100	100	100
11	100	100	100	100	100	96	100	96
12	79	100	100	100	81	100	100	90
13	99	98	92	100	99	98	99	92
14	79	100	50	80	65	100	50	81
15	100	50	100	100	100	50	100	100
16	100	100	100	100	100	100	100	100
Average	94	96	96	98	96	95	96	95

Code	Description
U	Unprocessed
SRL	Speech Referenced Limiting

Stimulus number	Noise description
1	Power drill
2	Hammer
3	Lawn mower
4	Pressure cleaner

Table B-6. *SRL MKI: Exp. 2 – Speech quality ratings.*

Experiment 3 - Telephone Headset Application								
	Processing Condition & Noise Type							
Subject	**U 1**	**U 2**	**U 3**	**U 4**	**SRL 1**	**SRL 2**	**SRL 3**	**SRL 4**
1	4.0	3.6	4.0	3.9	4.0	3.6	4.0	3.9
2	3.6	4.1	2.8	4.0	3.8	3.9	3.0	4.0
3	4.0	4.0	4.0	4.0	4.1	4.0	4.0	3.8
4	4.1	4.6	4.0	4.0	4.0	3.9	4.0	4.0
5	4.0	3.8	4.0	4.0	4.0	3.7	4.0	4.0
6	4.0	4.0	4.0	3.9	4.0	4.0	4.0	4.0
7	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
8	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
9	3.0	3.1	4.0	4.0	2.9	3.0	3.9	4.0
10	3.2	2.9	3.0	4.0	3.0	2.9	4.0	4.0
11	4.4	4.0	3.9	4.0	4.0	4.0	3.6	4.0
12	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
13	4.1	4.0	4.0	4.0	4.1	4.1	4.0	4.2
14	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
15	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
16	3.5	3.5	3.0	3.1	3.3	3.6	3.0	3.0
Average	3.9	3.8	3.8	3.9	3.8	3.8	3.8	3.9

Code	Description
U	Unprocessed
SRL	Speech Referenced Limiting

Stimulus number	Noise description
1	Fax machine
2	Feedback
3	'On hold' music
4	Fault noise

Table B-7. *SRL MKI: Exp. 3 – Speech loudness ratings.*

Experiment 3 - Telephone Headset Application								
	Processing Condition & Noise Type							
Subject	**U 1**	**U 2**	**U 3**	**U 4**	**SRL 1**	**SRL 2**	**SRL 3**	**SRL 4**
1	6.1	6.3	6.5	6.5	4.1	3.5	4.2	4.1
2	6.8	6.9	6.3	6.8	4.3	3.0	2.0	5.2
3	7.0	6.4	7.0	7.0	4.7	3.0	4.0	5.7
4	6.7	6.0	6.7	6.6	4.7	3.3	4.5	4.7
5	6.7	6.1	6.2	6.8	3.8	2.9	3.8	4.4
6	7.0	5.1	7.0	7.0	3.2	1.0	4.0	2.8
7	7.0	7.0	6.2	6.0	4.8	4.0	4.9	4.0
8	6.0	6.8	6.2	6.6	2.7	2.9	3.3	4.0
9	5.0	4.8	6.2	5.0	1.9	3.0	3.0	2.9
10	5.1	4.7	5.9	6.0	3.9	1.8	3.0	4.0
11	6.9	6.8	6.4	6.4	3.4	4.0	4.5	5.0
12	6.0	6.1	6.0	5.9	4.0	2.8	2.8	4.0
13	7.0	6.8	7.0	7.0	5.3	4.0	4.8	4.7
14	7.0	6.1	7.0	7.0	1.0	1.0	4.0	3.1
15	5.8	4.7	5.8	6.0	4.0	3.3	3.4	4.2
16	6.1	5.7	5.1	6.1	3.0	3.3	2.4	3.6
Average	6.4	6.0	6.3	6.4	3.7	2.9	3.7	4.1

Code	Description
U	Unprocessed
SRL	Speech Referenced Limiting

Stimulus number	Noise description
1	Fax machine
2	Feedback
3	'On hold' music
4	Fault noise

Table B-8. *SRL MKI: Exp. 3 – Noise loudness ratings.*

Experiment 3 - Telephone Headset Application

Subject	Processing Condition & Noise Type U 1	U 2	U 3	U 4	SRL 1	SRL 2	SRL 3	SRL 4
1	100	100	100	100	100	100	100	100
2	100	98	97	94	100	97	96	81
3	100	100	100	100	100	100	100	100
4	100	100	100	100	100	100	100	100
5	100	100	100	100	100	100	100	100
6	98	97	95	100	97	98	93	100
7	100	100	100	100	100	100	100	100
8	100	100	99	100	100	100	100	100
9	100	100	100	100	100	100	100	100
10	100	100	100	100	100	100	100	100
11	100	100	100	100	100	100	100	100
12	100	100	100	100	100	100	100	100
13	98	99	99	100	98	100	99	100
14	79	100	100	100	59	100	100	100
15	100	100	100	100	100	100	100	100
16	100	100	100	100	100	100	100	100
Average	98	100	99	100	97	100	99	99

Code	Description
U	Unprocessed
SRL	Speech Referenced Limiting

Stimulus number	Noise description
1	Fax machine
2	Feedback
3	'On hold' music
4	Fault noise

Table B-9. *SRL MKI: Exp. 3 – Speech quality ratings.*

Appendix C: SRL MKII file processing

File processing for SRL MKII subjective evaluation

The processing resulted in 748 sound files being generated (22 speech & noise files x 2 processing methods x 17 fixed (maximum) limiting levels). In addition to these sound files, another ensemble of files was required for the training session and Experiments 1 to 4, containing only the speech part of the original stimuli. The speech part was extracted and saved separately and then processed in the same manner as the speech and noise. This resulted in a further 748 files being produced for the speech only (22 speech-only files x 2 processing methods x 17 fixed (maximum) limiting levels).

In addition to these two ensembles, a group of reference stimuli was produced for the training session and Experiments 1 to 4. This comprised the unprocessed speech and noise with the noise adjusted so its RMS level matched the RMS level of the speech. Together each set comprised 70 files as follows:

1. The reference speech and noise (with noise RMS = speech RMS).
2. The SRL-processed speech and noise, with 17 maximum fixed-limit values.
3. The FRL-processed speech and noise, with 17 maximum fixed-limit values.
4. The reference speech only.
5. The SRL-processed speech only, with 17 maximum fixed-limit values.
6. The FRL-processed speech only, with 17 maximum fixed-limit values.

The set was of the same form for 14 out of the 22 original stimuli. There were twice as many speech-only quality files involved in Experiment 4 resulting in 105 files in the four sets. The remaining four stimuli sets (Experiment 5) comprised only items 2 and 3 above. The total number of training and assessment stimuli was:

$$14 * 70 + 4 * 105 + 4 * 34 = 1536 \text{ files}$$

Appendix D: SRL MKII subjective assessment data

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	REF 1	REF 2	REF 3	REF 4
F	1	3.7	3.7	3.7	3.2	3.7	3.7	3.7	3.3	3.7	3.7	3.7	3.2
F	2	3.0	3.0	5.0	2.9	3.0	3.0	5.0	2.9	3.0	3.0	5.1	4.0
F	4	3.6	3.9	4.1	4.6	3.6	3.9	4.2	4.6	3.6	3.8	4.2	4.6
F	5	3.7	3.6	5.0	4.0	3.7	3.7	5.1	4.1	3.7	3.6	5.1	4.0
F	6	4.0	4.1	5.3	3.7	3.9	5.0	5.0	3.1	4.0	4.9	5.1	3.4
F	7	4.0	3.0	4.0	3.1	4.0	3.0	3.9	3.0	4.0	3.1	3.9	2.9
F	8	4.4	4.0	4.1	4.1	4.5	4.0	4.0	4.0	4.5	4.0	4.0	4.1
M	3	4.4	3.0	3.0	4.9	4.4	3.0	3.0	4.9	4.4	3.2	3.0	5.0
M	9	3.9	3.9	4.0	4.0	4.0	4.0	4.1	4.0	4.1	3.9	4.0	4.1
M	10	4.2	3.9	4.2	4.5	4.2	4.0	4.2	4.5	4.1	3.9	4.2	4.7
M	11	3.0	3.0	4.0	4.0	3.0	3.0	4.0	3.9	2.9	3.0	4.0	4.0
M	12	4.0	3.5	4.0	3.8	3.9	4.0	3.9	3.9	4.0	3.9	3.8	4.0
M	13	4.0	3.2	3.9	5.0	4.0	3.2	4.0	5.0	4.0	3.1	4.0	5.0
M	14	4.1	3.9	4.0	4.0	4.0	4.1	4.0	4.0	4.0	4.1	4.1	4.0
M	15	3.3	3.9	4.5	4.3	3.5	3.9	4.6	4.3	3.4	4.1	4.6	4.2
M	16	4.0	3.0	4.0	4.0	4.1	3.0	4.0	3.9	4.0	3.0	4.0	4.1

Over all stimuli	FRL	SRL	REF	All
Female Average	3.9	3.9	3.9	3.9
Male Average	3.9	3.9	3.9	3.9

Basic statistics	FRL				SRL				REF				FRL	SRL	REF
Max	4.4	4.1	5.3	5.0	4.5	5.0	5.1	5.0	4.5	4.9	5.1	5.0	5.3	5.1	5.1
Mean	3.8	3.5	4.2	4.0	3.8	3.6	4.2	4.0	3.8	3.6	4.2	4.1	3.9	3.9	3.9
Min	3.0	3.0	3.0	2.9	3.0	3.0	3.0	2.9	2.9	3.0	3.0	2.9	2.9	2.9	2.9

Differences in condition	FRL - SRL				FRL - REF				SRL - REF				FRL-SRL	FRL-REF	SRL-REF
Max	-0.1	-0.9	0.2	-0.1	-0.1	-0.9	0.2	-0.1	0.0	0.0	-0.1	0.0	0.2	0.2	-0.1
Mean	0.0	-0.1	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	0.0	0.0	-0.1	0.0	0.0	0.0
Min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0

Stimulus Number	Stimulus noise description
1	Baby crying
2	Plates clanging
3	Vacuum cleaning
4	Umpire's whistle

Table D-1. *SRL MKII: Exp. 1 – Hearing Aids – Speech loudness ratings.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	REF 1	REF 2	REF 3	REF 4
F	1	6.6	6.0	6.0	6.1	4.8	5.0	4.7	5.0	3.7	4.3	4.8	5.2
F	2	4.8	5.2	6.4	5.0	4.4	4.0	5.3	4.3	3.8	4.4	5.9	3.9
F	4	6.8	6.7	6.7	6.4	5.5	5.5	4.1	5.0	3.9	4.1	4.5	6.0
F	5	5.5	6.3	5.9	4.7	3.7	4.8	5.1	4.8	3.1	4.8	4.5	4.9
F	6	6.1	6.5	6.8	4.6	4.2	5.6	4.3	2.7	3.1	3.8	3.3	3.7
F	7	5.5	4.5	5.5	5.1	4.1	3.0	2.9	3.9	3.1	3.0	2.9	4.1
F	8	5.9	4.9	5.1	4.3	4.5	4.2	3.6	4.4	3.8	4.3	3.7	4.4
M	3	6.6	6.0	6.6	5.9	5.9	5.5	5.2	5.2	5.0	4.8	4.6	5.5
M	9	5.0	5.0	5.1	4.0	4.0	3.9	3.0	4.0	3.0	3.9	4.0	4.0
M	10	6.3	5.4	6.0	4.5	6.0	5.0	5.0	4.6	4.6	5.2	6.0	5.0
M	11	6.0	6.0	7.0	6.0	4.1	4.4	3.9	4.0	2.0	3.7	4.0	5.0
M	12	5.7	5.0	6.0	3.9	3.9	3.0	4.9	4.0	3.2	2.9	3.8	4.0
M	13	4.3	3.9	4.0	5.0	4.2	4.0	4.0	5.0	4.0	3.6	4.5	5.1
M	14	6.2	6.0	6.6	5.8	5.0	4.4	4.9	5.0	3.9	4.9	4.9	5.0
M	15	6.5	6.8	7.0	5.9	4.7	5.2	5.2	3.9	3.9	5.0	4.3	5.2
M	16	7.0	7.0	7.0	6.0	6.1	6.0	5.0	5.1	5.1	5.1	5.0	5.1

Over all stimuli	FRL	SRL	REF	All
Female Average	5.7	4.4	4.1	4.7
Male Average	5.8	4.6	4.4	4.9

Basic statistics	FRL				SRL				REF				FRL	SRL	REF
Max	7.0	7.0	7.0	6.4	6.1	6.0	5.3	5.2	5.1	5.2	6.0	6.0	7.0	6.1	6.0
Mean	5.9	5.7	6.1	5.2	4.7	4.6	4.4	4.4	3.7	4.2	4.4	4.8	5.7	4.5	4.3
Min	4.3	3.9	4.0	3.9	3.7	3.0	2.9	2.7	2.0	2.9	2.9	3.7	3.9	2.7	2.0

Differences in condition	FRL - SRL				FRL - REF				SRL - REF				FRL-SRL	FRL-REF	SRL-REF
Max	0.9	0.9	1.7	1.3	1.9	1.8	1.0	0.4	1.0	0.9	-0.6	-0.8	0.9	1.0	0.1
Mean	1.2	1.1	1.7	0.8	2.2	1.5	1.7	0.4	1.0	0.4	0.0	-0.3	1.2	1.5	0.3
Min	0.7	0.9	1.1	1.2	2.3	1.0	1.0	0.2	1.6	0.1	-0.1	-1.0	1.2	1.8	0.6

Noise less speech	FRL: Noise - Speech				SRL: Noise - Speech				REF: Noise - Speech				FRL Ave	SRL Ave	REF Ave
Mean	2.1	2.2	1.9	1.2	0.9	1.0	0.3	0.5	-0.1	0.6	0.2	0.7	1.8	0.6	0.3

Stimulus Number	Stimulus noise description
1	Baby crying
2	Plates clanging
3	Vacuum cleaning
4	Umpire's whistle

Table D-2. *SRL MKII: Exp. 1 – Hearing Aids – Noise loudness ratings.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	REF 1	REF 2	REF 3	REF 4	Over all stimuli			
F	1	-1.0	-0.1	-0.4	-0.1	0.0	-0.6	-0.3	0.1	-1.0	-0.8	0.0	-0.1				
F	2	-0.1	-0.1	-0.1	-0.5	0.0	-0.1	-0.1	-0.5	-0.1	-0.1	0.0	0.0				
F	4	-0.1	0.0	0.0	-0.1	-0.1	0.0	-0.1	0.0	-0.1	-0.1	0.0	0.0				
F	5	0.0	0.0	0.0	-0.2	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	-0.2				
F	6	0.2	-0.8	0.7	0.1	0.2	-0.3	-0.4	0.1	0.1	0.7	0.5	0.1	Female Average			
F	7	-0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	FRL	SRL	REF	All
F	8	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0	0.0	-0.1	-0.1	0.0	-0.1
M	3	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	-0.5	0.0	0.0	-0.3	-0.4				
M	9	0.0	0.0	0.0	-0.1	0.0	0.0	0.1	0.1	0.0	-0.1	0.1	0.0				
M	10	-0.3	0.1	0.2	1.0	0.0	0.0	-0.4	1.6	0.0	0.2	0.2	0.0				
M	11	-0.1	-0.1	0.0	0.0	-0.1	0.0	0.0	-0.1	0.0	-0.1	-0.1	0.0				
M	12	0.0	-0.1	-0.1	0.0	-0.1	0.2	0.0	0.0	0.1	-0.1	-0.1	0.0				
M	13	-0.1	-0.1	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	0.0	-0.1	0.0	-0.1				
M	14	0.0	0.0	0.1	-0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	Male Average			
M	15	0.0	0.0	-0.5	-0.1	-0.1	-1.0	-0.6	-0.4	-0.1	0.1	0.0	-0.5	FRL	SRL	REF	All
M	16	0.1	0.0	0.0	0.0	1.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Basic statistics		FRL				SRL				REF				FRL	SRL	REF	
Max		0.2	0.1	0.7	1.0	1.0	0.2	0.1	1.6	0.1	0.7	0.5	0.1	1.0	1.6	0.7	
Mean		-0.1	-0.1	0.0	0.0	0.0	-0.1	-0.1	0.0	-0.1	0.0	0.0	-0.1	0.0	0.0	0.0	
Min		-1.0	-0.8	-0.5	-0.5	-0.1	-1.0	-0.6	-0.5	-1.0	-0.8	-0.3	-0.5	-1.0	-1.0	-1.0	
Differences in condition		FRL - SRL				FRL - REF				SRL - REF				FRL-SRL	FRL-REF	SRL-REF	
Max		-0.8	0.0	0.6	-0.6	0.1	-0.6	0.2	0.9	0.9	-0.5	-0.4	1.5	-0.6	0.4	0.9	
Mean		-0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.1	-0.1	-0.1	0.1	0.0	0.0	0.0	
Min		-0.8	0.2	0.1	0.0	0.1	0.0	-0.2	0.0	0.9	-0.2	-0.3	0.0	0.0	0.1	0.0	

Stimulus Number	Stimulus noise description
1	Baby crying
2	Plates clanging
3	Vacuum cleaning
4	Umpire's whistle

Table D-3. *SRL MKII: Exp. 1 – Hearing Aids – Speech quality ratings.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	REF 1	REF 2	REF 3	REF 4	Over all stimuli			
F	1	3.8	1.7	2.8	4.0	3.7	1.7	2.9	3.8	3.8	1.7	2.7	3.9				
F	2	2.9	1.6	4.0	3.6	2.9	1.6	3.9	3.5	3.2	1.6	3.9	3.5				
F	4	3.3	1.9	3.7	3.8	3.2	2.0	3.6	3.7	3.2	2.1	3.7	3.8				
F	5	4.0	2.9	4.0	3.8	4.0	2.9	4.0	3.8	4.0	3.1	4.0	3.8				
F	6	4.0	2.0	4.2	4.0	4.0	2.0	4.8	4.1	3.9	2.1	4.8	4.5	Female Average			
F	7	3.9	3.1	3.9	3.9	3.9	2.9	4.0	3.9	4.0	3.0	4.0	4.0	FRL	SRL	REF	All
F	8	4.0	2.5	3.9	4.1	3.9	2.5	3.9	4.1	4.0	2.4	4.0	4.2	3.4	3.4	3.5	3.4
M	3	4.0	2.4	3.4	3.5	4.0	2.4	3.5	3.6	4.0	2.4	3.5	3.7				
M	9	4.1	2.0	4.0	4.0	4.1	2.0	4.0	4.1	4.0	2.1	4.0	4.0				
M	10	3.8	2.2	4.0	3.9	3.8	2.5	4.0	3.9	3.9	2.5	4.0	4.0				
M	11	4.0	3.0	4.0	4.1	4.0	3.0	3.9	4.0	4.0	3.0	4.0	4.0				
M	12	4.9	4.0	4.0	4.1	4.2	3.9	4.0	4.0	4.8	4.1	4.1	4.0				
M	13	4.0	3.9	4.1	4.0	4.0	3.9	4.1	3.8	4.1	3.9	4.1	3.9				
M	14	4.0	4.0	4.0	4.0	4.2	4.0	4.0	4.1	4.2	4.1	4.1	3.9	Male Average			
M	15	4.0	3.0	3.8	4.2	3.9	3.1	3.8	4.1	3.9	3.1	3.8	4.1	FRL	SRL	REF	All
M	16	4.0	2.0	3.0	3.9	4.0	2.0	3.0	4.0	4.0	2.0	3.0	4.1	3.7	3.7	3.7	3.7
Basic statistics		FRL				SRL				REF				FRL	SRL	REF	
Max		4.9	4.0	4.2	4.2	4.2	4.0	4.8	4.1	4.8	4.1	4.8	4.5	4.9	4.8	4.8	
Mean		3.9	2.6	3.8	3.9	3.9	2.7	3.8	3.9	3.9	2.7	3.9	3.9	3.6	3.6	3.6	
Min		2.9	1.6	2.8	3.5	2.9	1.6	2.9	3.5	3.2	1.6	2.7	3.5	1.6	1.6	1.6	
Differences in condition		FRL - SRL				FRL - REF				SRL - REF				FRL-SRL	FRL-REF	SRL-REF	
Max		0.7	0.0	-0.6	0.0	0.1	-0.1	-0.6	-0.3	-0.5	-0.1	-0.1	-0.3	0.1	0.1	-0.1	
Mean		0.1	0.0	0.0	0.0	0.0	-0.1	-0.1	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	
Min		0.0	0.0	-0.1	0.0	-0.3	0.1	0.2	-0.1	-0.2	0.1	0.2	0.0	0.0	0.1	0.1	

Stimulus Number	Stimulus noise description
1	Power drill
2	Hammer
3	Lawn mower
4	Pressure cleaner

Table D-4. *SRL MKII: Exp. 2 – Hearing Protectors – Speech loudness ratings.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	REF 1	REF 2	REF 3	REF 4
F	1	7.0	6.4	6.9	6.2	6.0	1.6	4.8	4.9	4.9	3.6	4.9	3.4
F	2	5.9	5.0	6.0	6.0	3.9	1.7	3.3	4.1	3.5	2.1	2.9	3.6
F	4	6.8	6.6	6.7	6.7	4.7	1.1	5.6	4.9	3.5	2.9	4.2	3.2
F	5	6.4	6.2	6.5	6.3	5.1	3.8	4.3	4.6	2.6	4.2	3.2	3.9
F	6	7.0	3.5	7.0	7.0	5.2	1.0	6.4	6.0	4.5	2.1	6.8	5.0
F	7	6.0	3.9	6.1	5.4	3.9	1.0	5.0	4.5	3.5	2.1	2.0	3.0
F	8	6.3	4.8	6.0	6.2	2.0	2.5	3.2	2.8	4.0	3.3	3.3	2.9
M	3	6.9	6.2	6.2	6.7	5.9	1.6	5.8	4.5	5.0	4.1	5.0	4.2
M	9	6.2	4.0	6.1	4.9	4.1	2.1	3.9	3.1	3.1	2.1	3.1	3.1
M	10	5.5	4.2	5.4	5.4	3.3	2.1	4.2	2.6	2.4	3.3	2.8	3.3
M	11	7.0	4.0	6.0	7.0	4.0	2.0	5.0	4.0	3.0	2.0	4.0	3.0
M	12	6.9	6.0	6.9	7.0	5.9	2.9	5.1	5.9	6.4	4.1	3.5	4.6
M	13	5.8	4.8	5.2	5.6	4.9	2.0	4.8	3.9	4.1	4.0	4.0	2.5
M	14	5.9	6.1	6.4	5.8	4.1	4.1	6.0	4.0	3.0	3.4	5.1	3.2
M	15	6.8	6.8	6.4	6.6	2.5	1.8	5.1	4.3	1.7	2.3	2.6	2.6
M	16	6.1	6.0	5.9	6.0	4.0	1.9	4.9	4.8	2.2	4.0	4.0	4.0

Over all stimuli

Female Average

FRL	SRL	REF	All
6.1	3.9	3.5	4.5

Male Average

FRL	SRL	REF	All
6.0	3.9	3.5	4.5

Basic statistics	FRL				SRL				REF				FRL	SRL	REF
Max	7.0	6.8	7.0	7.0	6.0	4.1	6.4	6.0	6.4	4.2	6.8	5.0	7.0	6.4	6.8
Mean	**6.4**	**5.3**	**6.2**	**6.2**	**4.3**	**2.1**	**4.8**	**4.3**	**3.6**	**3.1**	**3.8**	**3.5**	**6.0**	**3.9**	**3.5**
Min	5.5	3.5	5.2	4.9	2.0	1.0	3.2	2.6	1.7	2.0	2.0	2.5	3.5	1.0	1.7

Differences in condition	FRL - SRL				FRL - REF				SRL - REF				FRL-SRL	FRL-REF	SRL-REF
Max	1.0	2.7	0.6	1.0	0.6	2.5	0.2	2.0	-0.4	-0.2	-0.3	1.0	0.6	0.2	-0.3
Mean	**2.1**	**3.2**	**1.4**	**1.9**	**2.8**	**2.2**	**2.4**	**2.7**	**0.8**	**-1.0**	**1.0**	**0.8**	**2.1**	**2.5**	**0.4**
Min	3.6	2.5	2.0	2.3	3.8	1.6	3.2	2.4	0.3	-1.0	1.2	0.1	2.5	1.8	-0.7

Noise less speech	FRL: Noise - Speech				SRL: Noise - Speech				REF: Noise - Speech				FRL Ave	SRL Ave	REF Ave
Mean	**2.5**	**2.6**	**2.4**	**2.3**	**0.5**	**-0.6**	**1.0**	**0.4**	**-0.4**	**0.4**	**0.0**	**-0.5**	**2.4**	**0.3**	**-0.1**

Stimulus Number	Stimulus noise description
1	Power drill
2	Hammer
3	Lawn mower
4	Pressure cleaner

Table D-5. *SRL MKII: Exp. 2 – Hearing protectors – Noise loudness ratings.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	REF 1	REF 2	REF 3	REF 4
F	1	-0.4	-0.2	-0.3	-0.6	0.5	-1.0	-0.5	-0.9	-1.0	-0.1	0.0	-0.4
F	2	-0.1	0.0	-0.1	0.0	-0.1	-0.1	-0.1	0.0	-0.1	0.0	0.0	0.0
F	4	-0.1	-0.1	-0.1	0.0	0.0	-0.1	-0.1	-0.5	-0.1	-0.1	0.0	-0.1
F	5	0.0	0.0	-0.3	0.0	-0.1	-0.1	0.5	-0.1	0.0	0.0	-0.3	0.0
F	6	2.2	0.0	0.1	0.1	-0.2	0.0	0.4	-1.0	1.5	0.1	0.9	2.5
F	7	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	-1.0	0.0	-0.1	0.0	0.0
F	8	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	0.0
M	3	0.0	0.0	0.1	0.0	-0.1	0.1	0.4	0.0	0.0	0.0	0.0	0.0
M	9	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
M	10	0.5	-0.1	0.1	-0.2	0.0	-0.4	-0.2	-0.6	0.0	0.0	0.0	-0.4
M	11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0
M	12	-0.4	-0.6	0.0	-0.5	-0.6	0.0	0.0	-0.1	-0.1	-0.1	0.0	-0.3
M	13	-0.1	-0.1	-0.1	0.0	-0.1	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
M	14	0.0	0.0	-0.1	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	-0.1	0.2
M	15	-0.3	0.0	-0.1	0.0	0.0	0.0	0.0	-0.9	-0.1	-0.1	-0.1	-0.1
M	16	0.1	0.1	0.0	0.0	0.1	1.0	0.0	-0.1	0.0	-0.1	-0.1	-0.1

Over all stimuli

Female Average

FRL	SRL	REF	All
0.0	-0.2	0.1	0.0

Male Average

FRL	SRL	REF	All
0.0	0.0	0.0	0.0

Basic statistics	FRL				SRL				REF				FRL	SRL	REF
Max	2.2	0.1	0.1	0.1	0.5	1.0	0.5	0.1	1.5	0.1	0.9	2.5	2.2	1.0	2.5
Mean	**0.1**	**-0.1**	**-0.1**	**-0.1**	**0.0**	**0.0**	**0.0**	**-0.3**	**0.0**	**0.0**	**0.0**	**0.1**	**0.0**	**-0.1**	**0.0**
Min	-0.4	-0.6	-0.3	-0.6	-0.6	-1.0	-0.5	-1.0	-1.0	-0.1	-0.3	-0.4	-0.6	-1.0	-1.0

Differences in condition	FRL - SRL				FRL - REF				SRL - REF				FRL-SRL	FRL-REF	SRL-REF
Max	1.8	-0.9	-0.3	0.0	0.8	0.0	-0.8	-2.3	-1.0	0.9	-0.5	-2.3	1.2	-0.2	-1.4
Mean	**0.1**	**0.0**	**-0.1**	**0.2**	**0.1**	**0.0**	**-0.1**	**-0.2**	**0.0**	**0.0**	**0.0**	**-0.4**	**0.1**	**0.0**	**-0.1**
Min	0.2	0.4	0.2	0.5	0.6	-0.5	0.0	-0.2	0.4	-0.9	-0.2	-0.6	0.5	0.4	-0.1

Stimulus Number	Stimulus noise description
1	Power drill
2	Hammer
3	Lawn mower
4	Pressure cleaner

Table D-6. *SRL MKII: Exp. 2 – Hearing protectors – Speech quality ratings.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	REF 1	REF 2	REF 3	REF 4	Over all stimuli			
F	1	3.4	2.3	2.0	3.0	3.4	2.1	2.0	3.1	3.3	2.3	2.1	3.1				
F	2	2.0	2.8	3.1	1.7	2.0	2.8	3.2	1.8	1.9	2.8	3.3	1.8				
F	4	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0				
F	5	3.6	3.8	4.0	3.7	3.6	3.7	4.0	3.7	3.7	3.8	4.0	3.7				
F	6	4.1	2.1	4.0	4.0	4.2	2.0	4.6	3.8	4.0	2.0	4.8	4.6	Female Average			
F	7	3.5	3.1	3.0	3.0	3.5	3.1	3.0	3.0	3.5	3.1	3.0	3.0	FRL	SRL	REF	All
F	8	4.0	3.4	3.3	4.2	4.0	3.3	3.3	4.1	4.0	3.3	3.3	4.1	3.2	3.2	3.2	3.2
M	3	4.4	3.6	3.0	4.2	4.4	3.7	3.0	4.2	3.9	3.8	3.0	4.2				
M	9	3.9	4.0	4.0	4.0	4.0	4.1	4.0	4.0	4.0	4.0	4.0	3.9				
M	10	5.2	3.8	4.0	4.3	5.3	3.8	4.1	4.2	5.1	3.8	4.0	4.3				
M	11	3.0	3.0	3.0	3.0	3.0	3.0	3.0	2.9	3.0	3.0	3.0	3.0				
M	12	4.0	4.0	4.1	4.1	3.9	4.0	4.1	4.0	4.0	3.9	4.1	4.1				
M	13	4.2	3.8	4.0	4.0	4.2	3.8	4.0	4.0	4.2	3.8	4.0	4.1				
M	14	4.1	4.0	4.0	4.0	4.2	4.1	3.5	4.0	4.1	4.2	3.5	4.0	Male Average			
M	15	3.4	3.4	3.1	3.4	3.5	3.4	3.2	3.3	3.4	3.4	3.2	3.5	FRL	SRL	REF	All
M	16	4.0	3.0	3.9	4.0	3.9	3.0	4.0	4.0	4.0	3.0	4.0	4.0	3.8	3.8	3.8	3.8

Basic statistics	FRL				SRL				REF				FRL	SRL	REF
Max	5.2	4.0	4.1	4.3	5.3	4.1	4.6	4.2	5.1	4.2	4.8	4.6	5.2	5.3	5.1
Mean	3.7	3.3	3.5	3.6	3.8	3.3	3.5	3.6	3.7	3.3	3.5	3.6	3.5	3.5	3.5
Min	2.0	2.1	2.0	1.7	2.0	2.0	2.0	1.8	1.9	2.0	2.1	1.8	1.7	1.8	1.8

Differences in condition	FRL - SRL				FRL - REF				SRL - REF				FRL-SRL	FRL-REF	SRL-REF
Max	-0.1	-0.2	-0.5	0.1	0.1	-0.2	-0.7	-0.3	0.2	0.0	-0.2	-0.4	-0.1	0.1	0.2
Mean	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	-0.1	0.0	0.0	0.0
Min	0.0	0.1	0.1	-0.1	0.0	0.1	0.0	-0.1	0.0	0.0	-0.1	0.0	-0.1	-0.1	0.0

Stimulus Number	Stimulus noise description
1	Fax machine
2	Feedback
3	On hold' music
4	Fault noise

Table D-7. *SRL MKII: Exp. 3 – Telephone headsets – Speech loudness ratings.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	REF 1	REF 2	REF 3	REF 4	Over all stimuli			
F	1	6.3	6.8	6.7	7.0	4.5	4.5	4.9	5.2	4.1	5.0	3.7	4.1				
F	2	3.9	6.3	5.4	6.5	4.4	4.3	3.5	5.8	4.8	6.2	2.9	4.6				
F	4	5.7	5.6	6.2	6.5	3.3	1.9	3.6	4.0	2.9	3.3	3.0	3.0				
F	5	5.8	6.3	6.5	6.1	4.2	4.4	4.6	2.5	4.1	5.4	5.2	2.8				
F	6	7.0	5.7	7.0	7.0	4.1	1.3	6.2	5.4	5.2	3.7	5.4	4.0	Female Average			
F	7	4.5	4.8	5.1	5.0	3.5	2.1	4.0	3.0	2.9	3.0	4.0	2.4	FRL	SRL	REF	All
F	8	5.3	6.5	5.9	5.9	4.1	2.5	4.0	4.6	4.2	4.2	3.3	3.8	6.0	3.9	4.0	4.6
M	3	6.4	6.5	6.2	6.5	5.1	3.0	4.2	5.1	3.3	4.9	3.7	4.6				
M	9	6.0	4.0	5.0	6.0	3.9	2.1	4.0	4.0	4.0	4.0	4.0	3.1				
M	10	5.7	4.6	6.5	6.2	5.1	3.3	4.4	5.6	5.4	4.1	3.8	4.6				
M	11	6.0	6.0	6.0	6.0	4.1	3.0	3.0	4.0	4.1	5.0	3.0	3.0				
M	12	5.9	5.0	6.1	6.1	3.8	3.3	4.7	4.0	3.4	3.1	4.1	4.0				
M	13	4.3	3.1	5.6	5.9	4.2	3.2	4.9	4.4	4.1	4.0	4.7	3.9				
M	14	5.9	6.4	6.1	5.9	4.1	3.2	4.8	4.1	3.3	5.2	4.1	3.4	Male Average			
M	15	6.9	6.8	6.5	6.5	3.9	2.4	4.8	3.8	5.1	4.9	4.1	3.3	FRL	SRL	REF	All
M	16	7.0	7.0	5.9	6.9	5.9	3.9	4.0	5.9	4.9	5.0	4.0	5.0	5.9	4.1	4.1	4.7

Basic statistics	FRL				SRL				REF				FRL	SRL	REF
Max	7.0	7.0	7.0	7.0	5.9	4.5	6.2	5.9	5.4	6.2	5.4	5.0	7.0	6.2	6.2
Mean	5.8	5.7	6.0	6.2	4.3	3.0	4.3	4.5	4.1	4.4	3.9	3.7	5.9	4.0	4.1
Min	3.9	3.1	5.0	5.0	3.3	1.3	3.0	2.5	2.9	3.0	2.9	2.4	3.1	1.3	2.4

Differences in condition	FRL - SRL				FRL - REF				SRL - REF				FRL-SRL	FRL-REF	SRL-REF
Max	1.1	2.5	0.8	1.1	1.6	0.8	1.6	2.0	0.5	-1.7	0.8	0.9	0.8	0.8	0.0
Mean	1.5	2.7	1.7	1.8	1.7	1.3	2.1	2.5	0.1	-1.4	0.4	0.7	1.9	1.9	0.0
Min	0.6	1.8	2.0	2.5	1.0	0.1	2.1	2.6	0.4	-1.7	0.0	0.1	1.8	0.7	-1.1

Noise less speech	FRL: Noise - Speech				SRL: Noise - Speech				REF: Noise - Speech				FRL Ave	SRL Ave	REF Ave
Mean	2.1	2.4	2.6	2.7	0.5	-0.3	0.9	0.9	0.4	1.1	0.4	0.1	2.4	0.5	0.5

Stimulus Number	Stimulus noise description
1	Fax machine
2	Feedback
3	On hold' music
4	Fault noise

Table D-8. *SRL MKII: Exp. 3 – Telephone headsets – Noise loudness ratings.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	REF 1	REF 2	REF 3	REF 4
F	1	-1.1	0.0	-0.8	-0.3	-0.5	-0.1	-0.3	-0.5	-1.0	0.3	-0.8	0.0
F	2	-0.1	0.0	0.0	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	0.0	-0.1
F	4	-0.1	-0.1	-0.1	0.0	-0.1	0.0	-0.1	-0.1	-0.1	0.0	-0.1	-0.1
F	5	0.1	0.1	0.0	0.1	0.1	-0.1	-0.1	0.0	0.1	-0.2	-0.1	0.0
F	6	-0.1	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	-0.1	0.0	0.0
F	7	0.0	0.0	-0.1	-0.1	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0
F	8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	-0.1
M	3	0.2	0.1	0.1	0.0	0.3	1.0	0.1	0.0	0.2	0.7	0.1	-0.1
M	9	0.0	-0.1	-0.1	0.0	-0.1	0.0	0.0	0.0	0.0	-0.1	0.0	0.0
M	10	0.0	0.0	0.0	-0.5	0.4	0.0	0.0	-0.4	0.2	0.0	0.0	0.0
M	11	0.0	0.0	1.0	0.0	0.0	0.0	-1.0	0.0	-0.1	-0.1	-0.1	0.0
M	12	-0.1	0.0	0.0	0.0	-0.1	0.0	0.0	0.1	-0.2	-0.1	0.0	-0.1
M	13	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	0.0	-0.1	0.0	-0.1	0.0	0.0
M	14	-0.1	-0.1	0.1	-0.1	0.0	0.0	0.0	-0.1	-0.2	0.1	0.1	0.0
M	15	-0.4	-0.2	-0.1	0.0	-0.8	-0.1	0.0	0.0	-0.2	-0.1	-0.4	0.0
M	16	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1

Over all stimuli

Female Average			
FRL	SRL	REF	All
-0.1	-0.1	-0.1	-0.1

Male Average			
FRL	SRL	REF	All
0.0	0.0	0.0	0.0

Basic statistics	FRL				SRL				REF				FRL	SRL	REF
Max	0.2	0.1	1.0	0.1	0.4	1.0	0.1	0.1	0.2	0.7	0.1	0.1	1.0	1.0	0.7
Mean	-0.1	0.0	0.0	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	0.0	-0.1	0.0	0.0	0.0	0.0
Min	-1.1	-0.2	-0.8	-0.5	-0.8	-0.1	-1.0	-0.5	-1.0	-0.2	-0.8	-0.1	-1.1	-1.0	-1.0

Differences in condition	FRL - SRL				FRL - REF				SRL - REF				FRL-SRL	FRL-REF	SRL-REF
Max	-0.2	-0.9	0.9	0.0	0.0	-0.6	0.9	0.1	0.2	0.3	0.0	0.0	0.0	0.3	0.3
Mean	-0.1	-0.1	0.1	0.0	0.0	-0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Min	-0.3	-0.1	0.2	0.0	-0.1	0.0	0.0	-0.4	0.2	0.1	-0.2	-0.4	-0.1	-0.1	0.0

Stimulus Number	Stimulus noise description
1	Fax machine
2	Feedback
3	On hold' music
4	Fault noise

Table D-9. *SRL MKII: Exp.3 – Telephone headsets – Speech quality ratings.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	REF 1	REF 2	REF 3	REF 4
F	1	2.2	3.1	4.3	4.9	2.1	3.1	4.0	4.8	2.0	3.1	4.1	5.0
F	2	4.0	2.0	3.6	4.0	4.0	2.0	3.2	4.0	4.0	1.9	3.7	5.3
F	4	4.0	3.0	5.0	5.5	4.0	3.0	4.9	5.5	4.0	3.0	4.8	5.6
F	5	4.1	2.7	3.8	5.6	4.1	2.6	3.7	5.7	4.0	2.7	3.8	5.8
F	6	4.0	1.6	4.0	7.0	4.2	1.6	4.0	5.0	4.2	1.7	4.8	6.6
F	7	4.0	4.0	3.9	4.9	4.1	4.0	4.0	5.0	4.0	4.0	3.9	5.0
F	8	2.0	4.0	4.7	6.0	2.0	4.0	4.8	5.1	2.0	4.0	4.7	6.1
M	3	4.0	2.5	4.1	5.5	4.0	2.5	3.5	5.5	4.0	2.6	4.0	5.6
M	9	3.1	3.0	4.0	4.9	3.0	3.0	4.0	4.1	3.1	3.0	4.0	5.1
M	10	4.0	3.0	4.5	5.1	4.0	3.0	3.6	5.1	3.9	3.0	4.6	5.2
M	11	4.0	2.1	4.0	6.0	4.0	2.0	4.0	6.0	4.0	2.1	4.0	6.0
M	12	4.1	4.1	4.1	5.6	4.3	3.9	4.1	5.7	4.1	4.0	4.9	5.3
M	13	5.0	4.1	5.2	6.0	5.0	4.0	5.3	6.0	5.0	4.0	5.2	6.3
M	14	4.1	3.9	4.1	4.8	4.1	4.0	4.0	4.8	4.0	4.0	4.0	4.7
M	15	4.0	3.1	5.0	5.4	4.0	3.0	5.0	5.4	4.1	3.0	5.0	5.4
M	16	2.1	2.0	4.0	5.0	2.0	2.0	4.0	4.0	2.0	2.0	4.1	5.0

Over all stimuli

Female Average			
FRL	SRL	REF	All
4.0	3.9	4.1	4.0

Male Average			
FRL	SRL	REF	All
4.2	4.1	4.2	4.1

Basic statistics	FRL				SRL				REF				FRL	SRL	REF
Max	5.0	4.1	5.2	7.0	5.0	4.0	5.3	6.0	5.0	4.0	5.2	6.6	7.0	6.0	6.6
Mean	3.7	3.0	4.3	5.4	3.7	3.0	4.1	5.1	3.7	3.0	4.4	5.5	4.1	4.0	4.1
Min	2.0	1.6	3.6	4.0	2.0	1.6	3.2	4.0	2.0	1.7	3.7	4.7	1.6	1.6	1.7

Differences in condition	FRL - SRL				FRL - REF				SRL - REF				FRL-SRL	FRL-REF	SRL-REF
Max	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.4	0.0	0.0	0.1	-0.5	1.0	0.4	-0.5
Mean	0.0	0.0	0.1	0.3	0.0	0.0	-0.1	-0.1	0.0	0.0	-0.2	-0.4	0.1	0.0	-0.1
Min	0.0	0.0	0.4	0.0	0.0	-0.1	-0.1	-0.7	0.0	-0.1	-0.5	-0.7	0.0	-0.1	-0.1

Stimulus Number	Stimulus description
1	Speech @ 62 dB SPL & Noise @ 90 dB SPL
2	Speech @ 55 dB SPL & Noise @ 90 dB SPL
3	Speech @ 68 dB SPL & Noise @ 90 dB SPL
4	Speech @ 75 dB SPL & Noise @ 90 dB SPL

Table D-10. *SRL MKII: Exp. 4 – Dynamic speech level – Speech loudness ratings.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	REF 1	REF 2	REF 3	REF 4	Over all stimuli			
F	1	7.0	7.0	7.0	7.0	3.1	4.0	4.9	5.6	2.9	3.7	4.8	6.0				
F	2	6.3	6.7	6.7	6.8	4.5	1.6	4.0	5.3	2.7	1.2	3.2	6.5				
F	4	7.0	7.0	7.0	6.8	5.0	4.1	4.0	5.0	3.1	2.1	4.1	5.5				
F	5	6.7	6.3	6.6	6.7	4.3	3.4	3.1	3.8	2.1	1.5	2.1	4.5				
F	6	7.0	7.0	6.9	7.0	6.4	5.0	4.9	6.8	5.5	2.8	6.5	7.0	Female Average			
F	7	6.0	6.1	5.5	5.9	4.0	3.3	3.9	5.0	2.9	3.4	3.0	4.0	FRL	SRL	REF	All
F	8	6.6	7.0	6.8	6.8	3.0	3.9	3.6	4.0	2.0	2.0	4.0	4.0	6.7	4.3	3.7	4.9
M	3	7.0	7.0	7.0	6.5	5.0	5.0	4.7	4.9	2.9	3.9	5.3	4.8				
M	9	6.0	6.0	6.0	6.1	4.0	3.1	4.0	5.0	3.0	2.0	3.0	4.1				
M	10	6.4	5.9	6.2	6.2	3.3	3.4	4.0	5.0	3.4	2.4	4.3	5.7				
M	11	7.0	7.0	7.0	7.0	4.0	4.0	4.0	5.0	3.8	4.0	4.0	5.0				
M	12	6.9	6.9	6.9	6.8	3.5	3.2	4.4	5.3	3.1	2.7	4.1	4.6				
M	13	6.4	6.5	6.3	6.0	5.3	4.0	5.2	5.9	4.0	2.8	5.3	6.3				
M	14	6.7	7.0	6.0	6.3	4.1	4.7	3.3	3.9	3.3	4.0	3.3	4.1	Male Average			
M	15	6.9	7.0	7.0	6.9	4.5	4.5	3.5	4.9	2.7	2.8	2.8	4.1	FRL	SRL	REF	All
M	16	7.0	7.0	7.0	7.0	2.9	2.0	4.0	5.1	2.0	1.0	4.1	6.1	6.6	4.2	3.7	4.9

Basic statistics	FRL				SRL				REF				FRL	SRL	REF
Max	7.0	7.0	7.0	7.0	6.4	5.0	5.2	6.8	5.5	4.0	6.5	7.0	7.0	6.8	7.0
Mean	6.7	6.7	6.6	6.6	4.2	3.7	4.1	5.0	3.1	2.7	4.0	5.1	6.7	4.2	3.7
Min	6.0	5.9	5.5	5.9	2.9	1.6	3.1	3.8	2.0	1.0	2.1	4.0	5.5	1.6	1.0
Differences in condition	**FRL - SRL**				**FRL - REF**				**SRL - REF**				**FRL-SRL**	**FRL-REF**	**SRL-REF**
Max	0.6	2.0	1.8	0.2	1.5	3.0	0.5	0.0	0.9	0.9	-1.3	-0.2	0.2	0.0	-0.2
Mean	2.5	3.0	2.5	1.6	3.6	4.1	2.6	1.5	1.1	1.0	0.1	-0.1	2.4	2.9	0.5
Min	3.1	4.3	2.4	2.1	4.0	4.9	3.4	2.0	0.9	0.6	1.0	-0.1	3.9	4.5	0.6
Noise less speech	**FRL: Noise - Speech**				**SRL: Noise - Speech**				**REF: Noise - Speech**				**FRL Ave**	**SRL Ave**	**REF Ave**
Mean	3.0	3.7	2.3	1.2	0.5	0.7	0.0	-0.1	-0.6	-0.3	-0.4	-0.4	2.6	0.3	-0.4

Stimulus Number	Stimulus description
1	Speech @ 62 dB SPL & Noise @ 90 dB SPL
2	Speech @ 55 dB SPL & Noise @ 90 dB SPL
3	Speech @ 68 dB SPL & Noise @ 90 dB SPL
4	Speech @ 75 dB SPL & Noise @ 90 dB SPL

Table D-11. *SRL MKII: Exp.4 – Dynamic speech level – Noise loudness ratings.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	REF 1	REF 2	REF 3	REF 4	Over all stimuli			
F	1	0.0	-0.7	-0.4	-0.4	-0.4	0.0	-0.7	-0.3	-0.3	0.0	-0.1	0.0				
F	2	-0.1	-0.1	0.0	0.0	0.0	0.0	-0.6	0.0	0.0	-0.1	-0.1	-0.1				
F	4	0.0	0.0	0.0	-0.1	0.0	0.0	-1.1	-1.0	-0.1	0.0	-0.1	0.0				
F	5	0.1	-0.1	0.0	-0.2	0.1	0.0	-0.8	-0.2	0.0	0.0	-0.1	-0.5				
F	6	0.5	0.0	-1.4	-1.4	0.7	0.0	-2.3	-2.5	0.1	0.0	-0.6	0.7	Female Average			
F	7	0.0	-0.1	-0.1	-1.1	0.0	0.0	-0.9	-1.0	0.1	0.0	0.1	0.0	FRL	SRL	REF	All
F	8	0.0	0.0	0.0	0.0	0.0	0.0	-1.0	-0.8	0.0	0.0	0.0	0.0	-0.2	-0.5	0.0	-0.2
M	3	0.0	0.0	0.0	-0.6	0.0	0.0	-1.4	-1.3	-0.1	0.0	-0.4	-0.6				
M	9	0.1	-0.1	0.0	0.0	0.1	0.0	-0.9	-1.1	0.0	-0.1	0.0	0.0				
M	10	0.0	0.1	0.0	0.0	0.0	0.2	-0.8	-0.5	0.0	0.0	0.2	0.0				
M	11	0.0	0.0	0.0	0.0	0.0	0.0	-2.0	-1.0	0.0	0.0	0.0	0.0				
M	12	0.0	0.0	0.0	0.1	0.0	0.1	0.4	0.0	-0.1	0.0	0.0	0.0				
M	13	-0.1	0.0	-0.2	-1.9	-0.1	-0.1	-0.2	-1.9	-0.1	-0.1	0.0	-0.1				
M	14	-0.1	-0.1	0.1	0.0	-0.1	0.1	-0.5	-0.5	-0.1	0.0	0.0	0.0	Male Average			
M	15	0.0	-0.1	0.0	-0.3	0.0	0.0	-1.4	-1.9	0.0	0.0	0.0	0.0	FRL	SRL	REF	All
M	16	1.0	0.0	0.0	0.0	0.0	0.0	-2.0	-1.0	0.0	-0.1	0.0	0.0	-0.1	-0.5	0.0	-0.2

Basic statistics	FRL				SRL				REF				FRL	SRL	REF
Max	1.0	0.1	0.1	0.1	0.7	0.2	0.4	0.0	0.1	0.0	0.2	0.7	1.0	0.7	0.7
Mean	0.1	-0.1	-0.1	-0.4	0.0	0.0	-1.0	-0.9	0.0	0.0	-0.1	0.0	-0.1	-0.5	0.0
Min	-0.1	-0.7	-1.4	-1.9	-0.4	-0.1	-2.3	-2.5	-0.3	-0.1	-0.6	-0.6	-1.9	-2.5	-0.6
Differences in condition	**FRL - SRL**				**FRL - REF**				**SRL - REF**				**FRL-SRL**	**FRL-REF**	**SRL-REF**
Max	0.3	-0.1	-0.4	0.1	0.9	0.0	-0.2	-0.7	0.6	0.1	0.2	-0.8	0.3	0.3	0.0
Mean	0.1	-0.1	0.9	0.6	0.1	-0.1	-0.1	-0.3	0.0	0.0	-0.9	-0.9	0.4	-0.1	-0.4
Min	0.3	-0.6	0.9	0.5	0.2	-0.6	-0.7	-1.3	-0.1	0.0	-1.6	-1.9	0.5	-1.3	-1.8

Stimulus Number	Stimulus description
1	Speech @ 62 dB SPL & Noise @ 90 dB SPL
2	Speech @ 55 dB SPL & Noise @ 90 dB SPL
3	Speech @ 68 dB SPL & Noise @ 90 dB SPL
4	Speech @ 75 dB SPL & Noise @ 90 dB SPL

Table D-12. *SRL MKII: Exp. 4 – Dynamic speech level – Initial speech quality rating.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	REF 1	REF 2	REF 3	REF 4
F	1	-0.1	0.0	-0.1	0.8	-0.5	-0.8	-0.1	-0.5	-0.1	-0.4	0.0	-0.4
F	2	0.0	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	0.0	-0.1	-0.1	0.0	-0.1
F	4	-0.1	0.0	-0.1	-0.1	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	-0.1	-0.1
F	5	-0.1	0.0	-0.2	0.0	0.1	-0.1	-0.1	-0.1	0.0	0.0	0.0	-0.1
F	6	-0.1	0.0	0.9	-0.3	0.2	0.0	1.4	-0.4	0.0	0.0	2.0	-0.3
F	7	0.0	0.0	-0.1	-0.1	-0.1	-0.1	0.0	-0.1	0.0	0.0	0.0	-0.1
F	8	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.7	-0.1
M	3	1.0	0.0	-0.4	0.4	0.4	-0.2	0.0	-0.3	0.0	0.0	-0.2	0.0
M	9	-0.1	0.0	0.0	0.0	0.0	0.1	0.0	-0.1	-0.1	0.0	-0.1	0.0
M	10	0.0	0.1	0.0	0.0	0.3	0.1	0.0	-0.5	0.5	0.0	0.0	0.0
M	11	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	-0.1	0.0
M	12	0.0	0.1	0.0	0.2	0.0	0.1	0.0	0.1	0.0	0.1	-0.1	0.0
M	13	0.0	-0.1	0.0	-1.1	0.0	-0.1	0.0	-1.1	0.0	0.0	-0.1	0.0
M	14	-0.1	-0.1	0.1	-0.1	0.0	0.0	0.0	-0.1	0.1	0.0	0.0	0.1
M	15	-0.1	0.0	0.0	-0.1	0.0	0.0	0.0	-0.1	0.0	-0.2	-0.1	0.0
M	16	0.0	0.0	-0.1	-1.0	0.1	0.0	0.0	0.0	1.0	0.0	0.1	-1.0

Over all stimuli				
Female Average	FRL	SRL	REF	All
	0.0	-0.1	0.0	0.0
Male Average	FRL	SRL	REF	All
	0.0	0.0	0.0	0.0

Basic statistics	FRL				SRL				REF				FRL	SRL	REF
Max	1.0	0.1	0.9	0.8	0.4	0.1	1.4	0.1	1.0	0.1	2.0	0.1	1.0	1.4	2.0
Mean	0.0	0.0	0.0	-0.1	0.0	-0.1	0.1	-0.2	0.1	0.0	0.1	-0.1	0.0	0.0	0.0
Min	-0.1	-0.1	-0.4	-1.1	-0.5	-0.8	-0.1	-1.1	-0.1	-0.4	-0.2	-1.0	-1.1	-1.1	-1.0

Differences in condition	FRL - SRL				FRL - REF				SRL - REF				FRL-SRL	FRL-REF	SRL-REF
Max	0.6	0.0	-0.5	0.7	0.0	0.0	-1.1	0.7	-0.6	0.0	-0.6	0.0	-0.4	-1.0	-0.6
Mean	0.0	0.1	-0.1	0.1	0.0	0.0	-0.1	0.0	-0.1	0.0	-0.1	-0.1	0.0	0.0	-0.1
Min	0.4	0.8	-0.3	0.0	0.0	0.3	-0.2	-0.1	-0.4	-0.5	0.1	-0.1	0.0	-0.1	-0.1

Stimulus Number	Stimulus description
1	Speech @ 62 dB SPL & Noise @ 90 dB SPL
2	Speech @ 55 dB SPL & Noise @ 90 dB SPL
3	Speech @ 68 dB SPL & Noise @ 90 dB SPL
4	Speech @ 75 dB SPL & Noise @ 90 dB SPL

Table D-13. *SRL MKII: Exp. 4 – Dynamic speech level – Final speech quality rating.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4
F	1	3.6	1.5	2.8	4.1	2.9	1.2	1.1	3.6
F	2	2.6	2.4	2.9	4.9	2.6	2.4	2.1	4.9
F	4	2.9	1.9	1.3	3.9	2.9	2.0	1.2	3.6
F	5	3.5	2.9	1.7	3.7	3.5	3.1	1.5	3.8
F	6	4.0	3.0	1.0	6.3	4.0	2.8	1.3	5.7
F	7	4.0	3.0	3.1	4.0	4.0	3.0	3.0	3.9
F	8	3.8	3.0	2.0	4.3	3.7	3.0	2.0	4.0
M	3	4.4	3.0	2.0	5.5	3.8	2.7	1.2	5.0
M	9	4.0	1.0	1.0	4.0	4.0	1.0	1.0	4.0
M	10	4.2	2.8	3.6	5.8	3.4	3.3	1.7	5.1
M	11	4.0	3.0	3.0	5.0	4.0	3.0	2.0	5.1
M	12	3.9	4.0	4.0	4.9	4.0	3.9	4.0	4.4
M	13	3.1	4.0	3.1	4.7	3.0	4.0	3.1	4.7
M	14	4.1	3.3	3.7	5.5	4.0	3.2	3.7	5.1
M	15	3.7	2.9	3.2	5.1	3.6	3.0	2.6	4.4
M	16	4.0	2.0	1.9	5.9	4.0	2.1	2.0	5.1

Over all stimuli		
Female Average	FRL	SRL
	3.2	3.0
Male Average	FRL	SRL
	3.7	3.4

Basic statistics	FRL				SRL				FRL	SRL
Max	4.4	4.0	4.0	6.3	4.0	4.0	4.0	5.7	6.3	5.7
Mean	3.7	2.7	2.5	4.9	3.6	2.7	2.1	4.5	3.5	3.2
Min	2.6	1.0	1.0	3.7	2.6	1.0	1.0	3.6	1.0	1.0

Differences in condition	FRL - SRL				FRL-SRL
Max	0.4	0.1	0.0	0.6	0.6
Mean	0.2	0.0	0.4	0.3	0.2
Min	0.1	0.0	0.0	0.1	0.0

Stimulus Number	Stimulus description
1	Speech in a travelling car
2	Speech and hammering
3	Speech and pressure cleaner
4	Speech and alarm

Table D-14. *SRL MKII: Exp. 5 – Simultaneous noise – Speech loudness.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	Over all stimuli	
F	1	3.0	4.9	5.7	5.7	3.3	4.2	4.9	4.6		
F	2	4.0	5.6	5.9	6.0	3.0	5.0	5.7	6.0		
F	4	2.1	6.2	6.2	5.9	1.8	4.6	5.2	5.9		
F	5	4.1	6.5	6.5	5.4	2.8	4.6	5.6	5.0		
F	6	4.3	5.0	6.2	4.9	3.6	5.5	6.2	5.1	Female Average	
F	7	4.1	4.0	3.9	5.0	4.0	3.0	3.7	5.2	**FRL**	**SRL**
F	8	4.8	6.0	5.7	5.9	3.7	3.7	3.6	4.8	5.1	4.4
M	3	4.7	6.5	6.4	6.2	3.9	5.3	5.4	5.8		
M	9	4.0	5.1	6.1	5.1	3.0	2.1	4.1	4.1		
M	10	4.1	5.7	6.1	6.2	3.8	4.6	5.4	5.5		
M	11	6.0	5.0	7.0	6.0	4.0	3.0	6.0	4.0		
M	12	5.9	5.5	7.0	6.8	4.6	3.5	6.7	6.1		
M	13	2.7	5.1	6.0	5.9	3.0	4.9	5.8	6.1		
M	14	5.1	6.1	6.0	6.2	3.7	4.7	5.0	4.8	Male Average	
M	15	6.0	6.4	5.9	6.3	2.6	4.1	4.4	3.6	**FRL**	**SRL**
M	16	4.0	6.0	6.0	7.0	4.0	4.0	5.0	6.0	5.7	4.5
Basic statistics		**FRL**				**SRL**				**FRL**	**SRL**
Max		6.0	6.5	7.0	7.0	4.6	5.5	6.7	6.1	7.0	6.7
Mean		**4.3**	**5.6**	**6.0**	**5.9**	**3.4**	**4.2**	**5.2**	**5.2**	**5.5**	**4.5**
Min		2.1	4.0	3.9	4.9	1.8	2.1	3.6	3.6	2.1	1.8
Differences in condition		**FRL - SRL**								**FRL-SRL**	
Max		1.4	1.0	0.3	0.9					0.3	
Mean		**0.9**	**1.4**	**0.9**	**0.7**					**1.0**	
Min		0.3	1.9	0.3	1.2					0.3	
Noise less speech		**FRL: Noise - Speech**				**SRL: Noise - Speech**				**FRL Ave**	**SRL Ave**
Mean		**0.6**	**2.9**	**3.5**	**1.0**	**-0.2**	**1.4**	**3.1**	**0.6**	**2.0**	**1.2**

Stimulus Number	**Stimulus description**
1	Speech in a travelling car
2	Speech and hammering
3	Speech and pressure cleaner
4	Speech and alarm

Table D-15. *SRL MKII: Exp.5 – Simultaneous noise – Noise loudness.*

(maximums in red, minimums in blue)

Gender	Subject	FRL 1	FRL 2	FRL 3	FRL 4	SRL 1	SRL 2	SRL 3	SRL 4	Over all stimuli	
F	1	45	54	15	81	42	54	19	83		
F	2	61	42	45	70	52	53	45	79		
F	4	62	71	12	100	52	71	21	100		
F	5	45	44	37	70	51	40	46	63		
F	6	13	26	7	58	23	26	2	69	Female Average	
F	7	51	61	30	100	51	49	29	100	**FRL**	**SRL**
F	8	59	35	8	68	60	35	8	69	49	50
M	3	38	74	49	76	35	71	63	68		
M	9	40	41	8	49	40	41	9	70		
M	10	46	70	49	56	56	46	53	71		
M	11	50	65	11	60	70	85	26	80		
M	12	61	62	39	82	60	59	62	81		
M	13	40	84	25	93	40	84	25	93		
M	14	60	39	34	55	71	67	31	56	Male Average	
M	15	34	90	38	80	35	86	35	67	**FRL**	**SRL**
M	16	11	50	49	70	11	49	50	70	52	56
Basic statistics		**FRL**				**SRL**				**FRL**	**SRL**
Max		62	90	49	100	71	86	63	100	100	100
Mean		**45**	**57**	**29**	**73**	**47**	**57**	**33**	**76**	**51**	**53**
Min		11	26	7	49	11	26	2	56	7	2
Differences in condition		**FRL - SRL**								**FRL-SRL**	
Max		-9	4	-14	0					0	
Mean		**-2**	**0**	**-4**	**-3**					**-2**	
Min		0	0	5	-7					5	

Stimulus Number	**Stimulus description**
1	Speech in a travelling car
2	Speech and hammering
3	Speech and pressure cleaner
4	Speech and alarm

Table D-16. *SRL MKII: Exp. 5 – Simultaneous noise – Speech quality.*

(maximums in red, minimums in blue)

Appendix E: SRL file processor user's guide

SRL File Process Help

Contents

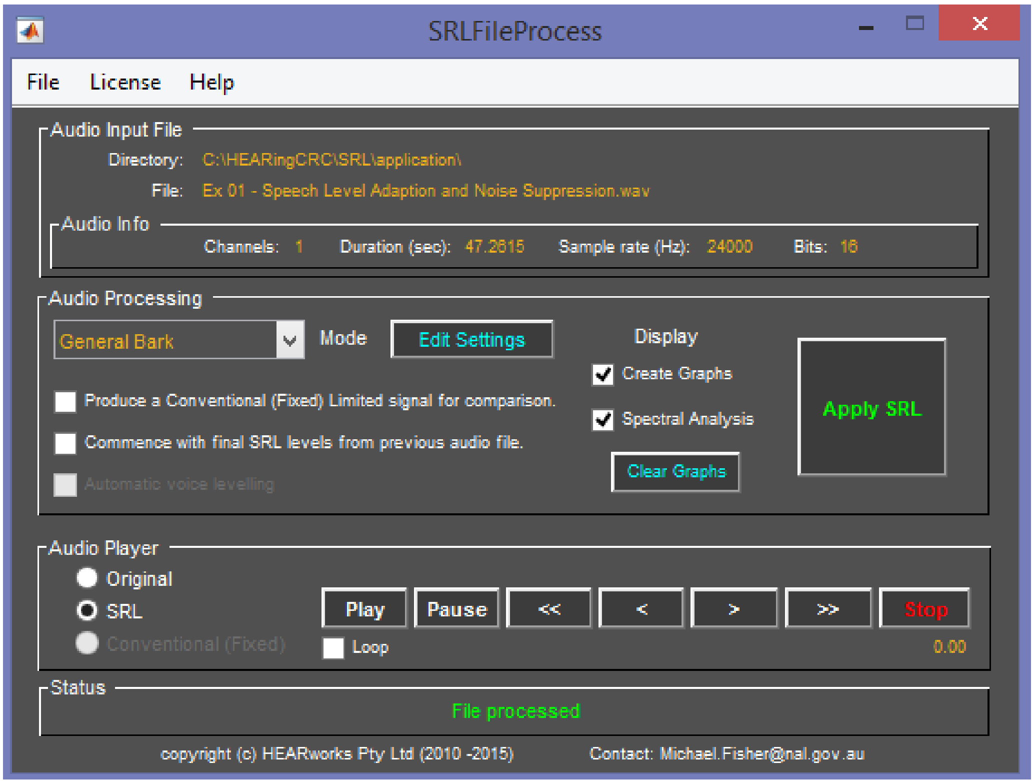
Updated 15/07/2016

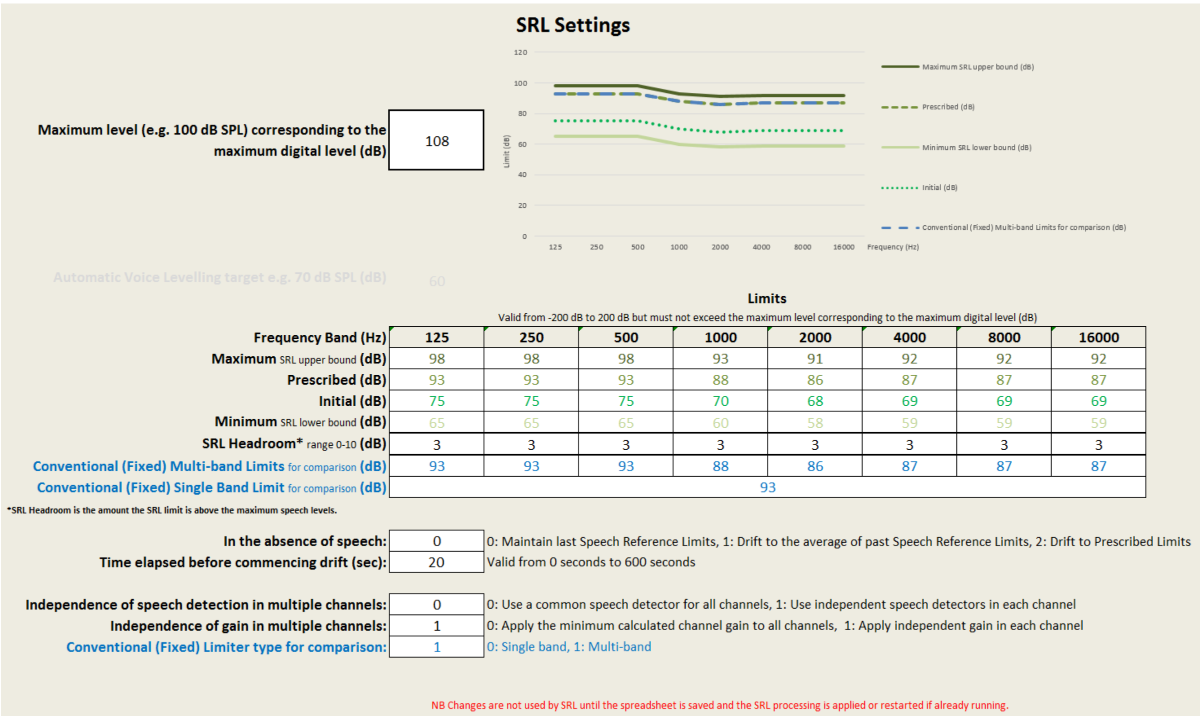
Michael J A Fisher

E: Michael.Fisher@nal.gov.au

1 Overview

This application has been created to demonstrate the features of Speech Referenced Limiting (SRL) audio control using audio files that you supply and/or audio files that are included in this demonstration package. SRL file based processing comprises of the SRL.exe app and the SRLsettings.xlsx spreadsheet, these are shown below.





Frequency Band (Hz)	125	250	500	1000	2000	4000	8000	16000
Maximum SRL upper bound (dB)	98	98	98	93	91	92	92	92
Prescribed (dB)	93	93	93	88	86	87	87	87
Initial (dB)	75	75	75	70	68	69	69	69
Minimum SRL lower bound (dB)	65	65	65	60	58	59	59	59
SRL Headroom* range 0-10 (dB)	3	3	3	3	3	3	3	3
Conventional (Fixed) Multi-band Limits for comparison (dB)	93	93	93	88	86	87	87	87
Conventional (Fixed) Single Band Limit for comparison (dB)	93							

*SRL Headroom is the amount the SRL limit is above the maximum speech levels.

In the absence of speech:	0	0: Maintain last Speech Reference Limits, 1: Drift to the average of past Speech Reference Limits, 2: Drift to Prescribed Limits
Time elapsed before commencing drift (sec):	20	Valid from 0 seconds to 600 seconds
Independence of speech detection in multiple channels:	0	0: Use a common speech detector for all channels, 1: Use independent speech detectors in each channel
Independence of gain in multiple channels:	1	0: Apply the minimum calculated channel gain to all channels, 1: Apply independent gain in each channel
Conventional (Fixed) Limiter type for comparison:	1	0: Single band, 1: Multi-band

NB Changes are not used by SRL until the spreadsheet is saved and the SRL processing is applied or restarted if already running.

2 The application

The application enables you to process wav files using SRL and also with conventional fixed-reference limiting (FRL). The application produces a display of the waveforms and enables you to play and save all the resulting audio signals. The steps of operation are:

- Opening a sound file for processing
- Selecting the appropriate processing options / settings
- Processing a sound file
- Listening to (and optionally viewing) the SRL-processed and unprocessed sound files (and optionally playing and viewing a FRL-processed file).
- Saving a SRL-processed sound file (optionally saving a FRL-processed file)

3 Opening a sound file for processing

Wave files (*.wav) may be opened by selecting File →Open Input File and finding the appropriate directory and name of the file to be opened. SRL will open *.wav files with any number of channels with any of the following sample rates:

- 8 kHz
- 11.025 kHz
- 12 kHz
- 16 kHz*
- 22.05 kHz
- 24 kHz*
- 32 kHz*
- 44.1 kHz
- 48 kHz*

*These are also internal processing rates. Files with sampling rates other than these are converted to the next highest internal processing rate for processing, listening and viewing and are converted back to their original sampling rate when saved.

4 Processing a sound file

Once an input file is loaded, it may be processed by clicking on the **Apply SRL** button. There are several options that effect the processing these are as follows:

> **Mode** This determines the mode of processing to the applied, such as the use of different frequency scales, e.g. Bark, ERB, Third Octave and Linear.

☐ **Produce a Conventional (Fixed) Limited signal for comparison.** This will result in the application producing two sets of output signals, one using SRL and one using conventional fixed-reference limiting (FRL). Both these signals can be played, displayed, analysed and saved as wav files.

☐ **Commence with final SRL levels from previous audio file.** If selected, the starting speech reference limits will be the last speech reference limits generated when processing the previous sound file, rather than the user defined set of limits specified in the *SRLsettings.xlsx* spreadsheet. This option is useful when the sound files represent a continuous or related set of sounds that are in separate files. This may be appropriate when testing subjects using a sequence of related sound files. If you need the speech reference limits appropriately set at the beginning of a sound file, then process the file twice with this option selected for the 2nd time you apply the processing.

☐ **Automatic voice levelling.** Disabled.

Edit Settings Clicking on this button will open the spreadsheet ***SRLsettings.xlsx*** using Excel (assuming you have it installed). This spreadsheet contains detailed settings for the application and these are read each time the application is applied to a wav file. See the section on the **SRL settings spreadsheet** for further details.

When applying SRL you can have it produce graphs. The graphing options are:

☐ **Create Graphs** At its most basic level this option will display the waveforms of the original (input) signal overlayed with the SRL-processed signal. It will also display the waveforms of the original (input) signal overlayed with its conventional fixed-reference limited (FRL) signal should the conventional (fixed) limiting option have been selected.

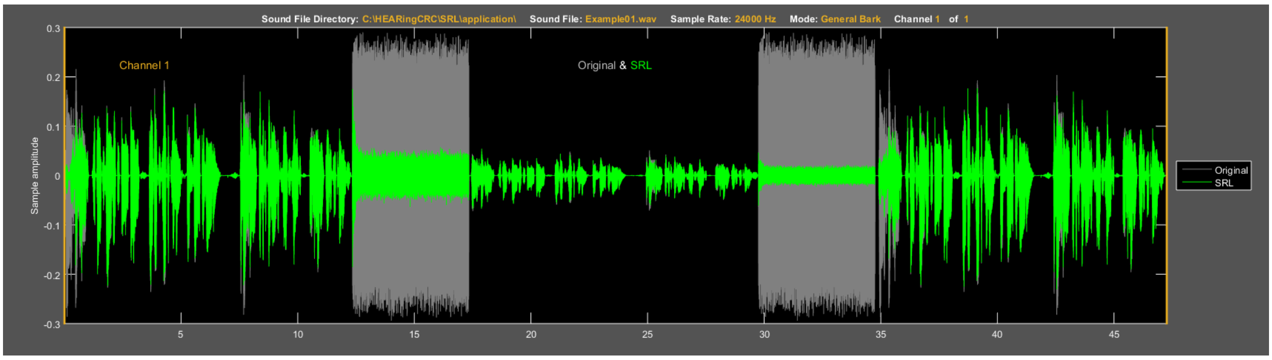
☐ **Spectral Analysis** This option will display the spectrograms of the original (input) signal and the SRL-processed signal and a spectrogram of the attenuation applied by SRL. It will also display a spectrogram of the conventional fixed-reference limited signal and the attenuation applied by the conventional fixed-reference limiter should the conventional (fixed) limiting option have been selected.

Clear Graphs Clicking on this button will clear all the graphs that have been previously produced.

5 The displays

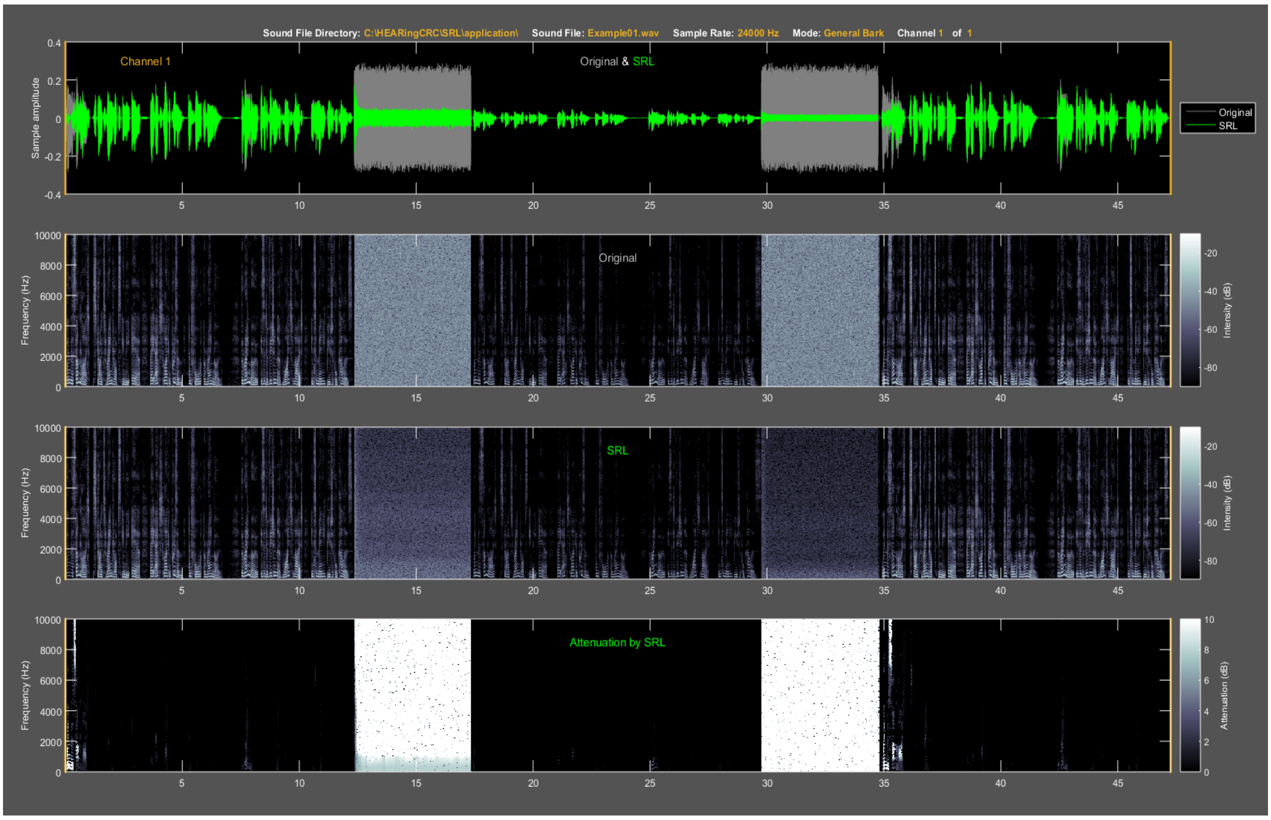
Four types of displays are produced:

1. This is a display of the original waveform (light grey) overlayed with the SRL-processed waveform (bright green).

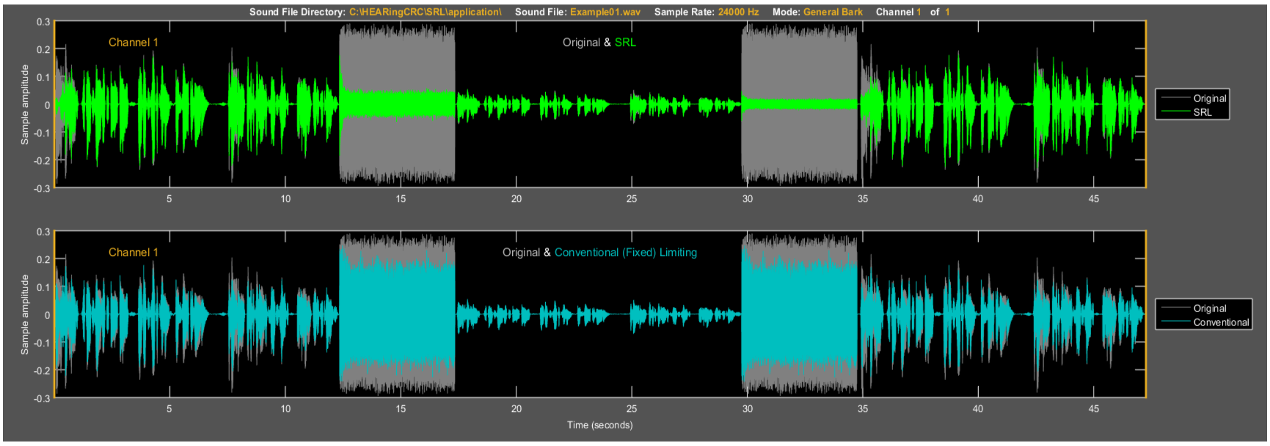


2. This is a display of the original waveform (light grey) overlayed with the SRL-processed waveform (bright green) plus spectrograms of the original signal, the SRL-processed signal, and the attenuation produced by SRL.

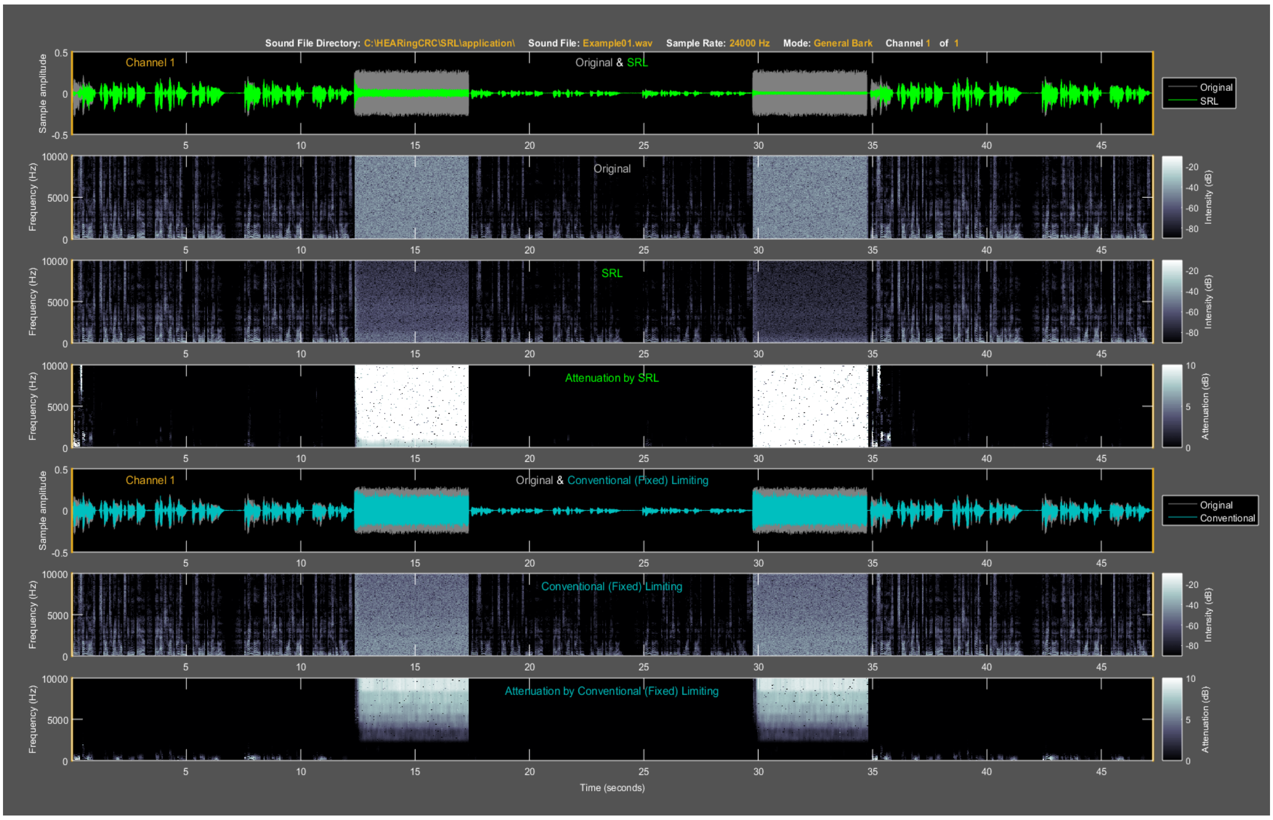
   Note the frequency range of the spectrogram is from 0 Hz to half the sampling frequency or 10 kHz, whichever is lower. The spectrogram's analysis time window is 40 milliseconds and the window overlap is 75%. Also note the maximum attenuation displayed is 10 dB, attenuation greater than this is simply shown as 10 dB.



3. This is a display of the original waveform (light grey) overlayed with the SRL-processed waveform (bright green) and the original waveform overlayed with the conventional (fixed) reference limiting processed waveform (bright blue).



4. This is a display of the original waveform (light grey) overlayed with the SRL-processed waveform (bright green) and the original waveform overlayed with conventional (fixed) reference limiting processed waveform (bright blue) plus spectrograms of the original signal, the SRL-processed signal, the attenuation produced by SRL, the conventional (fixed) reference limit processed signal, and the attenuation produced by the conventional (fixed) reference limiting.



5.1 Display Control

5.1.1 Save and Print

Display saving and printing of the display can be performed by clicking on the file and printer symbols. Further export options are available under the ‘File’ tab.

5.1.2 Zoom

Zooming in on a plot region can be performed by clicking on the positive magnifying glass symbol and highlighting the plot region to be zoomed in on. Zooming out can be performed by clicking on the negative magnifying symbol and then clicking on a previously zoomed plot.

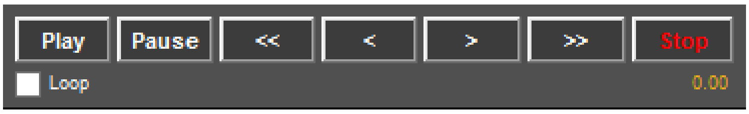
5.1.3 Scroll

Scrolling a plot, either horizontally or vertically, can be performed by clicking on the hand symbol and then clicking and holding it over a plot and moving it in the desired scrolling direction.

6 Playing sound files

The audio player will use the default sound output of the computer on which the application is run.

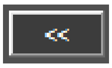
6.1 Standard audio player controls



These are as follows:

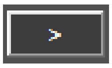
Play from the position of the file cursor

Pause play



Move the file cursor **2 seconds backwards**

Move the file cursor **0.1 seconds backwards**



Move the file cursor **0.1 seconds forwards**

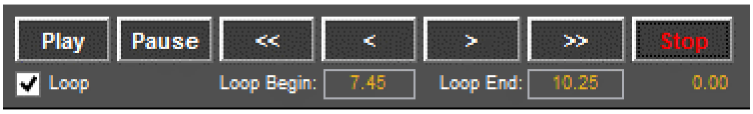
Move the file cursor **2 seconds forwards**

Stop play and return the cursor to the beginning of the file.

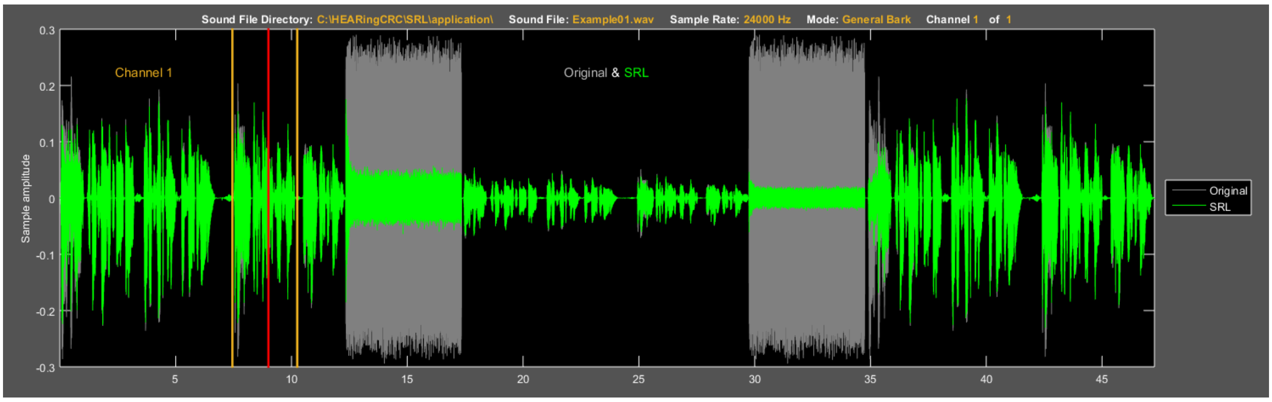
The position of the cursor is displayed in seconds in the box on the right 0.00.

6.2 Looping a section of audio

Clicking on the **☑ Loop** check box will bring up the **Loop Begin** and **Loop End** boxes. The loop beginning and ending times are in seconds. They are initially set to the beginning and end of the sound file. They may be edited to specify the desired loop beginning and ending as shown below.

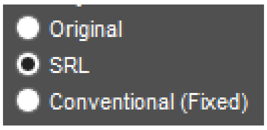


If a graph has been created, then the **Loop Begin** and **Loop End** points will be displayed as yellow vertical bars, as shown in the following figure. The moving cursor is displayed as a vertical red bar.



The looping function is particularly useful in investigating the change in the sound over a small time region.

6.3 Playback signal selection / comparison



The signal being played back can be switched at any time between original (input) signal and the SRL-processed signal. Combined with the looping function this enables focused investigation of the effect of SRL on the signal. If a conventional (fixed) limited signal has also been produced, then a three-way comparison can be made.

7 Saving a processed sound file

7.1 Saving a SRL-processed file

The SRL-processed signal maybe saved by selecting

File → Save SRL Processed File

The default file name and format will be identical to the original sound file but with the prefix SRL_.

7.2 Saving a FRL-processed file

The conventional (fixed) reference limiting, FRL-processed signal maybe saved by selecting

File → Save FRL Processed File

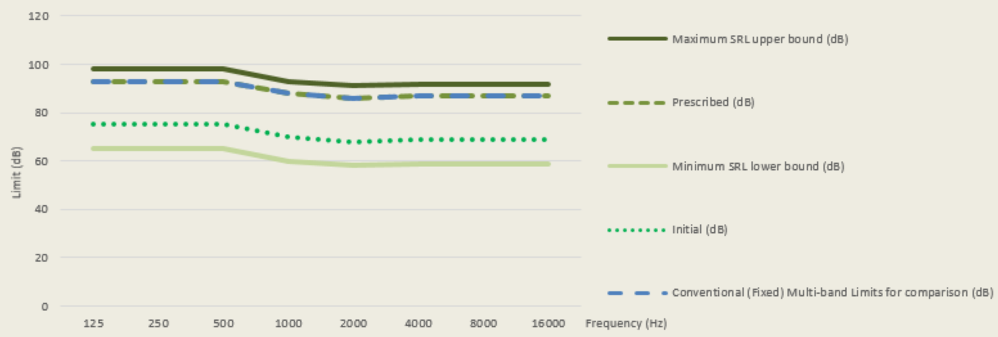
The default file name and format will be identical to the original sound file but with the prefix FRL_.

8 SRL settings spreadsheet

The SRL Settings Excel spreadsheet, ***SRLsettings.xlsx*** is shown in the figure on the following page. Assuming Excel is installed ***SRLsettings.xlsx*** may be opened by clicking on the **Edit Settings** button.

IMPORTANT: Modified settings will not be used until the spreadsheet is saved and SRL is applied or re-applied.

SRL Settings



Maximum level (e.g. 100 dB SPL) corresponding to the maximum digital level (dB): 108

Automatic Voice Levelling target e.g. 70 dB SPL (dB): 60

Limits

Valid from -200 dB to 200 dB but must not exceed the maximum level corresponding to the maximum digital level (dB)

Frequency Band (Hz)	125	250	500	1000	2000	4000	8000	16000
Maximum SRL upper bound **(dB)**	98	98	98	93	91	92	92	92
Prescribed (dB)	93	93	93	88	86	87	87	87
Initial (dB)	75	75	75	70	68	69	69	69
Minimum SRL lower bound **(dB)**	65	65	65	60	58	59	59	59
SRL Headroom* range 0-10 **(dB)**	3	3	3	3	3	3	3	3
Conventional (Fixed) Multi-band Limits for comparison **(dB)**	93	93	93	88	86	87	87	87
Conventional (Fixed) Single Band Limit for comparison **(dB)**	93							

*SRL Headroom is the amount the SRL limit is above the maximum speech levels.

In the absence of speech:	0	0: Maintain last Speech Reference Limits, 1: Drift to the average of past Speech Reference Limits, 2: Drift to Prescribed Limits
Time elapsed before commencing drift (sec):	20	Valid from 0 seconds to 600 seconds

Independence of speech detection in multiple channels:	0	0: Use a common speech detector for all channels, 1: Use independent speech detectors in each channel
Independence of gain in multiple channels:	1	0: Apply the minimum calculated channel gain to all channels, 1: Apply independent gain in each channel
Conventional (Fixed) Limiter type for comparison:	1	0: Single band, 1: Multi-band

NB Changes are not used by SRL until the spreadsheet is saved and the SRL processing is applied or restarted if already running.

8.1 Maximum level

Maximum level (e.g. 100 dB SPL) corresponding to the maximum digital level (dB)	108

Sound files have a maximum level, known as digital saturation, which is normally assigned a level of 0 dB and therefore the level of all other signals is negative; their value in dB represents how many dB they are below digital saturation. This maximum level can also be assigned to another level in dB and this can be used as the upper level for the processing. This therefore becomes the reference for all other levels. For example, the maximum may correspond to 110 dB SPL in a hearing aid, 90 dB SPL for a speech recording or 3 dBm0 in a telephone system. Consider, for example, a wav file that contains a recording of speech with a long-term level that is 30 dB below the maximum level of the sound file (i.e. at -30 dB). Say for example this speech when recorded had an average level of 65 dB SPL then the maximum level of the sound file (i.e. 0 dB) is 65 dB SPL + 30 dB = 95 dB SPL. Therefore, the maximum level of the file corresponds to 95 dB SPL and 95 is placed in the box. For a sound recording which is not referenced to any particular sound pressure level the maximum level can be set to 0 dB as is the case with most sound editors.

8.2 Limits

Limits

Valid from -200 dB to 200 dB but must not exceed the maximum level corresponding to the maximum digital level (dB)

Frequency Band (Hz)	125	250	500	1000	2000	4000	8000	16000
Maximum SRL upper bound (dB)	95	95	95	95	95	95	95	95
Prescribed (dB)	90	90	90	90	90	90	90	90
Initial (dB)	70	70	70	70	70	70	70	70
Minimum SRL lower bound (dB)	60	60	60	60	60	60	60	60
SRL Headroom* range 0-10 (dB)	3	3	3	3	3	3	3	3
Conventional (Fixed) Multi-band Limits for comparison (dB)	90	90	90	90	90	90	90	90
Conventional (Fixed) Single Band Limit for comparison (dB)	90							

*SRL Headroom is the amount the SRL limit is above the maximum speech levels.

While SRL's limits are adaptive and determined by the speech levels they each can be bounded within a range from a **Minimum** to a **Maximum**. They can start at an **Initial** values and in the absence of speech over time they can drift to **Prescribed** values. SRL's adaptive limits can be identical to the maximum levels of speech or can be a little higher by an amount equal to the specified **SRL Headroom** values. Having a slightly higher limit than the speech levels enables noise and non-detected speech to be a little higher in level than the preceding speech.

While SRL typically operates using a fine frequency structure, such as an approximation to the Bark scale, the limits specified here are at octave centre frequencies for ease of entry. These are interpolated to form limits for the finer frequency structure of SRL. Depending on the application the limits can be shaped. For example, they can be shaped to follow the typical speech spectrum for controlling the sound from a flat response microphone or they can follow the typically high-frequency emphasised response when used on a hearing-aid amplified signal.

The SRL limits are:

Maximum SRL upper bound **(dB)** These are the maximum allowable limits. These limits must not exceed the overall Maximum level. If there is no upper bound to the limiting, then all the values equal to the overall Maximum level.

Minimum SRL lower bound **(dB)** These are the minimum allowable limits. They must not exceed the Maximum limits. If there is no lower bound to the limiting, then all the values equal -200.

Initial (dB) These are the initial SRL limits, they should be set to the expected speech levels. These initial values must not be above the Maximum limits or be below the Minimum limits.

Prescribed (dB) If the option to 'Drift to Prescribed Limits' for 'In the absence of speech' is enabled, then these prescribed limits are used when speech has been absent for the period specified by 'Time elapsed before commencing drift (sec)'.

SRL Headroom* range 0-10 **(dB)** SRL's adaptive limits can be identical to the maximum levels of speech or can be a little higher as specified by the SRL Headroom values. Having a slightly higher limit than the speech levels enables noise and non-detected speech to be a little higher in level than then preceding speech. For really tight control of noise, at the expense of limiting brief increases in the maximum speech levels, these values should be set to 0 dB. A typical trade-off is to set these values to 2 or 3 dB.

The Conventional (Fixed) Limits are:

Conventional (Fixed) Multi-band Limits for comparison **(dB)** These are the limits used with a conventional (fixed) reference limiter. This limiter is identical in structure to SRL but uses fixed multiband limits. It is purely there to demonstrate the difference between having speech referenced adaptive limits and conventional fixed-reference limits. The limits must not exceed the overall Maximum level.

Conventional (Fixed) Single Band Limits for comparison **(dB)** This is a single-band limit used with a conventional fixed-reference single-band limiter. The structure is different from SRL in that it only has one broadband gain control. It is there to demonstrate the difference between a speech referenced adaptive multiband limiter and a conventional fixed-reference single-band limiter. The limit must not exceed the overall Maximum level.

8.3 Operational settings

The operational settings are:

In the absence of speech: This parameter decides what happens in the absence of speech. There are three options:

Setting: '0' - **Maintain last Speech Reference Limits.** This is mainly appropriate when testing the effect that the speech reference limits have on non-speech signals (or difficult to detect speech) over an extended duration.

Setting: '1' – **Drift to the average of past Speech Reference Limits.** This is appropriate when assessing SRL's application to a speech communications system, such as a telephone or two-way radio. After speech has been absent for some time it will limit non-speech sounds to the average level of past speech reference limits. Speech that occurs after a period of non-speech is quite likely to come from a new talker and therefore the average of past speech reference limits is more a likely match to the new speech than the last speech reference limits.

Setting '2' – **Drift to Prescribed Limits.** This is appropriate when assessing SRL's application to devices whose sound input comes from the field, such as a hearing-aids, cochlear implants, and level-dependent hearing-protectors. In this case, there may be extended periods in which there is no speech and therefore a set of fixed limits is more appropriate. In the case of a hearing-aid or cochlear implant these are the limits normally prescribed to prevent loudness discomfort for music and environmental sounds. In the case of level-dependent hearing-protectors these are the limits set to prevent loudness discomfort and excessive exposure.

Time elapsed before commencing drift (sec): This parameter determines the time before drifting the speech reference limits, to either: the average of past speech reference limits or to the prescribed limits, after speech becomes absent.

Independence of speech detection in multiple channels: There are two options:

Setting: ‘0’ – **Use a common speech detector for all channels.** This is appropriate when assessing SRL’s potential for use with some multiple channel devices such as some stereo communications headsets and some talk-thru hearing-protectors incorporating a microphone signal link between the ears. In this case it may be either desirable or more efficient to determine if speech is present using the combined left and right signal. Other examples are surround sound systems, such as 5.1, which may also benefit from a common SRL speech detector being used on the combined channel signal.

Setting: ‘1’ - **Use independent speech detectors in each channel.** This is appropriate when evaluating SRL’s potential for use with devices that have independent channels or there are multiple devices, such as bilaterally-fitted conventional hearing-aids and cochlear implants without a communications link between them. Further examples are multichannel sound processors in which the channels contain different talkers possibly talking simultaneously.

Independence of gain in multiple channels: There are two options:

Setting: ‘0’ – **Apply the minimum calculated channel gain to all channels.** This is an appropriate setting when assessing SRL’s potential for use multiple channel devices, such as hearing protectors or multiple devices, such as bilaterally-fitted hearing-aids and cochlear implants with a communications link between the left and right ear processors. In this case the same reduction in gain due to limiting can be applied to both channels equally and therefore localization cues are better preserved (ignoring secondary effects due to asymmetric recruitment).

Setting: ‘1’ - **Apply independent gain in each channel.** This is an appropriate setting when assessing SRL’s potential to be used with devices that have multiple independent channels or there are multiple devices, such as bilaterally-fitted conventional hearing-aids and cochlear implants without a communications link between them. This setting may also be desirable even though the channels can be linked. For example, with bilaterally fitted devices a user will obtain a better signal to noise ratio from channel independence when the noise from one side of the head exceeds the speech from the other.

Conventional (Fixed) Limiter type for comparison: There are two options for the type of limiter used as a comparison:

Setting: 0 - **Single band.** A single-band limiter is used for comparison. The single-band limiter is still indicative of many limiters used today in hearing protectors and sound processors.

Setting: 1 - **Multi-band**. A multiband limiter (with the same number of channels as the SRL scheme) is used for comparison. A multiband limiter is indicative of most limiters used today in hearing aids and cochlear implants.

9 License

9.1 Opening the license document

The SRL license document may be opened by selecting the menus bar tab

License → Open License Document

This will open the text file SRL_LICENSE.txt using Notepad.

Bibliography

1. Fisher M, Dillon H. Psychoacoustic and audio signal-processing aspects of acoustic shock protection. In: Milhinch JC, ed. *Risking acoustic shock seminar*. Perth, Australia 2001.

2. Milhinch JC. Acoustic shock injury: Real or imaginary. *Audiology Online.* America Audiology Network. http://www.audiologyonline.com/articles/acoustic-shock-injury-real-or-1172 2002.

3. Cole R. The history of acoustic shock in Australia: Regulators and standards. In: Milhinch JC, ed. *Risking acoustic shock seminar*. Perth, Australia 2001.

4. Dillon H, Fisher M. Acoustic shock. *Acoustics Australia.* 2002; 30(2):67-70.

5. Fisher M, Inventor; Hear IP Pty Ltd, assignee. AU 2002318963 B2, Digital signal processing system and method for a telephony interface apparatus 2007.

6. Patel JA, Broughton K. Assessment of the noise exposure of call centre operators. *Annals of Occupational Hygiene.* 2002; 46(8):653-661.

7. Fisher MJA, Inventor; HearIP Pty Ltd, assignee. AU2003209821 Method and system for controlling potentially harmful signals in a signal arranged to convey speech 2006.

8. Fisher MJA, Inventor; HearIP Pty Ltd, assignee. US7,565,283, Method and system for controlling potentially harmful signals in a signal arranged to convey speech 2009.

9. Fisher MJA, Inventor; HearIP Pty Ltd, assignee. ZL03805824.3 Method and system for controlling potentially harmful signals in a signal arranged to convey speech 2006.

10. Fisher MJA, Inventor; HearIP Pty Ltd, assignee. IN254168 Method and system for controlling potentially harmful signals in a signal arranged to convey speech 2012.

11. Fisher MJA, Inventor; HearIP Pty Ltd, assignee. GB2401765 Method and system for controlling potentially harmful signals in a signal arranged to convey speech 2006.

12. ANSI. ANSI S3.5 Methods for Calculation of the Speech Intelligibility Index. Washington, USA: American National Standards Institute; 1997.

13. Fletcher H, Munson WA. Loudness, Its Definition, Measurement and Calculation. *Bell System Technical Journal.* 1933; 12(4):377-430.

14. Florentine M, Fay RR, Popper AN. *Loudness*. New York, USA: Springer; 2011.

15. Fechner GT. *Elements of psychophysics*, 1860.

16. Stevens SS. To honour Fechner and repeal his law: A power function, not a log function, describes the operating characteristic of a sensory system. *Science (New York, NY).* 1961; 133(3446):80-86.

17. Stevens SS. Calculating loudness. *Noise control.* 1957; 3(5):11-22.

18. Stevens SS. The measurement of loudness. *The Journal of the Acoustical Society of America.* 1955; 27(5):815-829.

19. Stevens SS. A scale for the measurement of a psychological magnitude: loudness. *Psychological Review.* 1936; 43(5):405.

20. Couling J. TV Loudness: time for a new approach? *Dolby Laboratories, Inc. England, http://www. dolby. com/uploadedFiles/zz-—SharedAssets/English—PDFs/Professiona1/34—TV—Loudness—AES. pdf.* 2003.

21. ITU. ITU-R BS. 1770-2. Algorithms to measure audio programme loudness and true-peak audio level. Geneva, Switzerland: International Telecommunications Union; 2011.

22. Zwicker E, Scharf B. A model of loudness summation. *Psychological review.* 1965; 72(1):3.

23. Zwicker E, Fastl H. *Psychoacoustics: Facts and Models.* Berlin Heidelberg, Germany: Springer-Verlag; 1990.

24. Moore BC, Glasberg BR, Baer T. A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society.* 1997; 45(4):224-240.

25. Fastl H, Zwicker E. *Psychoacoustics: facts and models.* Berlin Heidelberg, Germany: Springer Verlag; 2006.

26. Florentine M, Epstein M. TO HONOR STEVENS AND REPEAL HIS LAW (FOR THE AUDITORY STSTEM). *Proceedings of Fechner Day.* 2006; 22(1):37-42.

27. ISO. International Standard ISO 226. Normal Equal-Loudness Level Contours. Geneva, Switzerland: International Organization for Standardization; 2003.

28. Zwislocki J. Theory of temporal auditory summation. *The Journal of the Acoustical Society of America.* 1960; 32(8):1046-1060.

29. Zwislocki JJ. Temporal summation of loudness: An analysis. *The Journal of the Acoustical Society of America.* 1969; 46(2B):431-441.

30. Buus S, Florentine M, Poulsen T. Temporal integration of loudness, loudness discrimination, and the form of the loudness function. *The Journal of the Acoustical Society of America.* 1997; 101(2):669-680.

31. Florentine M, Poulsen T. Temporal integration of loudness as a function of level. *The Journal of the Acoustical Society of America.* 1996; 99(3):1633-1644.

32. Katz J. *Handbook of clinical audiology.* Baltimore, USA: Lippincott Williams & Wilkins; 2002.

33. Shaw E. Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *The Journal of the Acoustical Society of America.* 1974; 56(6):1848-1861.

34. Shaw E, Teranishi R. Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source. *The Journal of the Acoustical Society of America.* 1968; 44(1):240-249.

35. Shaw EA. The external ear. *Auditory System*. Springer; 1974: 455-490.

36. Shaw E. Earcanal pressure generated by a free sound field. *The Journal of the Acoustical Society of America.* 1966; 39(3):465-470.

37. Carlile S. *The physical and psychophysical basis of sound localization*: Springer; 1996.

38. Brungart DS, Rabinowitz WM. Auditory localization of nearby sources. Head-related transfer functions. *The Journal of the Acoustical Society of America.* 1999; 106(3):1465-1479.

39. Brungart DS, Durlach NI, Rabinowitz WM. Auditory localization of nearby sources. II. Localization of a broadband source. *The Journal of the Acoustical Society of America.* 1999; 106(4):1956-1968.

40. Åhs F, Dunsmoor JE, Zielinski D, LaBar KS. Spatial proximity amplifies valence in emotional memory and defensive approach-avoidance. *Neuropsychologia.* 2015; 70:476-485.

41. Epstein M, Florentine M. Binaural loudness summation for speech presented via earphones and loudspeaker with and without visual cues. *The Journal of the Acoustical Society of America.* 2012; 131(5):3981-3988.

42. Rajan R, Johnstone B. Contralateral cochlear destruction mediates protection from monaural loud sound exposures through the crossed olivocochlear bundle. *Hearing research.* 1989; 39(3):263-277.

43. Kulkarni A, Colburn HS. Variability in the characterization of the headphone transfer-function. *The Journal of the Acoustical Society of America.* 2000; 107(2):1071-1074.

44. Zwicker E, Flottorp G, Stevens SS. Critical band width in loudness summation. *The Journal of the Acoustical Society of America.* 1957; 29(5):548-557.

45. Zwicker E. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *The Journal of the Acoustical Society of America.* 1961; 33(2):248-248.

46. Moore BC, Glasberg BR. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America.* 1983; 74(3):750-753.

47. ISO. International Standard ISO 532, Acoustics - Method for calculating loudness level. Geneva, Switzerland: International Organization for Standardization; 1975.

48. Glasberg BR, Moore BC. A model of loudness applicable to time-varying sounds. *Journal of the Audio Engineering Society.* 2002; 50(5):331-342.

49. Chalupper J, Fastl H. Dynamic loudness model (DLM) for normal and hearing-impaired listeners. *Acta Acustica united with Acustica.* 2002; 88(3):378-386.

50. ANSI. ANSI S3.4 Procedure for the Computation of Loudness of Steady Sounds. Washington, USA: American National Standards Institute; 2007.

51. Schlittenlacher J, Ellermeier W, Hashimoto T. Spectral loudness summation: Shortcomings of current standards. *The Journal of the Acoustical Society of America.* 2015; 137(1):EL26-EL31.

52. Anweiler A-K, Verhey JL. Spectral loudness summation for short and long signals as a function of level. *The Journal of the Acoustical Society of America.* 2006; 119(5):2919-2928.

53. Verhey JL, Kollmeier B. Spectral loudness summation as a function of duration. *The Journal of the Acoustical Society of America.* 2002; 111(3):1349-1358.

54. Röhl M, Kollmeier B, Uppenkamp S. Spectral loudness summation takes place in the primary auditory cortex. *Human brain mapping.* 2011; 32(9):1483-1496.

55. Hots J, Rennies J, Verhey J. Loudness of subcritical sounds as a function of bandwidth, center frequency, and level. *The Journal of the Acoustical Society of America.* 2014; 35(3):1313-1320.

56. IEC. International Standard IEC 61672-1:2013. Sound level meters - Part 1: Specifications. Geneva, Switzerland: International Electrotechnical Commission; 2013.

57. Soulodre GA, Lavoie MC. Development and Evaluation of Short-Term Loudness Meters. Paper presented at: Audio Engineering Society Convention 121 2006.

58. Zwicker E. Procedure for calculating loudness of temporally variable sounds. *The Journal of the Acoustical Society of America.* 1977; 62(3):675-682.

59. Rennies J, Verhey JL, Fastl H. Comparison of loudness models for time-varying sounds. *Acta Acustica united with Acustica.* 2010; 96(2):383-396.

60. Rennies J, Holube I, Verhey JL. Loudness of speech and speech-like signals. *Acta Acustica united with Acustica.* 2013; 99(2):268-282.

61. Rennies J, Wächtler M, Hots J, Verhey J. Spectro-Temporal Characteristics Affecting the Loudness of Technical Sounds: Data and Model Predictions. *Acta Acustica united with Acustica.* 2015; 101(6):1145-1156.

62. Halpern DL, Blake R, Hillenbrand J. Psychoacoustics of a chilling sound. *Perception & Psychophysics.* 1986; 39(2):77-80.

63. Byrne D, Dillon H, Tran K, et al. An international comparison of long-term average speech spectra. *The Journal of the Acoustical Society of America.* 1994; 96(4):2108-2120.

64. Coren S. Most comfortable listening level as a function of age. *Ergonomics.* 1994; 37(7):1269-1274.

65. Rogers DS, Harkrider AW, Burchfield SB, Nabelek AK. The influence of listener's gender on the acceptance of background noise. *Journal of the American Academy of Audiology.* 2003; 14(7):372-382.

66. Kellaris JJ, Rice RC. The influence of tempo, loudness, and gender of listener on responses to music. *Psychology & Marketing.* 1993; 10(1):15-29.

67. Thomas J, Jones D. Individual differences in noise annoyance and the uncomfortable loudness level. *Journal of Sound and Vibration.* 1982; 82(2):289-304.

68. Kardous CA, Morata TC. Occupational and recreational noise exposures at stock car racing circuits: An exploratory survey of three professional race tracks. *Noise Control Engineering Journal.* 2010; 58(1):54-61.

69. Fastl H. Audio-visual interactions in loudness evaluation. *Proceedings of the 18th International Congress on Acoustics.* 2004.

70. Parizet E, Koehl V. Influence of train colour on loudness judgments. *Acta Acustica united with Acustica.* 2011; 97(2):347-349.

71. Bregman AS. *Auditory scene analysis: The perceptual organization of sound*. MIT press; 1994.

72. Zendel BR, Alain C. Concurrent sound segregation is enhanced in musicians. *Journal of Cognitive Neuroscience.* 2009; 21(8):1488-1498.

73. Møller AR. *Hearing: its physiology and pathophysiology*. Academic Press; 2000.

74. Djupesland G. Advanced reflex considerations. *Handbook of clinical impedance audiometry. J Jerger.* New York, USA, American Electromedics Corp. 1975: 85-126.

75. Ramírez LM, Ballesteros LE, Pablo Sandoval G. Tensor tympani muscle: strange chewing muscle. *Medicina Oral, Patología Oral y Cirugía Bucal (Internet).* 2007; 12(2):96-100.

76. Liden G, Nordlund B, Hawkins J. Significance of the stapedius reflex for the understanding of speech. *Acta Oto-Laryngologica.* 1964; 57(sup188):275-279.

77. Puria S, Peake WT, Rosowski JJ. Sound-pressure measurements in the cochlear vestibule of human-cadaver ears. *The Journal of the Acoustical Society of America.* 1997; 101(5):2754-2770.

78. Klockhoff I, Anderson H. Reflex activity in the tensor tympani muscle recorded in man: Preliminary report. *Acta oto-laryngologica.* 1960; 51(1-2):184-188.

79. Mukerji S, Windsor AM, Lee DJ. Auditory brainstem circuits that mediate the middle ear muscle reflex. *Trends in amplification.* 2010: 1084713810381771.

80. Marsh R, Hoffman HS, Stitt CL. Temporal integration in the acoustic startle reflex of the rat. *Journal of Comparative and Physiological Psychology.* 1973; 82(3):507.

81. Lang PJ, Davis M, Öhman A. Fear and anxiety: animal models and human cognitive psychophysiology. *Journal of affective disorders.* 2000; 61(3):137-159.

82. Jewett DL, Williston JS. Auditory-evoked far fields averaged from the scalp of humans. *Brain.* 1971; 94(4):681-696.

83. Djupesland G, Zwislocki J. Effect of temporal summation on the human stapedius reflex. *Acta oto-laryngologica.* 1971; 71(1-6):262-265.

84. Salomon G, Starr A. Electromyography of middle ear muscles in man during motor activities. *Acta Neurologica Scandinavica.* 1963; 39(2):161-168.

85. Klockhoff I, Westerberg C. The tensor tympani muscle and tension headache. *Proc Ann Scand Migraine Soc.* 1971; 3(Suppl. 1).

86. Klockhoff I, Lindholm L, Westerberg C. "Spontaneous impedance fluctuation--a" tensor tympani syndrome" with special reference to tension headache." *Nordisk medicin.* 1971; 85(18):577-577.

87. Palva T. Occupational deafness in telephone exchange workers. *Acta oto-laryngologica.* 1957; 47(6):510-519.

88. Milhinch J. *Acoustic Shock Injury - A report on injury following acoustic incidents in call centres.* Melbourne, Australia 2001.

89. Lawton B. Audiometric findings in call centre workers exposed to acoustic shock. *Proceedings Institute of Acoustics.* 2003; 25(4):249-258.

90. Beastall R. Acoustic trauma in a telephone operator. *Occupational Medicine.* 1992; 42(4):215-216.

91. Guyot J-P. Acoustic trauma caused by the telephone. *ORL.* 1988; 50(5):313-318.

92. Parker W, Parker V, Parker G, Parker A. 'Acoustic shock': A new occupational disease? Observations from clinical and medico-legal practice. *International journal of audiology.* 2014; 53(10):764-769.

93. Westcott M, Sanchez TG, Diges I, et al. Tonic tensor tympani syndrome in tinnitus and hyperacusis patients: A multi-clinic prevalence study. *Noise and Health.* 2013; 15(63):117.

94. ETSI. ETSI TR 101 800 Acoustic safety of Terminal Equipment (TE); An investigation on standards and approval documents. Sophia Antipolis, France: European Telecommunications Standards Institute; 2000.

95. ITU. ITU-T Recommendation P.57 Artificial ears. *Objective measuring apparatus*. Geneva, Switzerland: International Telecommunications Union; 2011.

96. SA. *AS/CA S004:2013* Voice performance requirements for Customer Equipment. Sydney, Australia: Standards Australia; 2013.

97. CA. G616:2013 Acoustic safety for telephone equipment. Sydney, Australia: Communications Alliance; 2013.

98. Chalupper J. Aural Exciter and Loudness Maximizer: What's Psychoacoustic about-Psychoacoustic Processors? Paper presented at: Audio Engineering Society Convention 109 2000.

99. Ronan M, Ward N, Sazdov R. The Perception of Hyper-Compression by Untrained Listeners. Paper presented at: Audio Engineering Society Conference: 60th International Conference: DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech) 2016.

100. Croghan NB, Arehart KH, Kates JM. Quality and loudness judgments for music subjected to compression limiting. *The Journal of the Acoustical Society of America.* 2012; 132(2):1177-1188.

101. Taylor RW, Martens WL. Hyper-Compression in Music Production: Listener Preferences on Dynamic Range Reduction. Paper presented at: Audio Engineering Society Convention 136 2014.

102. Werrbach DR. Audio Processing for Broadcasting. Paper presented at: Audio Engineering Society Convention 3r 1988.

103. Orban R, Foti F. What Happens to My Recording When it's Played on the Radio? Paper presented at: Audio Engineering Society Convention 111 2001.

104. Lund T. Control of Loudness in Digital TV. Paper presented at: Proc. of the NAB-2006 Convention 2006.

105. Moore BC, Glasberg BR, Stone MA. 'Why Are Commercials so Loud?' Perception and Modeling of the Loudness of Amplitude-Compressed Speech. *Journal of the Audio Engineering Society.* 2003; 51(12):1123-1132.

106. Rosenheck AJ, Kronman JD. Design of a High Quality Public Address System for Aircraft Use. Paper presented at: Audio Engineering Society Convention 38 1970.

107. Stone MA, Moore BC. Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task. *The Journal of the Acoustical Society of America.* 2004; 116:2311.

108. Stone MA, Moore BC. Quantifying the effects of fast-acting compression on the envelope of speech. *The Journal of the Acoustical Society of America.* 2007; 121(3):1654-1664.

109. Stone MA, Moore BC, Füllgrabe C, Hinton AC. Multichannel fast-acting dynamic range compression hinders performance by young, normal-hearing listeners in a two-talker separation task. *Journal of the Audio Engineering Society.* 2009; 57(7/8):532-546.

110. Dillon H. *Hearing aids*. 2nd ed. Sydney, Australia: Boomerang Press; New York, USA: Thieme; 2012.

111. Souza PE. Effects of compression on speech acoustics, intelligibility, and sound quality. *Trends in amplification.* 2002; 6(4):131-165.

112. Moore BC. The choice of compression speed in hearing aids: Theoretical and practical considerations and the role of individual differences. *Trends in amplification.* 2008; 12(2):103-112.

113. Dillon H, Storey L. The National Acoustic Laboratories' procedure for selecting the saturation sound pressure level of hearing aids: theoretical derivation. *Ear and hearing.* 1998; 19(4):255-266.

114. Stelmachowicz P, Lewis D, Hoover B, Keefe D. Subjective effects of peak clipping and compression limiting in normal and hearing-impaired children and adults. *The Journal of the Acoustical Society of America.* 1999; 105(1):412-422.

115. ANSI. ANSI S3.22-2003 Specifications of hearing aid characteristic. New York, USA: American National Standards Institute; 2003.

116. Moore BC, Glasberg BR. A comparison of four methods of implementing automatic gain control (AGC) in hearing aids. *British journal of audiology.* 1988; 22(2):93-104.

117. Williamson MJ, Cummins KL, Hecox KE, Inventors. Adaptive, programmable signal processing and filtering for hearing aids: US patent 5,027,410 1991.

118. Kates JM. Principles of digital dynamic-range compression. *Trends in amplification.* 2005; 9(2):45-76.

119. Ludvigsen C, Inventor. Circuit arrangement for automatic gain control of hearing aids. US patent 5,687,241 1997.

120. Blamey PJ. Adaptive dynamic range optimization (ADRO): a digital amplification strategy for hearing aids and cochlear implants. *Trends in amplification.* 2005; 9(2):77-98.

121. Blamey PJ, James CJ, McDermott HJ, Martin L, Wildi K, Inventors. Adaptive dynamic range optimization sound processor. US patent 7,366,315 B2 2008.

122. Glover R, Inventor. USApp 20140301588 Hearing dose management 2012.

123. Stade WR, Inventor; Plantronics, assignee. US 5,229,721 Micropower amplifier/transducer driver with signal expansion 1993.

124. Schneider T, Brennan R, Balsiger P, Heubi A. An ultra low-power programmable DSP system for hearing aids and other audio applications. Paper presented at: Proc. ICSPAT 1999.

125. ITU. ITU-T P.58 Head and Torso Simulator for Telephonometry. Geneva, Switzerland: International Telecommunications Union; 2011.

126. Bernstein ER, Brammer AJ, Yu G. Improving speech understanding in communication headsets: Simulation of adaptive subband processing for speech in noise. *International Journal of Industrial Ergonomics.* 2013; 43(6):526-535.

127. Brammer AJ, Yu G, Bernstein ER, Cherniack MG, Peterson DR, Tufts JB. Understanding speech when wearing communication headsets and hearing protectors with subband processing. *The Journal of the Acoustical Society of America.* 2014; 136(2):671-681.

128. Wiggins IM, Seeber BU. Effects of dynamic-range compression on the spatial attributes of sounds in normal-hearing listeners. *Ear and hearing.* 2012; 33(3):399-410.

129. Wiggins IM, Seeber BU. Linking dynamic-range compression across the ears can improve speech intelligibility in spatially separated noise. *The Journal of the Acoustical Society of America.* 2013; 133(2):1004-1016.

130. Plomp R. The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *The Journal of the Acoustical Society of America.* Jun 1988; 83(6):2322-2327.

131. Hornsby BW, Ricketts TA. The effects of compression ratio, signal-to-noise ratio, and level on speech recognition in normal-hearing listeners. *The Journal of the Acoustical Society of America.* 2001; 109(2964).

132. IEC. International Standard IEC 61260:1995, Electroacoustics — Octave-band and fractional-octave-band filters. *Electroacoustics*. Geneva, Switzerland: International Electrotechnical Commission; 1995.

133. Kates JM, Inventor. Hearing aid signal-processing system. US patent 4,852,175 1989.

134. Crochiere R, Rabiner L. Multirate digital signal processing Englewood Cliffs: Prentice-Hall; 1983.

135. Patterson R, Nimmo-Smith I, Holdsworth J, Rice P. An efficient auditory filterbank based on the gammatone function. Paper presented at: a meeting of the IOC Speech Group on Auditory Modelling at RSRE 1987.

136. Patterson R, Moore BCJ. Auditory filters and excitation patterns as representations of frequency resolution. In: Moore BCJ, ed. *Frequency Selectivity in Hearing*. London: Academic; 1986: 123-177.

137. IEC. International Standard IEC 60318, An artificial ear, of the wideband type, for the calibration of earphones used in audiometry. Geneva, Switzerland: International Electrotechnical Commission; 1970.

138. IEC. International Standard IEC 60711, Occluded-ear simulator for the measurement of earphones coupled to the ear by ear inserts. Geneva, Switzerland: International Electrotechnical Commission; 1981.

139. Kuwano S, Namba S, Kato T, Hellbrück J. Memory of the loudness of sounds in relation to overall impression. *Acoustical Science and Technology.* 2003; 24(4):194-196.

140. Pearsons KS, Bennett RL, Fidell S. *Speech levels in various noise environments*: Office of Health and Ecological Effects, Office of Research and Development, US EPA; 1977.

141. Allen JB, Rabiner LR. A unified approach to short-time Fourier analysis and synthesis. *Proceedings of the IEEE.* 1977; 65(11):1558-1564.

142. Schneider T, Brennan R. A multichannel compression strategy for a digital hearing aid. Paper presented at: Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference 1997.

143. Kates JM. *Digital hearing aids*. San Diego, USA: Plural Pub.; 2008.

144. On Semiconductor. Introduction to Audio Processing Using the WOLA Filterbank Coprocessor. *https://www.onsemi.com/pub/Collateral/AND8383-D.P, February.* 2009.

145. Fisher M. Spectral enhancement project: Report 1 - Spectral modification scheme. Sydney, Australia: National Acoustic Laboratories; 1994.

146. MathWorks Inc. MATLAB The language of technical computing. http://www.mathworks.com: MathWorks, Inc; 2016.

147. Kernighan BW, Ritchie DM, Ejeklint P. *The C programming language. Vol 2.* Englewood Cliffs: Prentice-Hall; 1988.

148. TI. TMS320C6727, TMS320C6726, TMS320C6722 Floating-Point Digital Signal Processors. http://www.ti.com/lit/ds/symlink/tms320c6726.pdf, Texas, USA: Texas Instruments; 2007.

149. Moore BC, Glasberg BR, Plack C, Biswas A. The shape of the ear's temporal window. *The Journal of the Acoustical Society of America.* 1988; 83(3):1102-1116.

150. Harris FJ. On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE.* 1978; 66(1):51-83.

151. Proakis JG, Manolakis DG. *Digital signal processing: principles, algorithms, and applications.* Englewood Cliffs, USA: Pentice-Hall; 1996.

152. Schaub A. *Digital hearing aids*. New York, USA: Thieme; 2008.

153. Schaub A, Inventor. Hearing aid. US patent 6,580,798 2003.

154. Thomas LD, Inventor; National Research Development Corporation, assignee. Method and apparatus for automatic signal level adustment. US patent 4,947,133 1990.

155. TI. *Code Composer Studio Development Tools v3. 3 Getting Started Guide. SPRU509H, May.* Texas, USA: Texas Instruments; 2008.

156. TI. Lyrtech professional audio development kit. Texas, USA http://www.ti.com/lit/ml/sprt350/sprt350.pdf: Texas Instruments; 2005.

157. Cox RM, Alexander GC, Taylor IM, Gray GA. The contour test of loudness perception. *Ear and hearing.* 1997; 18(5):388-400.

158. ITU. ITU-R Recommendation BS.1534-3 Method for the subjective assessment of intermediate quality level of audio systems. Geneva, Switzerland: International Telecommunication Union. Radiocommunication Service; 2014.

159. ITU. ITU-R Recommendation BS.1116-1 Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. Geneva, Switzerland: International Telecommunication Union. Radiocommunication Service; 1997.

160. Riley R. *Audio editing with Adobe audition*: PC Publishing; 2008.

161. Fisher RA. *Statistical methods for research workers*: Genesis Publishing Pvt Ltd; 1925.

162. Dunn OJ. Multiple comparisons among means. *Journal of the American Statistical Association.* 1961; 56(293):52-64.

163. Patuzzi R, Milhinch J, Doyle J. Acute Aural Trauma in telephone headset and handset users. Paper presented at: NeuroOtological Society of Australia National Conference, Melbourne, Australia 2000.

164. Dillon H, Fisher M. Acoustic shock - an overview. Paper presented at: XXVI International Congress of Audiology, Melbourne; 2002.

165. Westcott M. Acoustic shock injury (ASI). *Acta Oto-Laryngologica.* 2006; 126(S556):54-58.

166. Blamey PJ, Macfarlane DS, Steele BR. An intrinsically digital amplification scheme for hearing aids. *EURASIP Journal on Applied Signal Processing.* 2005; 2005:3026-3033.

167. AES. AES5-2008 AES recommended practice for professional digital audio — Preferred sampling frequencies for applications employing pulse-code modulation. New York, USA: Audio Engineering Society; 2008.

168. EBU. Specification of the digital audio interface (the AES/EBU interface). Vol Tech 3250-E - Third edition. Geneva, Switzerland: European Broadcast Union; 2004.

169. ITU. ITU Recommendation G. 711 Pulse Code Modulation (PCM) of voice frequencies. Geneva, Switzerland: International Telecommunications Union; 1988.

170. ITU. ITU Recommendation G.722: 7 kHz audio-coding with 64 kbit/s. Geneva, Switzerland: International Telecommunications Union; 2012.

171. Durrant JD, Lovrinic JH. *Bases of hearing science*. Baltimore, USA: Williams & Wilkins; 1984.

172. ITU. ITU-T Recommendation P. 501 Test signals for use in telephonometry. *Series P: Terminals and subjective and objective assessment methods*. Geneva, Switzerland: International Telecommunications Union; 2009.

173. Dillon H. *Hearing aids*. New York, USA: Thieme, Sydney, Australia: Boomerang Press; 2001.

174. Schaub A, Leber R, Inventors. Loudness-controlled processing of acoustic signals. US patent 6,370,255 2002.

175. Gatehouse S, Naylor G, Elberling C. Linear and nonlinear hearing aid fittings–1. Patterns of benefit: Adaptación de auxiliares auditivos lineales y no lineales–1. Patrones de beneficio. *International journal of audiology.* 2006; 45(3):130-152.

176. Hansen M. Effects of multi-channel compression time constants on subjectively perceived sound quality and speech intelligibility. *Ear and hearing.* 2002; 23(4):369-380.

177. Neuman AC, Bakke MH, Mackersie C, Hellman S, Levitt H. The effect of compression ratio and release time on the categorical rating of sound quality. *The Journal of the Acoustical Society of America.* 1998; 103(5):2273-2281.

178. Jenstad LM, Souza PE. Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility. *Journal of Speech, Language and Hearing Research.* 2005; 48(3):651.

179. Rabiner L, Schafer R. *Digital processing of speech signals.* USA: Prentice-Hall; 1978.

180. Harrington J, Cassidy S. *Techniques in speech acoustics.* Vol 8: Springer Science & Business Media; 1999.

181. Fisher M, Chau N, Mulas P, Sharma M. PB Word Lists for Australian English - Normative Performance in Quiet and Noise. *Audiology Australia XX National Conference*. Adelaide 2012.

182. Fairbanks G. The rainbow passage. *Voice and articulation drillbook.* New York, USA: Harper and Row; 1960; 2.

183. ITU. ITU-R Recommendation BS 1284-1, General methods for the subjective assessment of sound quality. Geneva, Switzerland: International Telecommunication Union. Radiocommunication Service; 2003.