UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING AND THE ENVIRONMENT

Institute of Sound and Vibration Research

Evaluation of the Sparse Coding Shrinkage Noise Reduction Algorithm for the Hearing Impaired

by

Jinqiu Sang

Thesis for the degree of Doctor of Philosophy

September 2012

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND THE ENVIRONMENT

Institute of Sound and Vibration Research

Doctor of Philosophy

EVALUATION OF THE SPARSE CODING SHRINKAGE NOISE REDUCTION ALGORITHM FOR THE HEARING IMPAIRED

Jinqiu Sang

Although there are numerous single-channel noise reduction strategies to improve speech perception in a noisy environment, most of them can only improve speech quality but not improve speech intelligibility for normal hearing (NH) or hearing impaired (HI) listeners. Exceptions that can improve speech intelligibility currently are only those that require *a priori* statistics of speech or noise. Most of the noise reduction algorithms in hearing aids are adopted directly from the algorithms for NH listeners without taking into account of the hearing loss factors within HI listeners. HI listeners suffer more in speech intelligibility than NH listeners in the same noisy environment. Further study of monaural noise reduction algorithms for HI listeners is required.

The motivation is to adapt a model-based approach in contrast to the conventional Wiener filtering approach. The model-based algorithm called sparse coding shrinkage (SCS) was proposed to extract key speech information from noisy speech. The SCS algorithm was evaluated by comparison with another state-of-the-art Wiener filtering approach through speech intelligibility and quality tests using 9 NH and 9 HI listeners. The SCS algorithm matched the performance of the Wiener filtering algorithm in speech intelligibility and speech quality. Both algorithms showed some intelligibility improvements for HI listeners but not at all for NH listeners. The algorithms improved speech quality for both HI and NH listeners.

Additionally, a physiologically-inspired hearing loss simulation (HLS) model was developed to characterize hearing loss factors and simulate hearing loss consequences. A methodology was proposed to evaluate signal processing strategies for HI listeners with the proposed HLS model and NH subjects. The corresponding experiment was performed by asking NH subjects to listen to unprocessed/enhanced speech with the HLS model. Some of the effects of the algorithms seen in HI listeners are reproduced, at least qualitatively, by using the HLS model with NH listeners.

Conclusions: The model-based algorithm SCS is promising for improving performance in stationary noise although no clear difference was seen in the performance of SCS and a competitive Wiener filtering algorithm. Fluctuating noise is more difficult to reduce compared to stationary noise. Noise reduction algorithms may perform better at higher input signal-to-noise ratios (SNRs) where HI listeners can get benefit but where NH listeners already

reach ceiling performance. The proposed HLS model can save time and cost when evaluating noise reduction algorithms for HI listeners.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, JINQIU SANG, declare that the thesis entitled EVALUATION OF THE SPARSE CODING SHRINKAGE NOISE REDUCTION ALGORITHM FOR THE HEARING IMPAIRED and the work presented in this thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as conference and journal papers and listed in ***Appendix A.***

Signed: ………………………………………………………………………………

Date:………………………………………………………………………………….

Acknowledgements

This research was supported by the European Commission within the ITN AUDIS (grant agreement number PITN-GA-2008-214699). My thesis would have been impossible without the support from ISVR and the generous support from the AUDIS project.

Firstly, foremost, I am especially grateful to my supervisors, Dr. Stefan Bleeck and Prof. Mark Lutman. Dr. Stefan Bleeck always encouraged me in every stage of accomplishment of my PhD research and guided me with creative ideas. His unique insights and relaxed way of supervision allowed me to shape my research the way I wanted. Prof. Mark Lutman is very experienced in hearing research and I was always amazed by his smart ideas and abundant knowledge. I was so lucky to have him keeping an insightful eye on my research and offering me the assurance that I was going in a sensible way. I would like to take this opportunity to thank Dr. Hongmei Hu who always provided constructive suggestions and help during my research. The study of developing and evaluating the hearing loss simulation model was done with her. I also would like to thank Dr. Chengshi Zheng who always provided advice on noise reduction algorithms, Prof. Timo Gerkmann who provided the competitive Wiener Filtering code for me to compare in my thesis, Dr. Guoping Li whose brilliant work in sparse coding for Cochlear Implant users inspired my thesis topic, Dr. Shouyan Wang, who encouraged me during many stages of my research and also gave me insightful advice for my future. Thanks are due to Prof. Paul White who always gave constructive criticisms during my reviews, Dr. Ben Lineton who coordinated the thesis committee, Prof. Mark Plumbley who served as my examiner, Zoe Bevis, Hannah Semeraro and Laurence Wilmshirst for proofreading my thesis.

I would like to take the opportunity to thank all AUDIS members, Prof. Rainer Martin, Prof. Arne Leijon, Prof. Jan Wouters, Dr. Bas van Dijk, Dr. Henning Puder, Masoumeh Azarpour, Rajyalakshmi Dokku, Jalal Taghia, Jalil Taghia, Nasser Mohammadiha, Qazi Obaid, Cees H. Taal, Vaclav Bouse, Falk-Martin Hoffmann, Agamemnon Krasoulis. I believe our friendship will last forever and the experiences in AUDIS will be an invaluable treasure for us.

I would like to thank my friends from ISVR, Yan Cang, Guanjian Ni, Rami Saba, Liping Zhang, Ying Guo for making the offices a fun environment.

Finally, my special thanks go to my grandmother, my parents and my fiancé for their unconditional love, understanding and encouragement during the three years I have been abroad. Their love makes me more confident and excellent. I know they have missed me a lot and the same is also very true for me. This thesis is dedicated to my family.

List of abbreviations

AGC	Automatic gain control
BKB	Bamford-Kowal-Bench sentence (British sentence corpus)
CI	Cochlear implant
CS-WF	A Wiener filtering approach with cepstral smoothing
DFT	Discrete Fourier Transform
fwsegSNR	Frequency-weighted segmental singal-to-noise-ratio
HA	Hearing aid
HASQI	Hearing aid speech quality index
HI	Hearing impaired
HLS	Hearing loss simulation
ICA	Independent component analysis
IPCR	Interpolated paired comparison rating
IS	Itakura-Saito
KLT	Karhunen Loeve Transform
K-SVD	K-means Singular Value Decomposition
LLR	Log-likelihood ratio
LSD	Least-significant difference
MAP	Maximum a posterior
MMSE	Minimum mean square error
MOS	Mean opinion score
MSC	Magnitude-squared coherence
NAL	National Acoustics Laboratory procedure
NCM	Normalized covariance metric
NH	Normal hearing
NMF	Non-negative matrix factorization

NPSD	Noise power spectral density
PCA	Principle component analysis
PESQ	Perceptual evaluation of sound quality
PSE	Point of subjective equality
PSM	Perceptual similarity measure
PTA	Pure tone audiometry
Roex	Rounded exponential
SCS	Sparse coding shrinkage
segSNR	Segmental Signal-to-noise-ratio
SII	Speech intelligibility index
SNR	Signal-to-noise-ratio
SPP	Speech presence probability
SRT	Speech reception threshold
SSN	Speech shaped noise
STFT	Short-time Fourier Transform
STMI	Spectral-temporal modulation index
STOI	Short-time objective intelligibility measure
STI	Speech transmission index
SVD	Singular Value Decomposition
VQ	Vector Quantization

Chapter 1 Introduction

1.1 Contribution to knowledge

Communication in the presence of background noise is known to be a major problem for hearing impaired listeners. Hearing impaired (HI) listeners usually need a signal-to-noise ratio (SNR) at least 3 or 6 dB higher than normal hearing (NH) listeners to reach the same intelligibility level. Therefore noise reduction algorithms are critical in order to alleviate the difficulty of speech communication in daily life for hearing aid users. In spite of large advances of noise reduction algorithms, most of the algorithms only improve speech quality rather than intelligibility for normal hearing listeners. The intelligibility effects of noise reduction algorithms for hearing impaired listeners need to be further evaluated and confirmed. Although speech intelligibility can be improved using noise reduction algorithms with a large microphone array or single-channel noise reduction algorithms with *a priori* knowledge of the speech and/or noise, they are not practical due to the cosmetic considerations of a small size of a hearing aid or the variation in environmental noise. Therefore the motivation of this thesis is how to improve and evaluate single-channel noise reduction algorithms for hearing aid users.

Most of the single-channel noise reduction algorithms have been developed and evaluated for telecommunication systems, (e.g. cell phones, automatic speech recognition devices, etc.) for NH listeners; little research has been done to develop or evaluate noise reduction algorithms specifically for hearing aid users. Usually a noise reduction algorithm, originally developed for NH listeners, is adopted directly for HI listeners. Due to the hearing loss factors that distort the speech perception in HI listeners, a noise reduction algorithm that is optimal for NH listeners may not necessarily be optimal for HI listeners. Previous research has shown that HI listeners are more sensitive to noise (Arehart et al., 2003) but are less sensitive to speech distortion (Schijndel et al., 2001). This inspired investigation as to whether a noise reduction algorithm with more pronounced noise reduction effects can help HI listeners more at the expense of speech distortion that can be tolerated by HI listeners. In other words, algorithms that can extract key information of speech might help HI listeners more. The concept of sparse stimuli for hearing impaired listeners has been proposed in cochlear implant processing (Li, 2008, Li and Lutman, 2008, Hu et al., 2011a, Sang et al., 2011a). The sparse strategy has been found to bring significant intelligibility improvement for CI (cochlear implant) users (Li, 2008) who have profound hearing losses. This progress inspires to develop and evaluate sparse stimuli for hearing aid users who have mild to severe hearing losses (Sang et al., 2012). To the author's knowledge, this research is the first to evaluate sparse stimuli effects for hearing aid users.

Most single-channel noise reduction algorithms in current hearing aids utilise Wiener filtering or spectral subtraction methods, due to their robust performance and simple calculation. Recent advance in computation could bring other noise reduction methods with more complicated procedures into consideration for use in hearing aids. We propose a noise reduction algorithm that is a statistically based method in the principal component space through principal component analysis (PCA). We assumed that a) speech components in PCA space are sparsely distributed, and b) noise components have a Gaussian distribution and we applied the principle of sparse coding shrinkage (SCS) to reduce noise more efficiently. SCS is one of the few algorithms that take account of impaired auditory perceptual characteristics to better serve the group of HI listeners. Subjective tests with both NH and HI listeners were performed to assess the difference in perception of sparse stimuli between NH and HI listeners. Research questions are as follows:

Can SCS bring different intelligibility effects between NH listeners and HI listeners?

Can SCS make different quality impressions between NH listeners and HI listeners?

What are the factors that determine the benefits of noise reduction algorithms to NH and HI listeners?

To compare with SCS, unprocessed noisy speech and a noise reduction algorithm called CS-WF (cepstral smoothing based Wiener filtering) were also evaluated as a baseline performance and a competitive algorithm respectively. CS-WF was chosen, because Wiener filters are used frequently in today's hearing aids and it is a competitive state of the art algorithm (Breithaupt et al., 2008, Gerkmann and Martin, 2009, Gerkmann and Hendriks, 2012). Since the *a priori* SNR of the Wiener filtering approach is estimated by the cepstral smoothing method, we refer this approach as CS-WF herein. This algorithm is regarded as an optimal noise reduction algorithm for NH listeners and its effects are uncertain in HI listeners. By comparing SCS with CS-WF, it can be shown which algorithm may lead to greater benefits for NH or HI listeners in terms of speech intelligibility and speech quality.

The subjective tests included speech recognition tests and speech quality tests. The speech recognition tests were performed through an adaptive procedure to measure the speech reception threshold (SRT), which is the SNR corresponding to 79.4% correct recognition (Dahlquist et al., 2005). The quality tests were performed through the interpolated paired comparison rating (IPCR) to measure the SNR gain in terms of quality (Dahlquist et al., 2005). The results of quality and intelligibility were put together and compared on the same scale to give an illustration of the overall effects of noise reduction algorithms for NH or HI listeners.

To compare noisy speech, CS-WF and SCS, objective measures were also performed although they are not as meaningful as subjective tests. Objective measures are quick and cheap, and therefore popular during the algorithm development stage and widely used in the signal processing community. Most objective measures are developed and validated with NH listeners. Even some objective measures, which are developed for HI listeners with additive parameters of hearing thresholds, are not necessarily representative for speech distortion brought by noise reduction algorithms. Five objective measures, including frequency weighted segmental SNR (fwsegSNR), perceptual evaluation of sound quality (PESQ), the hearing aid speech quality index (HASQI), the normalized covariance metric (NCM) and the short-time objective intelligibility measure (STOI) were performed in this thesis to compare with subjective results. FwsegSNR and PESQ are objective speech quality measures that reflect physical noise reduction effects and speech distortion effects respectively for NH listeners. HASQI is an objective speech quality measure for HI listeners with input parameters of hearing threshold levels at six frequencies (250, 500, 1000, 2000, 4000, 6000 Hz). NCM and STOI are objective speech intelligibility measures for NH listeners. To the author's knowledge, the research is the first to validate HASQI in evaluating noise reduction algorithms in speech quality for HI listeners with subjective tests.

Current objective intelligibility measures for HI listeners are rare and often unreliable in predicting intelligibility effects of noise reduction algorithms for HI listeners. We propose an evaluation methodology to evaluate noise reduction effects on HI listeners through a hearing loss simulation model with NH subjects listening. The assumption is that consequences of hearing impaired listening can be an additive combination of hearing impaired distortion with normal hearing listening. If hearing- impaired distortion could be simulated in an appropriate and realistic way, the effects of listening to speech with HI listeners can be approximated by asking NH listeners to listen to the same speech with the hearing loss simulation (HLS). In our group, we proposed a physiologically-inspired hearing loss simulation model that has been validated with noisy speech in respect of speech intelligibility (Hu et al., 2011b). In this thesis, we evaluate noise reduction algorithms with this HLS model. If the HLS model is realistic, the effects of noise reduction algorithms on the HLS model might approach the effects on HI subjects. Therefore, another subjective speech recognition test was performed by asking NH listeners to listen to speech processed with the HLS model. Comparison between the experiment with the HLS model and the HI subjects can show whether an auditory filter based HLS model can predict noise reduction effects for HI listeners with NH listeners. There are several advantages to this experiment if it can appropriately reflect effects of HI listening. Firstly, it is much easier to recruit NH listeners than HI listeners. Secondly, we can introduce one specific hearing loss level during hearing loss simulation experiments; thus, the variability of responses in HI listeners due to different hearing losses is minimised. The proposed evaluation methodology can be regarded as a combination of a subjective test with an objective model. Current purely objective measures for HI listeners have only been validated with clean speech or noisy speech rather than with noise reduction algorithms. A future perspective is to set up a purely objective speech intelligibility measure that includes the auditory filter based HLS model to assess noise reduction effects on HI listeners. The research is thought to be the first to assess the validity of the hearing loss simulation model to evaluate effects of noise reduction algorithms in speech intelligibility for HI listeners.

1.2 Introduction of hearing loss

It is estimated that, in UK, the number of people suffering from hearing loss was 10 million in 2011 and will increase to 14.5 million by 2031. The main headline statistics about hearing loss in UK is cited as follows from the website of Action on Hearing Loss[1] in bullet points.

- *There are more than 10 million people in the UK with some form of hearing loss, or one in six of the population.*
- *From the total 3.7 million are of working age (16 – 64) and 6.3 million are of retirement age (65+).*
- *By 2031, it is estimated that there will be 14.5 million people with hearing loss in the UK.*
- *More than 800,000 people in the UK are severely or profoundly deaf.*
- *There are more than 45,000 deaf children in the UK, plus many more who experience temporary hearing loss.*
- *More than 70% of over 70 year-olds and 40% of over 50 year-olds have some form of hearing loss.*
- *There are approximately 356,000 people with combined visual and hearing impairment in the UK.*
- *About two million people in the UK have hearing aids, but only 1.4 million use them regularly.*
- *At least four million people who don't have hearing aids would benefit from using them.*
- *On average it takes ten years for people to address their hearing loss.*

[1] http://www.actiononhearingloss.org.uk/your-hearing/about-deafness-and-hearing-loss/statistics.aspx

Hearing impairment is likely to restrict social interaction and educational and career opportunities and thus significantly degrade quality of life. Hearing impairment can impose a heavy social and economic burden on individuals, families, communities and countries. Estimates published in 2006 suggested that £13bn is lost to the UK economy every year through unemployment linked to hearing loss (Shield, 2006).

Despite the severe status of hearing loss, hearing research is still significantly underfunded, which affects the development of treatments and cures. Estimates also suggested that, in 2010, the UK spent £1.34 on research into hearing loss for every person affected, which compares to £14.21 for sight loss, £21.31 for diabetes, and £49.71 for cardiovascular research (RNID, 2010). This calls for larger investment into hearing research in UK.

Hearing loss can be defined in different levels usually diagnosed from individual's audiograms. Hearing loss levels can be categorized as mild, moderate, severe, and profound, depending upon how well a person can hear the intensities or frequencies most strongly associated with speech. Impairments in hearing can occur in only one ear or in both ears. Figure 1.1 indicates the definition of different hearing threshold levels (BSA, 2011) and an example of mild-to-moderate high frequency hearing losses in both ears of an individual. Mild hearing loss indicates hearing threshold levels between 20 dB HL and 40 dB HL; Moderate hearing loss indicates hearing threshold levels between 41 dB HL and 70 dB HL; Severe hearing loss indicates hearing threshold levels between 71 dB HL and 94 dB HL; Profound hearing loss indicates hearing threshold levels more than 95 dB HL.

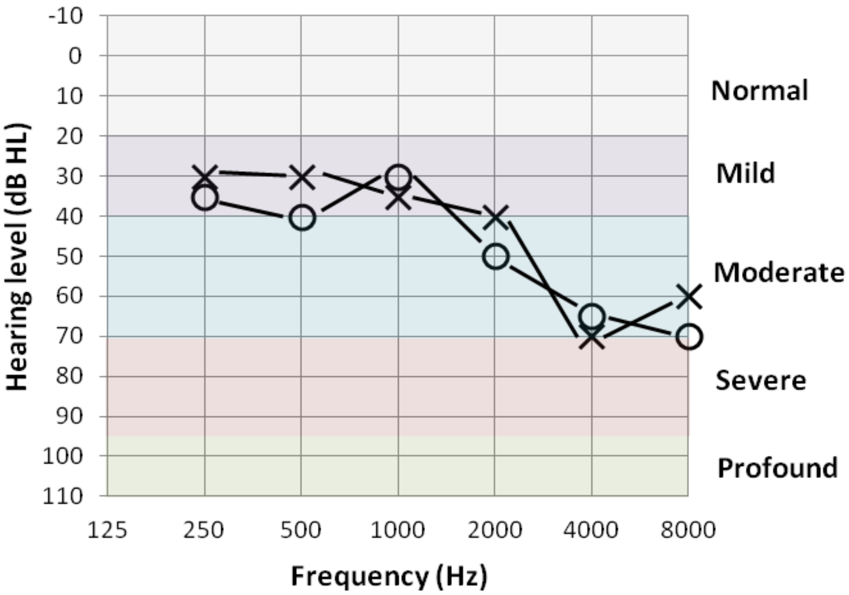


Figure 1.1: An example of typical mild-to-moderate high frequency hearing loss is shown in dB HL, where "O" indicates an audiogram of the right ear; "✕ "indicates an audiogram of the left ear. Different hearing loss levels are shown with different colours.

Hearing loss can also be categorised in different types, depending on which part of the hearing pathway is affected[2]. It is important to localize where the problem lies in the hearing pathway to determine an appropriate treatment and compensation strategies.

The four types of hearing loss are:

- Conductive hearing loss
- Sensorineural hearing loss
- Central hearing loss
- Mixed hearing loss

Purely conductive losses are those resulting from dysfunction of the ear canal or middle-ear structures, such that less acoustic energy reaches the auditory receptors in the normal cochlea. Today, most conductive losses are successfully treated through medical intervention and do not necessitate hearing aids to understand speech (Watson, 1991).

[2] http://ehealthmd.com/content/different-types-hearing-loss

Sensorineural losses involve reduced sensitivity or resolving power of the neural receptor mechanisms themselves, which is generally not amenable to medical intervention. Hearing aids and cochlear implants are designed primarily for individuals with sensorineural hearing loss (Watson, 1991). For people with mild to severe hearing losses, hearing aids are usually prescribed. For people with profound hearing losses, cochlear implants are more effective to help speech perception. In central hearing loss, the problem lies in the central nervous system, at some point within the brain. Some people with central hearing loss can hear perfectly well but have trouble interpreting or understanding speech. There is no effective treatment for central auditory processing disorders, other than educating the person, family, and friends, to manage the environment. Mixed hearing loss occurs when both conductive and sensorineural hearing losses are present in the same ear.

People with sensorineural hearing loss often have difficulty in understanding speech, especially when the speech is corrupted with noise. Previous studies have attempted to understand how sensorineural hearing loss affects perceptual difficulties through hearing loss simulations (Moore and Glasberg, 1993, Baer and Moore, 1993, Baer and Moore, 1994, Nejime and Moore, 1997, Hu et al., 2011b). They have pointed out that speech perception difficulties for people with sensorineural hearing loss is due to reduced audibility of the speech signal (Humes et al., 1987, Zurek and Delhorne, 1987) and to abnormalities in the perceptual analysis of the signal, even when it is well above the absolute threshold (Plomp, 1986, Dreschler and Plomp, 1980, Dreschler and Plomp, 1985, Glasberg and Moore, 1989). Previous hearing loss simulations have studied perceptual effects of different hearing loss factors, such as hearing threshold elevation, loudness recruitment, reduced frequency selectivity and reduced temporal resolution (Nejime and Moore, 1997, Hu et al., 2011b). Moore and Glasberg (1993) asked normal hearing listeners to listen to the speech with or without noise processed by a hearing loss simulation model. This way, they could understand how much the intelligibility of speech is affected by different hearing loss factors, and whether a compensation strategy could be devised for a particular hearing loss factor.

In this research, we focus on the group of hearing aid users who usually have mild to severe sensorineural hearing loss. In the next section, we introduce the basic configuration of a hearing aid and what we were going to study and focus on.

1.3 What is a hearing aid

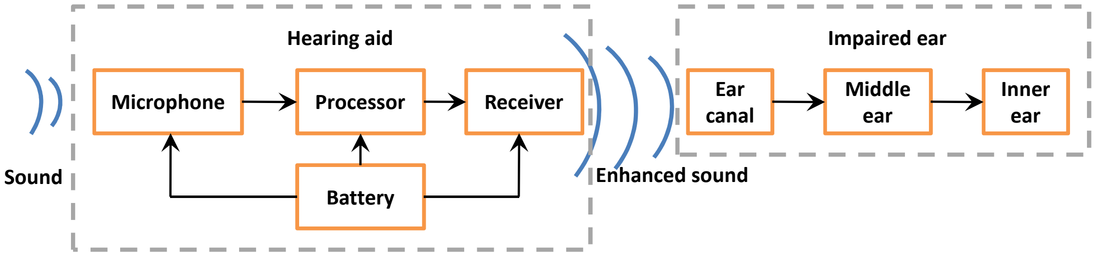


Figure 1.2: The basic configuration of a hearing aid.

Hearing aids use electronic amplification to enhance acoustic signals for hearing impaired people with mild, moderate to severe hearing losses. Hearing aids comprise one or more microphones, an electronic filter (processor), an earphone (also called a receiver), and a battery that serves as the power source. Figure 1.2 shows this configuration. Hearing aids have limited power in speech processing due to the limited size of a hearing aid and its processor chip. A complex signal processing strategy is therefore not practical because it would consume too much power. Hearing aids are usually configured with no more than two microphones each side due to the limited size, which inhibits the use of multi-channel hearing aids from using multi-channel signal processing with a large microphone array, which could improve speech enhancement performance more easily (Kates and Weiss, 1996, Levitt, 2001, Schum, 2003).

Development of strategies in hearing aids includes two aspects. On one hand, the strategies are inspired to compensate for impaired auditory factors to reach the level of normal hearing. For example, dynamic compression could compensate for hearing threshold elevation and loudness recruitment (Moore, 2007). But there are no optimal solutions to compensate for reductions in frequency selectivity and temporal resolution (Moore, 2007). On the other hand, signal processing strategies are developed to reduce interference, e.g. ambient noises, reverberation, feedback interference, etc., which could otherwise deteriorate speech perception for normal hearing listeners and generally render speech incomprehensible for hearing impaired listeners.

All these signal processing strategies may be combined in the hearing aids to better serve the hearing aid users. As we develop and evaluate noise reduction strategies for hearing impaired listeners, we first introduce existing strategies to compensate for impaired auditory factors attempting to reach the level of normal hearing in Section 1.4. We then introduce current single-channel noise reduction strategies and its issues for hearing aids application in Section 1.5. We later introduce how to evaluate noise reduction strategies for hearing aid users in Section 1.6. The review in Section 1.4 and Section 1.5 inspired to develop a noise reduction strategy that could not only reduce additive noise, but also indirectly compensate for impaired auditory factors such as reduced frequency resolution and reduced temporal resolution. Section 1.6 introduces appropriate evaluation methods for HI listeners and also states some evaluation methods that are only suitable for NH listeners.

1.4 Introduction of hearing-impairment compensation strategies in hearing aids

Most sensorineural hearing impairments occur as a result of damage to the cochlea. A simple solution is to restore normal firing pattern of auditory nerve fibres. In principle, this method may involve two processing stages in cascade. The first stage is to simulate the processing of the normal cochlea and the second stage is to perform the inverse of the processing of the damaged cochlea. However, combination of the two processing stages is difficult to realise in a perfect way as both the normal cochlear and the damaged cochlea are complicated nonlinear systems. An attempt can be made to realize an inverse-cochlear model in an approximate way, which is a promising perspective.

Current strategies in hearing aids are developed to compensate for reduced psychoacoustic abilities underlying hearing impairment, e.g. hearing threshold elevation, loudness recruitment, reduced frequency selectivity and reduced temporal resolution. They are developed by intentionally introducing distortions into the speech signal in order to enhance available cues. This section introduces some compensation strategies in hearing aids that take account of impaired auditory factors and strengthen important speech cues.

1.4.1.1 *Linear amplification*

The primary goal of linear amplification in a hearing aid is to compensate for the loss of audibility across frequencies. The linear gain in a hearing aid is independent of sound level. A well-established amplification procedure is the National Acoustic Laboratory (NAL) procedure (Dillon, 2001) which compensates for threshold elevation. Specifically, a gain prescription is computed from an individual's audiogram to produce amplification with appropriate frequency dependent shaping. Because of the limited dynamic range associated with loudness recruitment, most hearing aids include some processing, such as peak clipping or compression limiting, to limit the maximum output levels.

1.4.1.2 *Automatic gain control (AGC)*

Automatic gain control (AGC) strategy has a nonlinear input-output function that operates by reducing overall gain when high-intensity sounds occur and amplifying overall gain when low-intensity sounds occur. The primary purpose of automatic control is to map the input signal into the limited dynamic range of the ear with a damaged cochlea. This method could compensate for loudness recruitment and threshold elevation. Although there have been quite a few techniques of AGC (Moore and Glasberg, 1988, Lunner et al., 1998) based on different rationales (Moore, 1996), there is no clear consensus on the best method.

1.4.1.3 *Base Increases at Low Levels (BILL)*

BILL is applied to adaptive high-pass filtering (HPF) hearing aids. The cut-off frequency of high-pass filtering increases as the level of the input signal increases (Killion et al., 1990). This strategy is applied to mildly or moderately sloping hearing losses with greater upward spread of masking which are more dependent on high-frequency gain for speech intelligibility.

1.4.1.4 *Treble Increases at Low Levels (TILL)*

TILL provides different amplification at different input levels. It acts as a high-pass filter at low input levels and as a flat, transparent filter at high input levels to overcome insertion loss (Tobin, 1997). The insertion loss here is the difference in sound pressure level at the eardrum before inserting a hearing aid and after inserting a hearing aid. The aim of this algorithm is to amplify weak high-frequency consonant energy at low input levels and suppress annoying high input levels. A comparative study between benefits of BILL and TILL is not available.

1.4.1.5 *Consonant amplification*

Previous studies (Montogomery and Edge, 1988, Guelke, 1987, Gordon-Salant, 1987) have demonstrated that amplifying consonants relative to vowels can improve the intelligibility of speech for hearing-impaired subjects.

1.4.1.6 *Time-scale modification*

Researchers have also investigated the possibility of changing the duration of speech segments. It can be understood that speech intelligibility could be increased if the speech is spoken in a slower speed. Revoile et al. (Revoile et al., 1986) concentrated specifically on the contrast between voiced and unvoiced final fricatives. They enhanced the contrast not by altering the consonant duration but by altering the duration of the vowel preceding it. Montgomery and Edge (Montogomery and Edge, 1988) investigated continuant consonants and voiceless stops. Their algorithm lengthened these consonants and shortened the vowel before the voiceless stops. Currently, time-scale modification needs off-line processing which cannot be processed real time.

1.4.1.7 *Spectral sharpening*

Reduced frequency resolution is one of the most important suprathreshold factors that affect speech intelligibility for hearing-impaired people in noise environments. Some attempts have been made to sharpen spectral peaks or formants of the speech signal (Baer and Moore, 1993, Alcantara et al., 1994, Dillon, 2001). Bustamante and Braida (1987) developed a scheme that can compress dynamic range and can also sharpen the spectral peaks of the speech signal through principal component analysis (Jolliffe, 1986). However, no intelligibility improvement has been shown in any of these efforts.

1.4.1.8 *Frequency lowering*

Many cues for recognition of consonants are at higher frequencies which are difficult to catch for people with severe high-frequency hearing impairment. Many researchers (Munoz et al., 1999) have tried to make these cues accessible for patients with relatively good low-frequency hearing by mapping high-frequency information to lower frequencies. However, no significant improvements in intelligibility results from frequency lowering have been demonstrated.

1.4.1.9 *Discussion*

Larson *et al.* (2000) demonstrated that the hearing aids provided significant benefits to hearing impaired listeners when compared to unaided listening conditions. However, this same study cited statistics from recent surveys indicating that only 65% of individuals with hearing aids report that they are satisfied with their hearing instruments (Larson et al., 2000). There is still a long way to go to compensate for hearing impairment and improve speech perception in hearing aids.

Among the existing compensation strategies in hearing aids, the most widely recognized strategy is AGC which compensates for threshold elevation and loudness recruitment. However, there are currently no appropriate solutions to suprathreshold hearing impairment, e.g. reduced frequency selectivity and reduced temporal resolution. Moreover, even if these strategies could perform well in clean speech, their performance might be deteriorated when listening to speech in noise. This is also one of the most frequent complaints of hearing aid users that they could not hear well in noisy environments. Therefore noise reduction strategies are critical to improve speech perception for hearing aid users.

1.5 Introduction of single-channel noise reduction strategies

For people with mild to severe hearing losses, current advanced hearing aids can help improve speech perception in quiet environments. However, one important reason why hearing-aid users do not like to use hearing aids is that the current hearing aids don't work well in background noise (Dillon, 2001, Alcantara et al., 2003). Hearing-impaired (HI) people typically require a speech-to-noise ratio that is 3-6 dB higher than normal-hearing people to achieve the same degree of speech intelligibility (Plomp, 1994, Alcantara et al., 2003). Therefore, noise reduction strategies in hearing aids are one critical factor to help HA users improve quality of life.

Although microphone arrays have been proved to be an option to improve speech intelligibility (Kates and Weiss, 1996, Levitt, 2001, Schum, 2003), their performance is only significant with a large microphone array. Additionally, the microphone array is usually effective when the target speech and interfering sounds are coming from different directions. However, due to the small size of a hearing aid, usually only one or two microphones are placed in a hearing aid. Accordingly, a large microphone array is not practical to be placed in a hearing aid. Some may also argue the importance of beam-forming with two microphones in hearing aids, rather than single-channel noise reduction strategies. Currently, most hearing aids are equipped with combination of single-channel noise reduction algorithms and beam-forming strategies (Widrow and Luo, 2003), which mutually determine the final noise reduction performance of hearing aids. There are still limitations in the current development stage of single-channel noise reduction algorithms, which are worthy of further study. Furthermore, there are several situations that require the use of single-channel strategies, e.g. telephone speech, in-the-ear hearing aids that are place deep in the ear canal, etc.

Previous research has shown that (Levitt, 1993, Levitt et al., 1993, Weiss and Neuman, 1993, Dahlquist et al., 2005) a greater signal-to-noise ratio (SNR) in hearing aids does not guarantee benefits to a hearing-impaired listener. Most noise reduction strategies could not improve speech intelligibility by improving SNR, because the processing removes part of the signal or distorts the speech in a way that reduces intelligibility.

Evaluations of single-channel noise reduction algorithms (Hu and Loizou, 2007) with NH listeners have shown no speech intelligibility improvement, except for the algorithms with the *priori* knowledge of speech and/or background noise (Kim and Loizou, 2010). Evaluations of noise reduction algorithms with hearing impaired listeners have shown various positive and negative effects (Arehart et al., 2003, Dahlquist et al., 2005, Elberling et al., 1993, Levitt et al., 1993, Levitt, 2001, Jamieson et al., 1995). Almost all of the algorithms could show improvement in speech quality, whereas only some of these algorithms could improve speech intelligibility. Besides differences in the noise reduction algorithms themselves, the varied intelligibility effects may be due to the noise type, the test type, the subject's hearing loss level, input SNR, and the subject's hearing aid experience. A spectral subtraction algorithm with an auditory masked threshold (Arehart et al., 2003) has shown intelligibility improvement for HI listeners in the background of voice communication channel noise and automobile highway noise with nonsensical syllable tests. Another nonlinear spectral subtraction algorithm (Lockwood and Boudy, 1992) did not show intelligibility improvement for HI listeners in background of speech shaped noise, babble noise by sentence tests (Dahlquist et al., 2005). Speech shaped noise and babble noise are more difficult to remove as they show the same long term average spectrum as the speech being used. However, the nonlinear spectral subtraction algorithm (Lockwood and Boudy, 1992) that did not show speech intelligibility improvement for hearing aid users showed intelligibility improvement for cochlear implant users, who are profound HI listeners (Verschuur et al., 2006), for the same speech shaped noise and for the same sentence tests. The reason was that the hearing aid users with mild-to-moderate hearing impairments are more sensitive to speech distortion brought by noise reduction algorithms compared to CI users with profound hearing impairments. The more severe the hearing loss is, the more tolerant the listener was to speech distortion (Schijndel et al., 2001). Thus, more benefits can be acquired from noise reduction algorithms for people with more severe hearing losses. It is therefore worthwhile to evaluate some updated noise reduction algorithms for HI listeners.

This section introduces some classical, as well as some state-of-the-art, single-channel algorithms.

1.5.1 **Spectral subtraction algorithm**

Spectral subtraction algorithms are based on a simple principle. If noise is additive, the clean signal spectrum can be estimated by subtracting an estimate of the noise spectrum from the noisy speech spectrum. The noise spectrum can be estimated, and updated. The assumption here is that noise is stationary or quasi-stationary. The phase of the clean signal is estimated to be the same as the phase of the noisy signal. The enhanced signal is obtained by computing the inverse discrete Fourier transform of the estimated signal spectrum with the estimated phase of the clean signal.

Basic spectral subtraction algorithms were introduced by Boll (Boll, 1979) while further modification methods of spectral subtraction has been done by many researchers (Gustafsson et al., 2001, Zenton et al., 1998, Kang and Fransen, 1989). One problem of spectral subtraction is musical noise that is introduced by subtraction of the noise spectrum from the noisy spectrum while the noise spectrum is estimated. Some modifications of spectral subtraction can suppress the musical noise distortion. The most common method is to subtract the estimated noise spectrum multiplied by an over-subtraction factor from the noisy spectrum, and a spectral floor is set to keep the speech spectrum above a minimum value.

1.5.2 **Wiener filtering**

The Wiener filtering approach (Wiener, 1949) is used to reduce the noise present in a signal. The principle of the Wiener filtering is based on minimum-mean-square-error (MMSE) between the estimated and desired clean signal. Using a Gaussian distribution assumption of a stationary clean signal and noise, the optimal Wiener filter in the spectral domain can be derived as the cross power spectrum between the clean signal and the noisy signal divided by auto power spectrum of the noisy signal. Specific details can refer to (Oppenheim and

Verghese, 2010).

Wiener filtering can be implemented either iteratively or noniteratively in mathematical derivation. Autoregressive (AR) speech production models (Jae and Oppenheim, 1978) are examples of iterative models. Performance of this filtering approach is significantly improved with imposed constraints (Hansen and Clements, 1991, Pellom and Hansen, 1998). Several noniterative Wiener filtering algorithms were proposed to get good estimates of the *a priori* SNR with the use of low-variance spectral estimators (multitaper based) (Yi and Loizou, 2004).

Another recently developed Wiener filtering algorithm estimates the *a priori* SNR by temporal cepstrum smoothing with bias compensation (Breithaupt et al., 2008, Gerkmann and Martin, 2009), which can reduce the level of musical noise and suppress the non-stationary noise effectively. Since the *a priori* SNR of the WF approach is estimated by the cepstral smoothing method, we refer this approach as CS-WF herein. CS-WF is selected as a comparison algorithm in our evaluation study.

Although the Wiener filters are considered to be optimal in the mathematically mean-square sense, they are not necessarily the best estimators of the clean signal spectrum in a perceptual sense. However, Wiener filtering algorithms are recognized to produce robust noise reduction performance and can be chosen as competitive noise reduction algorithms.

1.5.3 **Subspace algorithms**

Subspace algorithms can be traced back to early work in principal component analysis. These ideas were first brought into engineering sciences by Pisarenko (Pisarenko, 1973) and later by Schmidt (Schmidt, 1986).

Subspace algorithms are based on the principle that the vector space of the noisy signal can be decomposed into a subspace that is occupied primarily by the clean signal and a subspace occupied primarily by the noise signal. Therefore, the clean speech could be estimated by discarding the component of the noisy vector residing in the noise subspace. The decomposition of the vector space of the noisy signal into "signal" and "noise" can be done using Singular Value Decomposition (SVD) or Karhunen Loeve Transform (KLT) algorithms.

In practice the signal components are further processed after decomposition. Some processing criteria are based on perceptually motivated measures, while others are based on the mathematically tractable square error.

The majority of the subspace algorithms were originally formulated under the assumption that the additive noise is white. Research was also conducted to handle coloured noise with subspace algorithms (Hu and Loizou, 2003).

1.5.4 **Statistical-model-based methods**

Statistical-model-based methods are nonlinear estimators of the magnitude rather than the complex spectrum of the signal. These nonlinear models estimate the probability density function of the noise and the speech coefficients with some transform such as the Discrete Fourier Transform (DFT), principal component analysis (PCA) or independent component analysis (ICA). These methods can be implemented with different optimization criteria, for example a maximum-likelihood estimator (Yoshioka, 2008), a Minimum Mean Square Error (MMSE) magnitude estimator, or a log-MMSE estimator (Ephraim and Malah, 1983, Ephraim and Malah, 1985). Perceptually motivated distortion measures are taken into account in some of these estimators. These estimators are often combined with soft-decision gain modifications that take the probability of speech presence into account. These methods have shown better performance in reducing musical noise.

1.5.4.1 *Sparse coding shrinkage principle in ICA space*

Some attempts have been made to develop speech enhancement algorithms based on the sparse coding shrinkage principle through ICA (Potamitis et al., 2001b, Zou et al., 2008, Sang et al., 2011a, Potamitis et al., 2001a). The sparse coding shrinkage principle assumes a super Gaussian distribution of signal and a Gaussian distribution of noise and suggests several shrinkage functions to suppress noise from the observed noisy signal. Super Gaussian distribution indicates the distribution is in a peaked shape rather than a bell shape of Gaussian distribution. A sparse transform is derived from training data with ICA to map signals to a space where coefficients of speech signals are sparsely distributed, and the coefficients of noise signals have a Gaussian distribution and the distribution of signal is estimated in the sparse space. The performance of sparse coding shrinkage is most efficient in white noise. The disadvantage of sparse coding shrinkage in ICA space is that it needs prior knowledge from training speech or even training noise.

1.5.4.2 *Sparse coding strategies*

Recently, there has been significant development in sparse coding strategies, exploring sparse representations in the context of denoising and classification (Aharon et al., 2006, Zhou et al., 2009, Zhou et al., 2011, Mairal et al., 2010). Sparse coding strategies that model data vectors as sparse linear combinations of basis vectors are widely used in machine learning, neuroscience, signal processing and statistics. Sparse applications exploit the fact that most signals of interest are sparsely represented in an appropriate dictionary or base. An appropriate dictionary could be derived by selecting one from a prespecified set of linear transforms ('off-the-shelf') or adapting the dictionary to a set of training signals ('on-the-shelf'). However, most previous research utilises 'off-the-shelf' wavelet and cosine-transform dictionaries (Ji et al., 2008), but recent research has demonstrated the significant advantages of dictionary learning matched to the signals of interest (Aharon et al., 2006, Zhou et al., 2009, Mairal et al., 2009, Hoyer, 2002, Hoyer, 2004).

There have been several attempts to develop speech enhancement algorithms with sparse coding strategies, e.g. K-SVD (Aharon et al., 2006), online dictionary learning (Mairal et al., 2010)), non-negative matrix factorization (NMF) (Mohammadiha et al., 2012). Sigg applied KSVD in speech enhancement (Sigg et al., 2010). For Gaussian noise, they derived a dictionary from training speech data; for structured noise (nongaussian), they not only learn the dictionary from training speech, but also learn another dictionary from the training noise. The speech and noise matrix are transformed in the short-time Fourier transform (STFT) magnitude domain for sparse coding. Results show that this method can improve speech quality compared to VQ-based (vector quantization based) methods (Sigg et al., 2010) and spectral subtraction (Loizou, 2007). They did not show the performance of KSVD on intelligibility. The disadvantage of sparse coding strategies with on-the-shelf dictionaries is that they need to estimate sparse distribution through training speech data or training noise data. Therefore the method is not very robust in different speech or noise environments.

Another state-of-the-art sparse coding strategy was developed to conduct fast dictionary learning for sparse representations of speech signals (Jafari and Plumbley, 2011). They showed that SNR can be improved with this strategy. However, their evaluation is only objective SNR that is not reliable to predict speech intelligibility or speech quality (Ma et al., 2009). Moreover, subjective evaluation was not conducted. Future research is promising to investigate the performance of such dictionary learning algorithm in speech enhancement thoroughly.

1.5.5 **Discussion**

Previous evaluations of noise reduction strategies in hearing aids mainly focused on spectral subtraction (Levitt et al., 1993, Elberling et al., 1993, Alcantara et al., 2003, Dahlquist et al., 2005) or Wiener filtering algorithms (Levitt et al., 1993). It is worthwhile to test other algorithms that might need more computation, but can be handled with advanced high speed computation technology. Moreover, most noise reduction algorithms were originally developed to improve speech perception for normal hearing subjects, and were later adopted for hearing aid users. Because of hearing loss factors discussed below, algorithms that are optimal for NH listeners may not be optimal for HI listeners. The algorithms for HI may allow more noise reduction even if greater speech distortion is introduced, as HI listeners are more tolerant to speech distortion, but more sensitive to background noise.

We will introduce our development of a noise reduction algorithm with sparse coding shrinkage in PCA space in Chapter 3. It assumes an approximately sparse distribution of speech components in PCA space and performs a shrinkage function in the principal components of noisy speech. The PCA method does not need to accurately estimate the distribution with training data, as is assumed a moderately super-Gaussian distribution. This algorithm is expected to utilize greater levels of noise reduction at the expense of increasing speech distortion to match the speech perception requirements of hearing impaired listeners.

1.6 Introduction of objective evaluation measures

The most accurate speech evaluation measures are through subjective listening tests. However, subjective tests can be time consuming and expensive. Subjective tests recruiting HI listeners become more difficult compared to recruiting NH listeners. Therefore, many objective measures have been developed and used during the signal strategy development stage or prior to subjective listening tests. Some objective measures include predictions for both NH and HI listeners whilst most objective measures include predictions only for NH listeners. If an objective measure has only been validated with NH listeners, it is uncertain whether it is suitable for predicting performance for HI listeners. Furthermore, if an objective measure is only validated for a particular type of speech distortion, e. g. low bit rate, it might not be reliable for evaluating speech distortion introduced by noise reduction algorithms. In this section, contemporary objective measures will be introduced in two categories: speech quality and intelligibility measures respectively. Whether the objective measures have been validated with NH or HI listeners will be stated.

1.6.1 **Objective speech intelligibility measures**

Contemporary objective speech intelligibility measures have been divided into different categories depending on their principal methods, which include articulation-index-based (AI-based) measures (French and Steinberg, 1947, Kryter, 1962, Pavlovic, 1987), speech-transmission-index-based (STI-based) measures (Houtgast and Steeneken, 1973, Houtgast and steeneken, 1985) and coherence-based measures (Kates and Arehart, 2005, Arehart et al., 2007). The articulation index can range from 0 to 1, representing the proportion of the average signal that is audible to an individual. Several mathematical evaluations can be used as an index to predict speech intelligibility, although reliability of these evaluations needs to be investigated further.

The AI measures were further modified to produce the speech intelligibility index (SII) (ANSI, 1997). For a given speech-in-noise condition, the SII is calculated from the speech spectrum, the noise spectrum, and the listener's hearing threshold. Both speech and noise signals are filtered into frequency bands. Within each frequency band, the audibility factor that indicates the degree to which the speech is audible, is derived from the signal-to-noise ratio in that band,. The SII is determined by accumulation of the audibility factor across the different frequency bands, weighted by the band-importance function, which results in a rating value between zero and one. The SII can be regarded as the proportion of the total speech information available to the listener. However, as there might be redundancy in the available information, this may not correlate well with subjective test of intelligibility. Moreover, as estimation of speech and noise cannot be accurate after noise reduction strategies, SII may not be correlated with subjective intelligibility results of noise reduction strategies.

The STI-based measures were further refined with various methods. One example is the normalized covariance metric (NCM) (Hollube and Kollmeier, 1996) that has been validated as one of the most reliable objective quality measures on noise reduction algorithms for NH listeners (Ma et al., 2009). NCM computes the STI as a weighted sum of transmission index (TI) values determined by the covariance between the reference and processed envelope signals in each frequency band. A short-time objective intelligibility (STOI) measure was proposed (Taal et al., 2011a), based on a correlation coefficient between the temporal envelopes of the clean and degraded speech, in short-time overlapping segments. In contrast to other conventional intelligibility models which tend to rely on global statistics across entire sentences, STOI is based on shorter time segments (386 ms). STOI showed high correlation with speech intelligibility of noise reduction algorithms for NH listeners.

One example of coherence-based measures is the magnitude-squared coherence (MSC) function (Kates, 1992). It was used to assess distortion in hearing aids by calculating the normalized cross-spectral density of the reference and the processed signals. Extensions of MSC measure (Kates and Arehart, 2005), called coherence SII (CSII) measures, were proposed to assess the effects of hearing-aid distortions (e.g. peak clipping) on speech intelligibility. CSII proposes to use the SII index as the base measure, and to replace the SNR term with the signal-to-distortion ratio term, which is computed using the coherence between the input and output signals. CSII can evaluate speech intelligibility for NH and HI listeners with hearing thresholds as input arguments.

1.6.2 **Objective speech quality measures**

Contemporary objective quality measures may be divided broadly into two distinct categories, namely, intrusive methods and non-intrusive methods. Intrusive quality measures shown in Figure 1.3 compare the difference between the degraded speech and the original clean speech and then calculate a rating value as output. Non-intrusive methods, on the other hand, calculate quality predictions from the degraded speech only. Significant advances in the field of objective speech quality measures have been made in recent decades; however, only few researchers concentrated on development and validation of objective quality measures for hearing impaired listeners in the application of hearing aids. Even if some objective quality measures were developed to correlate with subjective impression of hearing impaired listeners, they seldom have been validated to assess noise reduction algorithms. This section will first introduce the generalized objective quality measures for NH listeners, and then introduce some objective quality measures that are adapted with arguments of hearing loss characteristics in order to evaluate speech quality impression for HI listeners.

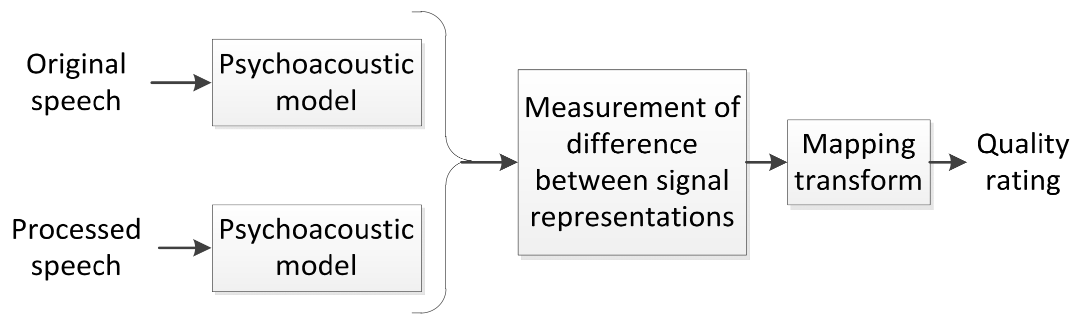


Figure 1.3: Principle and flowchart of intrusive objective measures.

1.6.2.1 *Objective speech quality measures for NH listeners*

When speech is corrupted with noise, an obvious choice to measure speech quality is to measure the level of the clean speech and the embedded noise. Hence a frame-based segmental SNR was proposed (Noll, 1976). Later frequency-weighted segmental SNR (fwsegSNR) was proposed (Tribolet et al., 1978), which calculates the frequency weighted average power difference between signal and noise.

Klatt proposed a weighted spectral slope (WSS) distance measure based on weighted differences between the spectral slopes in each band (Klatt, 1982). This measure was designed to penalise heavily differences between the spectral peak locations. It was motivated by psychoacoustic studies, where subjects assigned the largest distance to pairs of vowels that differed in formant frequencies.

Log-likelihood ratio (LLR) and Itakura-Saito (IS) measures, are two measures that use linear prediction coefficients to predict quality, and are often used to evaluate speech-enhancement algorithms (Quackenbush et al., 1988) for NH listeners. In the Barks spectral distortion (BSD) measure (Wang et al., 1992), both the original and processed signals undergo several stages of the auditory processing and then are compared in the Euclidean distance between their loudness spectra, leading to the so-called "loudness spectra".

Rix and Hollier (2000) developed the perceptual analysis measurement system (PAMS) to evaluate the perceived speech quality of telephone networks. PAMS takes two key network properties, linear filtering and variable bulk delay, into consideration to suit end-to-end measurement. Beerends and Stemerdink (1994) also developed the perceptual speech quality measure (PSQM), which optimized level compression for music and speech coding. PSQM was adopted in 1996 as the International Telecommunication Union's Recommendation ITU-T P.861. Later, PAMS (Rix and Hollier, 2000) and an updated and extended version of PSQM (Beerends and Stemerdink, 1994) were combined to produce the Perceptual Evaluation of Speech Quality (PESQ) (Rix et al., 2001), which superseded ITU's recommendation of PSQM

(ITU-T P.862). PESQ is the objective measure recommended by ITU-T for speech quality assessment of narrow-band handset telephony and narrow-band speech codecs. The basic components of PESQ include time alignment, a psychoacoustic model which maps signals into perceived loudness, disturbance processing, a cognitive model, aggregation of the disturbance in frequency and time, and finally, a mapping to the predicted subjective score. Despite the relative success of PESQ, other researchers proposed different psychoacoustic models and some reproduced internal information available to process in neural stages (Hansen and Kollmeier, 1997, Hansen and Kollmeier, 1999, Dau et al., 1996a, Dau et al., 1996b).

Objective quality criteria, such as segmental SNR, PESQ (perceptual evaluation of speech quality), fwsegSNR (frequency weighted segmental signal to noise ratio) and LLR (log likelihood ratio) have been evaluated with subjective quality rating tests through NH listeners. (Hu and Loizou, 2008). They found that PESQ, LLR and fwsegSNR yielded high correlation with subjective quality tests, while segmental SNR which has been widely used for evaluating the performance of speech enhancement algorithms yielded poor correlation with overall quality.

1.6.2.2 *Adaptation of objective measures for hearing loss*

Several efforts have been made to develop objective speech quality measures for HI listeners. PESQ was extended for HI listeners by adjusting for hearing loss and level variations (Beerends et al., 2008). Beerends developed the perceptual hearing aid quality measure (PHAQM) and developed the model of auditory comfort for hearing impaired persons – revised version (MCHI-R) (Bramslow, 2004, Bramslow, 2008).

Huber and Kollmeier (2006) proposed a measure called PEMO-Q, which is based on Dau et al.'s psychoacoustic model of the "effective" peripheral auditory system (Dau et al., 1996a, Dau et al., 1996b). PEMO-Q (Huber and Kollmeier, 2006) computes a perceptual similarity measure (PSM) as the weighted average of the linear cross correlation coefficient across modulation channels, as well as an instantaneous version of PSM by computing PSM every 10 ms. PEMO-Q was reported to be able to predict quality for narrow-band speech as well as wide-band audio signals. PEMO-Q was also reported to be able to predict listening effort reduction by noise reduction algorithms in hearing aids (Huber et al., 2010).

There have been several recent objective measures that specifically predict quality in the context of hearing aid distortion. Parsa and Jamieson (2001) proposed a measure based on an "auditory distance" parameter, which computes the distance between the hearing aid response and their model output (Parsa and Jamieson, 2001). Recently, Kates and Arehart (2010) developed an objective measure for evaluating distortions introduced specifically by hearing aids for both NH and HI listeners. This metric (Kates and Arehart, 2010), called the Hearing Aid Speech Quality Index (HASQI), aims to capture the aspects of quality deemed important for rating speech processed by hearing aids.

1.6.3 **Objective measures to be used in our thesis**

In this section, we explain specifically the objective speech quality and intelligibility measures that are used in this thesis (Chapter 3) to predict effects of noise reduction algorithms for NH or HI listeners. The objective measures include PESQ, fwsegSNR, HASQI, NCM and STOI. Although PESQ and fwsegSNR have not been validated through HI listeners, they are well-recognized quality measures for NH listeners. FwsegSNR has been validated to reflect the physical noise reduction effects with NH listeners. PESQ has been recommended by ITU-T to assess speech quality of narrow-band handset telephony and narrow-band speech codec and has been validated to assess speech quality by noise reduction algorithms with NH listeners. Although some effort has been made to develop objective quality measures for HI listeners, their performance in evaluating speech quality effects by noise reduction algorithms for HI listeners are still uncertain. HASQI is a speech quality metric for HI listeners and it incorporates hearing threshold values into the evaluation model. NCM and STOI are two objective intelligibility measures for noise reduction algorithms validated with NH listeners. The reason to present objective measures only validated with NH listeners is to compare these objective results with subjective results from NH listeners.

We have chosen two objective quality measures for NH listeners, one objective quality measure for HI listeners and two objective intelligibility measures for NH listeners. We have not chosen any objective intelligibility measure for HI listeners as there are few publicly available codes that are reliable to reflect intelligibility effects of noise reduction algorithms for HI listeners.

1.6.3.1 *Perceptual Evaluation of Speech Quality (PESQ)*

The perceptual evaluation of speech quality (PESQ) (Rix et al., 2001, ITU, 2000) was selected as the ITU-T recommendation P.862. This measure includes distortions commonly encountered when speech goes through telecommunication networks. The original and degraded signals are first level-equalized to a standard listening level, and filtered through a standard telephone handset filtering system. The signals are aligned in time to correct for time delays, and then processed through an auditory transform to obtain the loudness spectra. The difference between the loudness spectra is computed. The PESQ produces a score between 1.0 and 4.5, with high values indicating better quality. PESQ is validated as in good correlation with subjective quality measure (Rix et al., 2001, Hu and Loizou, 2008) for NH listeners. The MATLAB code of PESQ was adopted from (Loizou, 2007).

1.6.3.2 *Frequency-weighted segmental SNR (fwsegSNR) Measure*

The frequency-weighted segmental SNR (fwsegSNR) was computed using the following equation (Hu and Loizou, 2008)

$$fwsegSNR = \frac{10}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{K} W(j,m)\log_{10}\dfrac{X(j,m)^2}{(X(j,m)-\hat{X}(j,m))^2}}{\sum_{j=1}^{K} W(j,m)} \tag{1.1}$$

where $W(j,m)$ is the weight placed on the jth frequency band, K is the number of bands, M is the total number of frames in the signal, $X(j,m)$ is the critical-band magnitude (excitation spectrum) of the clean signal in the jth frequency band at the mth frame, and $\hat{X}(j,m)$ is the corresponding spectral magnitude of the enhanced signal in the same band. The critical-band spectra $X(j,m)$ were obtained by multiplying the FFT magnitude spectra by 25 overlapping

Gaussian-shaped windows (Loizou, 2007) spaced in proportion to the ear's critical bands and summing up the power within each band. Similar to the implementation in (Hu and Loizou, 2008), the excitation spectra were normalized to have an area of unity. This measure has been found to be in good correlation with both subjective quality and intelligibility measures (Hu and Loizou, 2008). The MATLAB code of fwsegSNR was adopted from (Loizou, 2007).

1.6.3.3 *Hearing Aid Speech Quality Index (HASQI)*

Kates and Arehart (2010) proposed the Hearing Aid Speech Quality Index (HASQI) to evaluate speech quality with distortions introduced by hearing aids for both NH and HI listeners. This metric starts with a cochlear model that incorporates aspects of impaired hearing and then extracts signal features related to quality judgments. One set of features measures effects of noise and nonlinear distortion on speech quality, and the other set of features measures the effects of linear filtering. The final index is the multiplicative combination of the nonlinear effects and linear filtering effects. The MATLAB code of HASQI was provided by James M. Kates.

1.6.3.4 *NCM*

The NCM measure was computed as follows (Ma et al., 2009). The stimuli were first processed through K band pass filters. The normalized covariance between the envelope of the original clean signal and the envelope of the processed signal is then calculated in each frequency band. The SNR in each band is computed with the normalized covariance. The transmission index (TI) in each band is computed by linearly mapping the SNR values between 0 and 1 to the TI. Finally, the transmission indices are averaged across all frequency bands with band-importance weights to produce the NCM index. The MATLAB code of NCM was provided by P. Loizou.

1.6.3.5 *STOI*

The short-time objective intelligibility (STOI) measure (Taal et al., 2011a) is based on short-time correlations between the clean speech and the processed speech, in contrast to other conventional intelligibility models that tend to rely on global statistics across entire sentences. STOI is calculated as follows. Both clean and processed stimuli are first decomposed into DFT-based, one-third octave bands. Next, short-time temporal envelope segments of the clean and processed speech are compared by means of a correlation coefficient. Before comparison, the short-time processed speech temporal envelopes are first normalized and clipped. The short-time intermediate intelligibility measures are then averaged to a rating value, which is expected to reflect speech intelligibility. The MATLAB code of STOI was downloaded from the following website[3].

1.6.4 **Discussion**

As objective measures are usually used during the algorithm development stage, as an intermediate evaluation tool prior to time-consuming subjective tests, they need to be reliable to guide new speech enhancement algorithms to be improved in the right direction.

Most contemporary objective measures are only suitable for NH rather than HI listeners. Objective measures for HI should take account of impaired auditory factors, e.g. threshold elevation; reduced frequency selectivity; reduced temporal resolution. Although there have been several efforts to develop objective measures for HI listeners (Kates and Arehart, 2010, Hollube and Kollmeier, 1996), they have only been validated with clean speech or speech in noise rather than with the noise reduction algorithms, which may introduce some distortion that makes the objective evaluations more difficult.

According to current availability and reliability of objective measures for NH or HI subjects, only five objective measures were chosen, including two objective quality measures (fwsegSNR, PESQ) and two objective intelligibility measures (NCM, STOI) for NH listeners, and one objective intelligibility measure (HASQI) for HI listeners. As we focus on subjective tests in this thesis, objective measures are only used to be compared and validated with subjective tests.

[3] http://siplab.tudelft.nl/content/software-and-data-resources

An alternative way to obtain objective measures for HI subjects is to simulate HI subjects' listening using NH subjects. The rationale of this method is as follows. If the consequences of the hearing impaired listening can be viewed as an additive combination of the hearing impaired distortion with the normal hearing listening, and if the hearing impaired distortion can be simulated accurately through the hearing loss simulation model (Baer and Moore, 1993, Moore and Glasberg, 1993, Nejime and Moore, 1997, Hu et al., 2011b), the effects of asking HI listeners to listen to speech can be approximated by asking NH listeners to listen to the same speech that processed with the hearing loss simulation. An auditory-based hearing loss simulation model is introduced in Chapter 2. The experiment to evaluate noise reduction algorithms with a combination of the hearing loss simulation model and NH listeners' participation is introduced in Chapter 4.

1.7 Introduction of subjective evaluation tests

A subjective evaluation test uses either NH or HI subjects to listen to speech, and to give feedback by either repeating what they have heard or giving a specific judgment. It is more time consuming but is more ecologically valid if a sufficient number of subjects take part in the tests compared to objective tests.

Quality and intelligibility are two different important attributes of speech and criteria should be set up to evaluate whether noise reduction algorithms could improve speech quality or intelligibility. Quality is highly subjective and variable in nature and is difficult to judge which is better as individual listeners have different perceptual criteria to assess sound quality. Intelligibility can be measured by asking listeners to identify the words. It is easier to perform subjective tests on intelligibility compared to quality. This section will introduce the basic methods for subjective speech intelligibility and quality tests.

1.7.1 **Subjective speech intelligibility tests**

There are numerous factors affecting the reliability of subjective intelligibility tests. Factors to be considered include good representation of all major speech phonemes, equal quality of test lists and control of contextual information. Speech tests proposed over the years could be divided into three categories: (1) recognition of syllables made up of meaningless combinations of speech sounds (Miller and Nicely, 1955, Fletcher and Steinberg, 1930), (2) recognition of single meaningful words (Voiers, 1983, Fairbanks, 1958, House et al., 1965), (3) recognition of meaningful sentences containing all contextual information among words (Nilsson et al., 1994). Each of these tests has its advantages and disadvantages. Choice of a speech test type depends on specific situations.

In most of the subjective intelligibility tests, speech intelligibility is often quantified in terms of percentage of identified words (or syllables) correct. Percentage intelligibility is often measured at fixed speech or noise levels, or fixed input SNRs. Such intelligibility measures are inherently limited by subjective floor or ceiling performance at different input SNRs. A more reliable measure for assessing speech intelligibility that is not sensitive to the presentation level of speech or noise or input SNR is recommended as the speech reception threshold (SRT).

SRT is an important criterion to assess speech intelligibility. The SRT can be measured either in quiet or in noise. In a quiet environment, it is defined as the presentation level at which listeners identify words with a fixed percentage that are correct. In noise, the SRT is defined as the signal-to-noise ratio at which listeners identify words with a fixed percentage threshold of accuracy. A practical and efficient method is developed for obtaining the SRT (Dirks et al., 1982, Levitt, 1971). It uses an adaptive method known as the up-down procedure which is designed to be accurate and efficient. The one-up-one-down procedure can find the threshold of 50% correct whereas the three-up-one-down procedure can find the threshold of 79.4% correct. In the three-up-one-down procedure, the subject must make three consecutive correct judgement s in order to cause the staircase magnitude to go up. Any incorrect judgment, at any time, causes the staircase to go down. This staircase settles at the magnitude where the threefold probability is 50% or the probability is 79.4%. The three-up-one-down procedure is adopted in the subjective speech recognition tests in Chapter 3&4 in order to find the threshold of input SNR which corresponds to 79.4% correct recognition.

1.7.2 **Subjective speech quality tests**

A subjective sound quality test is to ask listeners to listen to speech and give feedback, which is much more time consuming than objective sound quality measures.

The methods for evaluating subjective speech quality can be broadly classified into two categories: those based on relative preference tasks (Munson and Karlin, 1962, Hecker and Williams, 1966, Combescure et al., 1982) and those based on assigning a numerical value to the quality of the speech stimuli (Hu and Loizou, 2007). In relative preference tests, listeners are presented with a pair of speech stimuli consisting of the test stimuli and the reference stimuli. Listeners are asked which stimuli they prefer. In the rating tests, listeners are asked to rate the quality of the stimuli on a numerical scale, typically, a five-point scale with 1 indicating poor quality and 5 indicating excellent quality. As individual listeners have different perception or criteria of rating quality, there will be a large variance in rating tests.

A new method called Interpolated Paired Comparison Rating (IPCR) (Dahlquist et al., 2005) can be regarded as a combination of the relative preference task and the stimuli rating task. In IPCR tests, listeners are presented with a pair of speech stimuli and asked to give a rating value based on how much they prefer one stimulus B to the other stimulus A. If they prefer B to A, they give a rating distance value between 0 and 10; if they prefer A to B, they give a rating distance value between -10 and 0. It was developed to quickly derive the difference in signal-to-noise ratios between enhanced and unprocessed speech that give equal subjective sound quality impression. One direct but time-consuming way to measure such SNR-gain is to adaptively compare between enhanced stimuli with a fixed SNR and an unprocessed signal with a variable SNR. IPCR is performed by comparison between processed stimuli with fixed SNR and unprocessed stimuli with only two different SNRs. A rating value between -10 and 10 is given as subjective impression feedback of preference distance between the enhanced stimuli and unprocessed stimuli. Interpolation and extrapolation was used to find the SNR-gain of subjective equality of quality impression. Accordingly, IPCR is an efficient way to test subjective SNR-gain for noise reduction algorithms and can be compared with objective SNR improvement on the same scale in dB. Specific details were described (Dahlquist et al., 2005) and also discussed in Chapter 4. In the thesis, we tested subjective speech quality effects using noise reduction algorithms through IPCR tests.

1.7.3 **Discussion**

Strategy evaluation is important as it reflects the reliable perception results. Inaccurate evaluation will mislead the judgement and development of algorithms. Different signal processing strategies serve different purposes. To evaluate whether the signal processing strategy can achieve its developing motivation, we can set up an appropriate test to detect signal processing effects. For speech intelligibility or quality tests, many factors should be considered, e.g. speech materials, noise types, SNRs, hearing threshold levels of subjects, age of subjects, instructions and information sheets for subjects, specific procedure and testing order, etc. After testing, appropriate statistical analysis needs to be performed. Implications and tips for further development of strategies can also be acquired from subjective testing results.

1.8 Outline of this thesis

Figure 1.4 gives an overview of the remaining chapters in this thesis. The different colours indicate different studies, all of which are combined to investigate whether noise reduction algorithms can benefit the HI listeners. We first developed a hearing loss simulation (HLS) model to understand hearing loss factors and consequences. A deep understanding of the hearing loss mechanism can motivate appropriate strategies to compensate for hearing loss factors. Then, we developed a noise reduction algorithm called sparse coding shrinkage (SCS) that is motivated to present sparse stimuli to hearing impaired listeners. During the development stage, the algorithm is evaluated with objective measures. During test stage, evaluations are performed through subjective tests with NH and HI listeners. Evaluations are also performed through the HLS model with NH subjects to mimic speech perception in HI listeners. Through the evaluation results, several questions are going to be answered: whether the results of evaluation with the HLS model correlate with the evaluation with HI listeners; whether the results with HI listeners differ from the results with NH listeners in speech intelligibility or quality with background of different noise types; whether the objective measures correlate with the subjective results; whether noise reduction algorithms benefit HI subjects.

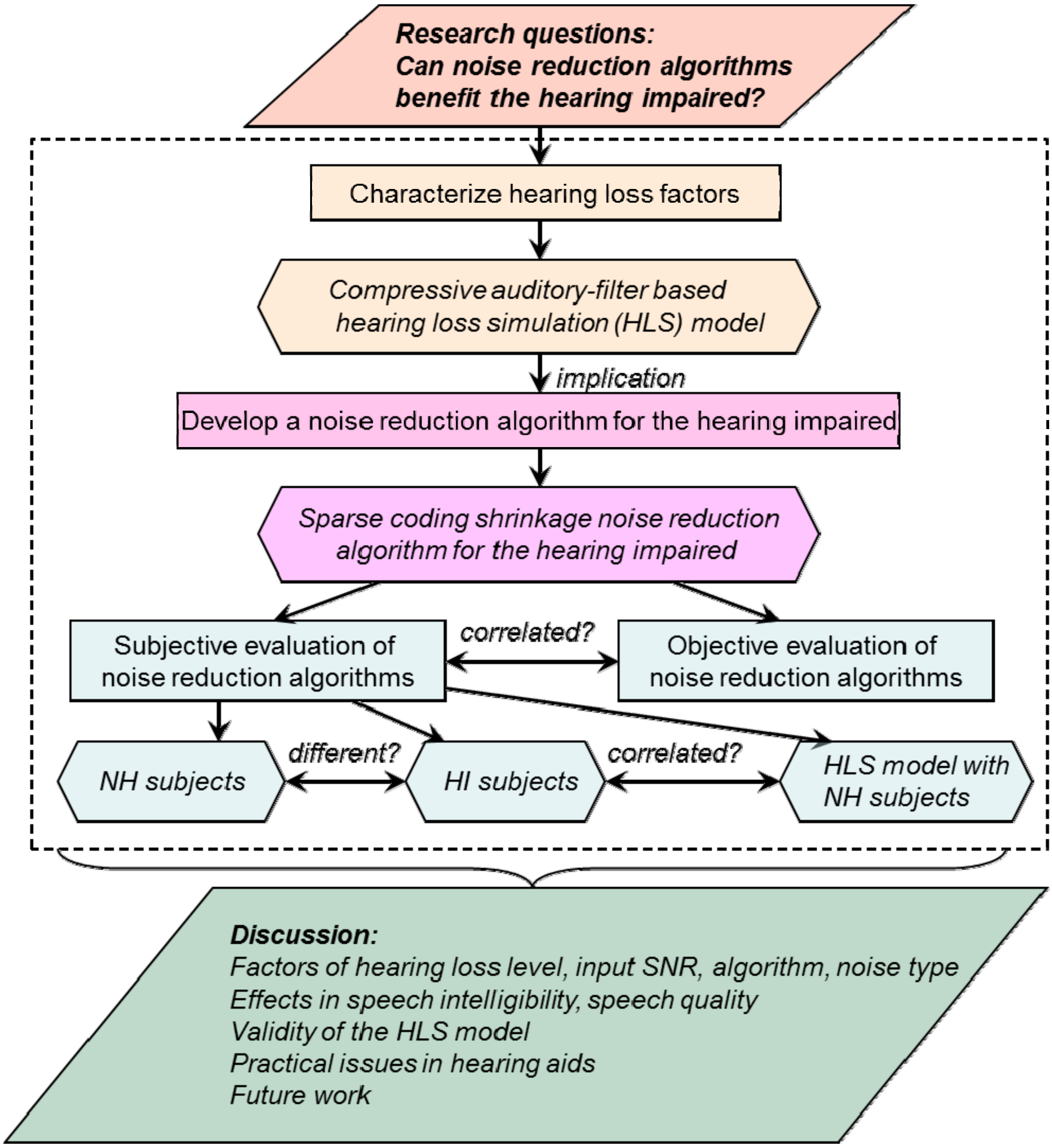


Figure 1.4: Outline of the remaining chapters of this thesis. The three studies are divided in

different colours shown in the dashed rectangular diagram. The rectangle indicates the topic of each study. The hexagon indicates the specific content of each study. Research questions and discussion are presented at the beginning and the end of this flowchart.

In Chapter 2, an auditory-filter-based hearing loss simulation model is developed to characterize the hearing loss factors and predict perceptual consequences of hearing impairment. This HLS model can simulate threshold elevation, loudness recruitment, reduced frequency selectivity and reduced dynamic range. This hearing loss simulation was developed in our group (Hu et al., 2011b). This model differs from previous HLS models in its adoption of a nonlinear physiologically-inspired auditory filter bank, called the gamma-chirp filter bank. The gamma-chirp filter bank adds a level-dependent asymmetric correction to the basic gammatone channel frequency response, thereby providing a yet more accurate approximation to the auditory frequency response (Irino and Patterson, 2001, Patterson et al., 2003, Irino and Patterson, 2006). Understanding the hearing loss mechanism through the HLS model should be a priority as it can give implications for development of noise reduction algorithms for HI listeners.

In Chapter 3, a sparse coding shrinkage noise reduction algorithm is proposed for hearing impaired subjects. In our group, sparse stimuli have shown benefits for cochlear implants users who are profoundly hearing impaired listeners (Li and Lutman, 2008, Hu et al., 2011a). This inspires us to further evaluate whether sparse stimuli are also beneficial for mildly to severely hearing impaired listeners who are usually hearing aid users. The noise reduction algorithm is performed by applying a sparse coding shrinkage function (Hyvärinen, 1999) to the noisy speech components in PCA space. Objective measures are presented in this chapter as a quick assessment of noise reduction algorithms in speech intelligibility and speech quality. However, most of the objective measures have only been validated with NH listeners; they might not be reliable to evaluate noise reduction algorithms for HI listeners.

In Chapter 4, subjective tests are presented to evaluate noise reduction algorithms for NH and HI listeners in speech quality and speech intelligibility. SCS is compared with CS-WF and noisy speech. CS-WF is selected as it is a competitive state-of-the-art noise reduction algorithm for NH listeners (Breithaupt et al., 2008, Gerkmann and Hendriks, 2011, Gerkmann and Hendriks, 2012). The subjective results will be analysed to answer several questions: whether noise reduction algorithms show different effects between HI subjects and NH subjects; whether SCS and CS-WF show different perceptual effects for HI or NH subjects; whether there is any advantage of sparse stimuli for HI subjects.

In Chapter 5, we will perform an experiment to evaluate noise reduction effects through the HLS model using NH subjects. The motivation is to validate the HLS model in predicting speech intelligibility effects for HI listeners. The assumption is that consequences of hearing impaired listening can be an additive combination of hearing impaired distortion with normal hearing listening. By simulating hearing impaired distortion in an appropriate and realistic way, the effects of listening to speech with HI listeners can be approximated by asking NH listeners to listen to the same speech processed with the hearing loss simulation. Through comparison between the subjective results with the HLS model in Chapter 5 and with HI subjects in Chapter 4, conclusions can be made in two aspects: whether the HLS model can predict noise reduction effects for HI listeners and whether noise reduction algorithms can bring intelligibility benefits for HI listeners if the HLS model is realistic in predicting intelligibility. The purpose of this experiment is to set up a test platform to predict noise reduction algorithms for HI listeners through NH listeners which can be much easier due to the difficulty in recruiting HI listeners and the large variance among HI listeners. The evaluation platform using the HLS model and NH subjects can predict intelligibility effects with noise reduction algorithms for the specific level of hearing impairment.

In Chapter 6, we discuss several factors that determine the performance of noise reduction algorithms. The effects include the hearing loss level, the input SNR, the algorithm itself, the noise type, etc. We also discuss the validity of the HLS model in predicting intelligibility performance of noise reduction algorithms for HI listeners. Limitation of this research is pointed out. Future work is also suggested. In Chapter 7, general conclusions are drawn.

Chapter 2 Characterization and prediction of perceptual consequences of hearing impairment

While beneficial, hearing aids do not restore normal audition. The varied successes of signal processing strategies for hearing aid design reflect our incomplete understanding or compensation of impaired auditory profiles that cause decreased speech perception in individual hearing impaired listeners, especially in noisy environments.

Prior to developing the right compensation strategy for HI listeners one needs to understand the hearing loss mechanism. One way to understand hearing loss mechanism and quantify hearing loss consequences is to develop a hearing loss simulation (HLS) model. A realistic HLS model can characterize hearing loss factors quantitatively and predict perceptual consequences of hearing impairment. A HLS model based on a compressive gammachirp auditory filter bank was developed in our group (Hu et al., 2011b) and described in this chapter. The HLS model will also be used in Chapter 3 to represent speech distortion by hearing impairment in speech spectrograms.

The HLS model can simulate different hearing loss factors such as threshold elevation, loudness recruitment and reduced frequency selectivity. One distinct character of the HLS model compared to other HLS models is its adoption of a compressive gammachirp auditory filter bank. Compressive gammachirp filters are physiologically motivated in the sense that they simulate the main passive and active mechanical processes in the basilar membrane (Irino and Patterson, 2001, Patterson et al., 2003), while most other HLS models use a FFT filter bank or a gammatone filter bank. The simulations of threshold elevation, loudness recruitment and reduced frequency selectivity were contributed by the author.

If the HLS model is realistic, the effects of normal hearing (NH) subjects listening to speech processed with the HLS could approximate the effects of hearing impaired (HI) subjects listening to unprocessed speech. The HLS model is expected to have implications for advanced compensation strategies and noise reduction algorithms in hearing aids. The HLS model can be used as an evaluation platform to assess effects of signal processing strategies for HI subjects through NH listeners. This platform can be adopted as a quick way prior to final subjective evaluations of signal processing strategies using HI subjects.

2.1 Introduction

2.1.1 **Introduction of impaired auditory factors**

2.1.1.1 *Hearing Threshold Elevation*

Hearing threshold is the minimum sound level of a pure tone that an individual ear can hear with no other sound present. Individual hearing threshold is usually measured through pure tone audiometry (BSA, 2011), which is a subjective measure of individual response to pure tone stimuli with different frequencies (250, 500, 1000, 2000, 4000, 8000 Hz). For HI listeners, one of their impaired auditory factors is shown as hearing threshold elevation. Individual hearing threshold has been used as a traditional reference for selecting an appropriate hearing aid.

The relationship between hearing threshold and speech perception has been studied by many researches through simulation studies (Fabry and van Tasell, 1986, Humes et al., 1987, Zurek and Delhorne, 1987, Dubno and Schaefer, 1992, Patterson et al., 1982, Lutman, 1991). Simulation of threshold elevation can be achieved by selective filtering or using masking noise. The stimuli processed with threshold elevation are tested by normal-hearing subjects. Provided the simulation is accurate, this makes it possible to study the effect of hearing threshold elevation in isolation. Most of these studies have implied that hearing threshold elevation is not the only factor for the relatively poor speech perception in noise of the hearing-impaired subjects.

2.1.1.2 *Loudness recruitment*

When hearing loss is present, the perception of loudness is altered. Sounds at low levels are no longer audible to the HI listeners but are still audible to the normal hearing (NH) listeners. However, sounds at high levels often are perceived as having the same loudness for a HI listener as they would for a NH listener. This phenomenon indicates that loudness grows more rapidly for HI listeners than normal listeners with the audible sound level. This phenomenon is called loudness recruitment. Loudness recruitment is associated with threshold elevation. Simulation of threshold elevation is also often accompanied with simulation of loudness recruitment in previous studies (Moore and Glasberg, 1993).

2.1.1.3 *Reduced frequency resolution*

Frequency resolution is defined as hearing ability to detect signal at one frequency in the presence of sound at a different frequency. It involves separating out spectral components of a sound. It reflects the sharpness of auditory filter shapes. For hearing impaired people, their auditory filter shapes are blunt (Moore, 2007) and thus their frequency resolution ability is limited. Measurement of frequency resolution can be performed through psychophysical tuning curves (PTC) or notch noise masking (Lutman et al., 1991).

Reduced frequency resolution can induce severe disruption of speech perception in noise which is one of the most severe difficulties of HI listeners (Moore, 2007). There are currently no optimal solutions to compensate for reduced frequency resolution. Some research has attempted frequency sharpening initially to compensate for broadened auditory filter shapes but no intelligibility improvement has been shown (Baer and Moore, 1993, Alcantara et al., 1994, Dillon, 2001).

Correlation studies (Glasberg and Moore, 1989) were used to investigate the relation between the auditory factors and speech perception. Correlation studies implied that in high levels of background noise, speech intelligibility is determined more by supra-threshold psychoacoustic factors such as frequency resolution than by hearing threshold.

2.1.1.4 *Temporal resolution*

Temporal resolution is defined as the ability to detect changes over time. In previous research (Moore, 2007), a temporal resolution model was described in terms of a four-stage model consisting of an array of band-pass filters (the auditory filters) each followed by a compressive non-linearity, a sliding temporal integrator and a decision device. It was argued that the cochlear hearing loss could affect the first two stages, the filters and the non-linearity, but that it usually did not affect the last two.

Some researchers (Drullman et al., 1994, Hou and Pavlovic, 1994) have taken the approach of simulating the effects of reduced temporal resolution to estimate the relative importance of temporal modulations at different rates for speech intelligibility. Drullman carried out a study of the effects of the temporal envelope smearing on speech intelligibility (Drullman et al., 1994). Their result suggests that reduced temporal resolution doesn't reduce speech intelligibility for most hearing-impaired people. Schijindel et al. (1994) also found that the detection thresholds for HI listeners with respect to temporal information were not significant differently different from those for NH listeners. That is also why we did not take into account reduced temporal resolution in our novel HLS model described in 2.2. Our HLS model was initially developed to predict intelligibility of speech in noise with sensorineural hearing impairment.

2.1.2 **Introduction of auditory filter banks**

Auditory filter banks are non-uniform band-pass filter banks designed to imitate the frequency resolution in the human auditory system. Derivation of auditory filters is based on psychoacoustic and physiological measurements (De Boer and Kuyper, 1968), leading to approximations of the auditory filter frequency response in terms of a Gaussian function (Patterson, 1976), a "rounded exponential" (roex) (Patterson et al., 1982), and more recently the gammatone filter bank (Patterson et al., 1995). Psychophysical measurements of the auditory filter shape, however, indicate that the filter is approximately symmetric only at low stimulus levels, but asymmetric at high stimulus level with the low-frequency skirt shallower than the high frequency skirt (Lutfi and Patterson, 1984, Rosen and Baker, 1994). These findings also are consistent with physiological observations of basilar membrane motion (Patterson and Moore, 1986). The gamma-chirp filter bank further adds a level-dependent asymmetric correction to the basic gammatone channel frequency response, thereby providing a yet more accurate approximation to the auditory frequency response (Irino and Patterson, 2001, Patterson et al., 2003, Irino and Patterson, 2006). This gammachirp filter bank is firstly adopted in a hearing loss simulation model in our research (Hu et al., 2011b) to imitate nonlinear asymmetric human auditory filtering more realistically compared to other auditory filter banks (Moore and Glasberg, 1993).

2.2 Development of Hearing Loss Simulation (HLS) Model

2.2.1 Framework of compressive gamma-chirp auditory filter based HLS

The most obvious perceptual consequence of cochlear damage is hearing threshold elevation (Moore, 2007). There are also supra-threshold effects of hearing loss, specifically loudness recruitment, reduced spectral and temporal resolution, and reduced dynamic range. A complete simulation of hearing loss should incorporate all the important auditory deficits that have been observed in HI listeners. Some effort has been made (Nejime and Moore, 1997) to combine different psychoacoustic factors together. However, in their simulation the signal has been filtered twice with two different filter banks. This double filtering is in contradiction with what is known in the cochlear mechanism that signals are only filtered once in the basilar membrane. Our HLS system was developed by combining a gammachirp auditory filter bank with traditional HLS methods: the input signal was divided into 26 sub-bands by the gammachirp auditory filter bank and then digital signal processing methods were applied in each band to realise reduced frequency selectivity, loudness recruitment and hearing threshold elevation.

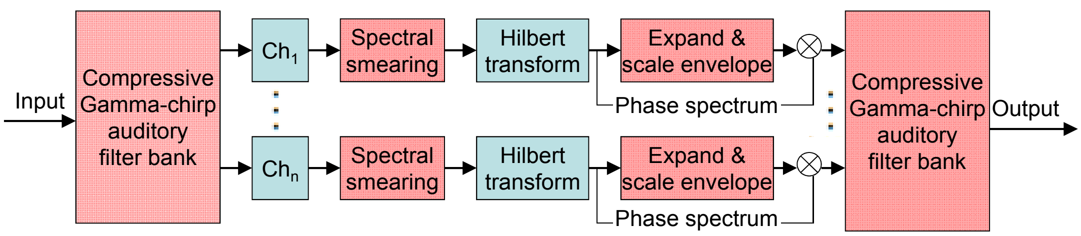


Figure 2.1: Framework of a compressive-gammachirp-auditory-filter based hearing loss simulation model.

Figure 2.1 describes the framework of compressive gammachirp auditory filter based HLS model. It is assumed that impaired ear listening can be regarded as the additive combination effects of impaired speech distortion and normal ear listening. If the HLS model can mimic the impaired speech distortion, the output signal in Figure 2.1 is expected to produce the same excitation pattern in a normal ear as an impaired ear would receive the input signal in Figure 2.1.

2.2.2 **Gammachirp auditory filer**

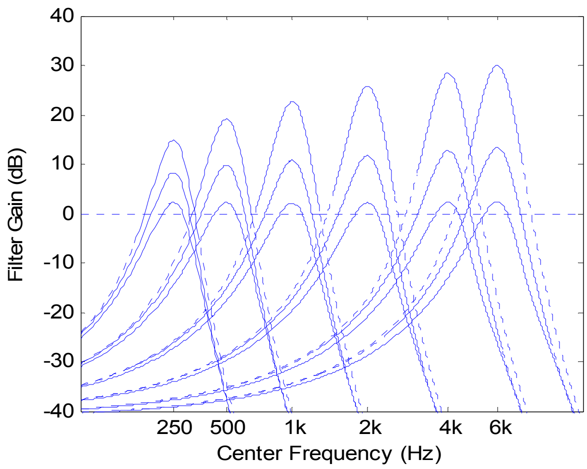


Figure 2.2: An example of six compressive gammachirp filters with different centre frequencies (250, 500, 1000, 2000, 4000 and 6000 Hz) at three sound pressure levels (40, 60, 80 dB). The lower the sound pressure level, the sharper the filter.

As mentioned in section 2.1.2, the gammachirp filter was introduced by Irino and Patterson (Irino and Patterson, 1997) and successively improved (Irino and Patterson, 2001, Patterson et al., 2003) to describe auditory filters. Irino and Patterson (Irino and Patterson, 2001, Patterson et al., 2003) demonstrated that the dynamic compressive gammachirp filter provides a good description of observed human auditory filter shapes. Its main advantages against traditional auditory filters are that its amplitude spectrum is asymmetric and level dependent. Compressive gammachirp filters are physiologically motivated in the sense that they simulate the main passive and active mechanical processes in the basilar membrane. It is therefore a functional model to simulate and to describe observed auditory filter shapes and excitation patterns. Gammachirp filter banks that are based on NH listeners' responses are well developed and are used in our model directly (Irino and Patterson, 2001, Patterson et al., 2003). Although the shapes of the auditory filters in HI listeners are different from those in NH listeners, our correction of the broadened auditory filter shape will be processed in the stage of spectral smearing in Figure 2.1. As currently there were no broadened Gamma-chirp filters for HI subjects, the correction is necessary to simulate the broadened shape of the impaired auditory filters.

In our model the auditory filter bank is simulated by 26 compressive gammachirp filters with centre frequencies spaced logarithmically between 100 Hz and 8000 Hz. Figure 2.2 shows six examples of the filter shapes at different centre frequencies (250, 500, 1000, 2000, 4000 and 6000 Hz) at three different sound pressure levels (40, 60, 80 dB) (Irino and Patterson, 1997). This figure shows the nonlinear level-dependent and asymmetric property of gammachirp filters which can better mimic the physiological mechanism of the basilar membrane. Conventional gammatone filters (Patterson et al., 1995) are symmetric and level-independent which can only simulate the filtering of sound in the basilar membrane at low sound levels.

The analytic complex form of the gammachirp auditory filter is (Irino and Patterson, 1997)

$$g_c(t) = at^{n_1-1}\exp(-2\pi b_1 ERB(f_{r1})t) \times \exp(j2\pi f_{r1}t + jc_1 \ln t + j\phi_1) \tag{2.1}$$

where t>0; a is the amplitude; n_1 and b_1 are parameters defining the envelope of the gamma distribution; c_1 is the chirp factor (Patterson et al., 1995). f_{r1} is the asymptotic frequency; $ERB(f_{r1})$ is the equivalent rectangular bandwidth (Moore et al., 1990); ϕ_1 is the initial phase; and $\ln t$ is the natural logarithm of time. For specific details refer to Irino and Patterson (1997).

An auditory filter bank consists of non-uniform band-pass filters and is designed to imitate the frequency resolution of human hearing. In our HLS model, gammachirp filter banks are used with centre frequencies spaced logarithmically between 100 Hz and 8000 Hz.

2.2.3 **Simulation of hearing threshold elevation and loudness recruitment**

Simulation of threshold elevation and loudness recruitment is implemented in the block of expanding and scaling spectral envelope in Figure 2.1. It is a modified method based on the principle of envelope expansion and scaling in a previous simulation (Moore and Glasberg, 1993). In the envelope of signal from each auditory filter band, the level of processed stimulus is calculated as L_p in dB HL, the level of unprocessed stimulus as L_u in dB HL, the hearing threshold level as L_T in dB HL and the threshold of discomfort as L_C (also called the uncomfortable loudness level) in dB HL. Figure 2.3 (a) shows the assumed audiogram of the moderate high frequency sloping hearing loss. The hearing thresholds are 40 dB HL at low frequencies (0.25, 0.5, 1 kHz) and 55 dB HL at high frequencies (2, 4, 8 kHz). The uncomfortable levels are 110 dB HL across all the frequencies. Figure 2.3 (b) shows the principal relationship between the processed envelope and unprocessed envelope in our HLS model. This input/output (unprocessed/processed) function can simulate the phenomenon that a hearing impaired individual could not hear the sound until the sound reaches the hearing threshold level and shows the similar uncomfortable loudness as a normal hearing individual when the sound level is very high. In Moore's simulation model (Moore and Glasberg, 1993), the input L_u and the output L_p are in unvaried linear relationship which is not realistic when the sound is too loud for a listener to tolerate. Therefore the proposed nonlinear model imposes a threshold at the uncomfortable level and sound will be clipped when it is over the uncomfortable loudness level. The nonlinear model also imposes the sound output level to be near 0 dB HL to mimic little sound perception as long as the sound level is below the hearing threshold. Therefore, the model can simulate the subjective perception of very loud and very quiet sounds.

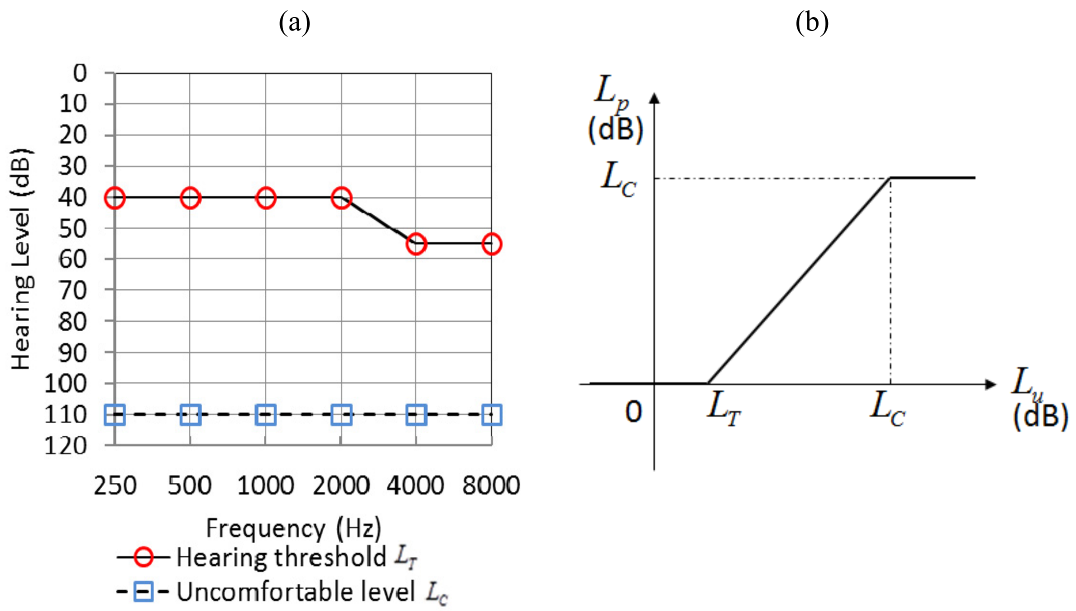


Figure 2.3: Illustration of the mechanism how to simulate the threshold elevation and loudness recruitment. (a) The assumed audiogram with hearing thresholds and uncomfortable levels across audiometric frequencies in the HLS model; (b) the input-output function to simulate threshold elevation and loudness recruitment in the HLS model. The hearing threshold and uncomfortable level are illustrated in both (a) and (b).

2.2.4 **Simulation of reduced frequency resolution**

Spectral smearing is implemented to simulate reduced frequency resolution. Reduced frequency resolution means the ability to detect signal at one frequency in the presence of sound at a different frequency. In each channel, the waveform of the input stimuli was processed frame by frame. For each of the analysis/synthesis frames in each channel, the short-term spectrum was calculated using a Hamming window and a fast Fourier transform (FFT). The length of the Hamming window was 128 samples at a sample rate of 16 kHz, corresponding to an input frame size of 8 ms. The size of FFT was chosen to be reasonably small so as to limit computational complexity, while still sufficiently long to encompass a typical pitch period and give reasonable frequency resolution. Spectral smearing was performed by replacing each component of the power spectrum with a weighted sum of the surrounding components (Baer and Moore, 1993). The weighting function was similar to the shape of the broadened impaired auditory filter centred on the component. The broadened roex function which simply simulates broadened auditory filter shape in the frequency domain was selected as the weighting function as in Baer and Moore (1993).

2.2.5 **Objective hearing loss simulation results**

Figure 2.4 shows waveforms and the spectrograms of the original clean and the noisy Bamford-Kowal-Bench (BKB) sentence: "*she drinks from her cup"* processed with or without the hearing loss simulation. The BKB sentences consist of 21 lists with 16 sentences in each list. These sentences can be used to test speech recognition by calculating the proportion of recognized keywords. The waveforms and spectrograms are shown for (a) original clean speech; (b) the same clean speech processed with the hearing loss simulation; (c) the same clean speech with speech shaped noise (0 dB input SNR); (d) the same noisy speech processed with the hearing loss simulation. Speech shaped noise is a stationary noise which is produced by filtering the white Gaussian noise with a filter that shows the similar frequency response as the spectrum of the speech being used. Through this way, the speech shaped noise has the similar long term spectrum as the speech being used. All the waveforms are plotted with the same amplitude scale to indicate the relative amplitudes. All the spectrograms are plotted with the same colour scale (colour scale corresponds to amplitude power level scale) to also indicate the relative levels. Comparison in waveforms between (a) and (b) or between (c) and (d) shows that perceived sound levels are reduced due to threshold elevation and loudness recruitment. Comparison in spectrograms between (a) and (b) or between (c) and (d) shows that spectral information is smeared and important speech formants are blurred in (b) and (d) due to reduced frequency selectivity. Comparison of spectrograms between (b) and (d) shows that HLS model is very sensitive to noise and the information of noisy speech will be blurred and deteriorated severely by the HLS model. Assuming the HLS model is realistic, this also implies the effects of noise on hearing impaired ear and why HI listeners complain communication in noise environment. Through informal listening, although clean speech processed with the HLS model can still be understood; noisy speech (0 dB SNR) with the HLS model is difficult to understand. Listening to speech (with or without noise) with the HLS model could give us an impression of the difficulty experienced by HI listeners when listening to speech. During the signal processing stage, the perceptual consequences of the enhanced speech for HI listeners can be predicted by processing the enhanced speech with the HLS model and playing to the normal ear. Therefore, the HLS model is potentially an effective tool in evaluating speech perceptual effects for the impaired ear through speech spectrograms or playing to NH listeners.

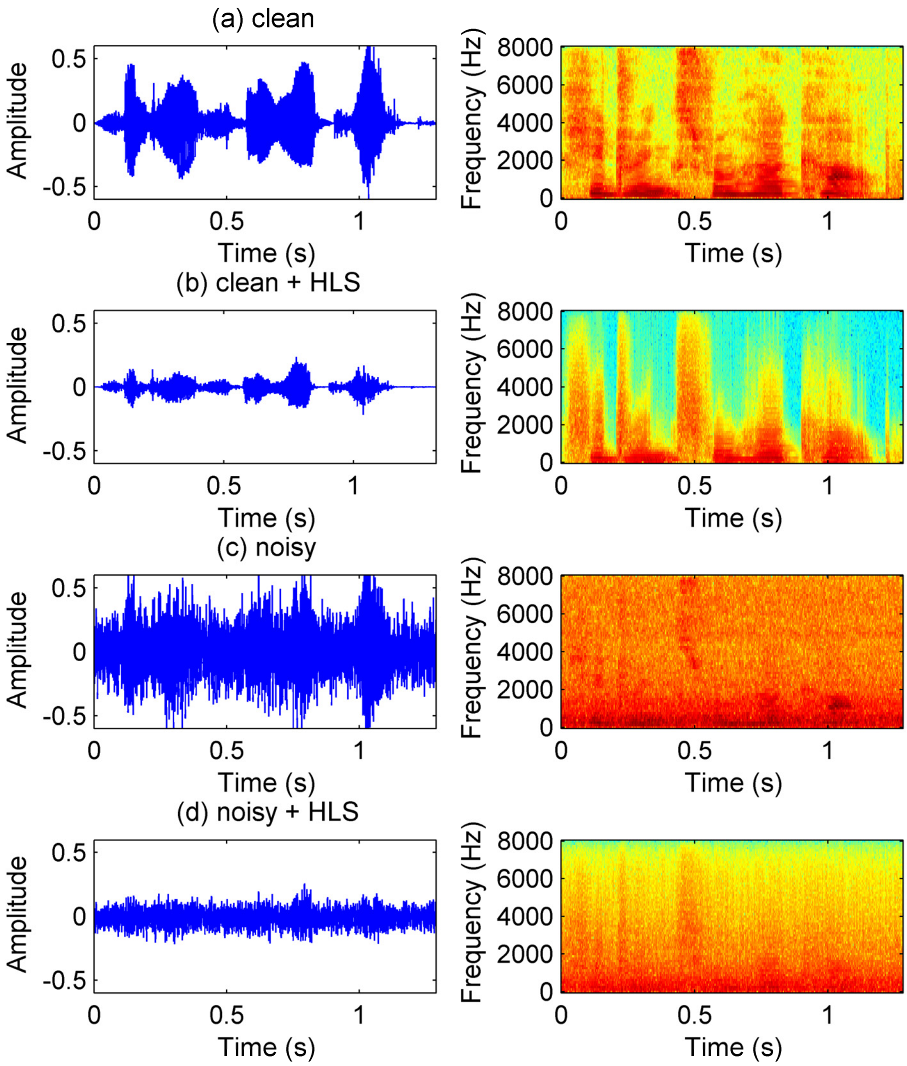


Figure 2.4: The demonstration of the speech in stationary noise with/without the HLS model. The BKB sentence is "*she drinks from her cup*". The waveforms and spectrograms are shown for (a) original clean speech; (b) the same clean speech processed with the hearing loss

simulation; (c) the same clean speech with speech shaped noise (0 dB input SNR); (d) the same noisy speech processed with the hearing loss simulation.

2.3 Discussion

This chapter proposed a novel HLS system by combining the physiologically motivated compressive gammachirp auditory filter bank with traditional methods to simulate both threshold and supra-threshold hearing loss factors. The purpose of the research was to build a versatile and computationally inexpensive HLS system that is simple to operate and maintain, yet has strong theoretical foundations and realistic simulation consequences. Figure 2.4 suggests that such a system can potentially simulate various perceptual aspects of sensorineural hearing loss, including reduced frequency selectivity, increased hearing thresholds and loudness recruitment. Furthermore, the compressive gammachirp-based HLS system can simulate the compressive, level dependent characteristic of the basilar membrane. The adoption of the gammachirp auditory filters into the HLS model is novel. This proposed HLS model has been validated in evaluating intelligibility of speech corrupted with speech shaped noise or babble noise.

When the noisy speech is processed with a noise reduction algorithm, perceptual distortions will be introduced to both speech and noise for an impaired ear which is different from perception for a normal ear. Whether the HLS model can appropriately quantify the impaired perceptual distortion brought by noise reduction algorithms needs validation experiments. Further work in Chapter 5 will be done to validate the HLS model in assessing the effects of noise reduction algorithms for HI listeners. The validation experiment in Chapter 5 will be performed by evaluating noise reduction algorithms with the HLS model and NH subjects. The result in Chapter 5 will be compared with the results with HI subjects in Chapter 4. If the intelligibility results from both experiments show similar performance, the HLS model can be proposed as realistic in predicting intelligibility effects for HI subjects.

Chapter 3 Development of sparse coding shrinkage algorithm

Hearing impaired people struggle more to understand speech that is corrupted with noise than normal hearing listeners. Current commercial hearing aids still have difficulty reducing noise and improving speech perception for hearing impaired (HI) listeners. This also demonstrates why hearing aid users often don't like to use hearing aids in noisy environments. In this chapter we have developed a single-channel sparse coding shrinkage noise reduction strategy for hearing aid (HA) users in noisy environments. This strategy takes into account superthreshold auditory deficits such as reduced frequency selectivity and, not only reduces background noise but also extracts key information from speech. The proposed strategy is compared with a competitive state-of-the-art noise reduction algorithm and unprocessed speech, through objective evaluations, in this chapter.

3.1 Introduction

Some hearing impairment compensation strategies have been well developed (e.g. automatic gain control), but noise reduction strategies need improving for use with hearing aids. Current hearing aids can help HI listeners in clean environments, but still have difficulty improving users' speech perception in noisy environments. This is due to a shortage of reliable and practical noise reduction algorithms for hearing aids. Because of the small size of hearing aids, multi-microphone noise reduction algorithms which can improve speech intelligibility are not suitable. Usually a hearing aid consists of only one or two microphones with single-channel noise reduction algorithms and beam forming algorithms together. The performance of single-channel noise reduction algorithms is critical in hearing aids.

Previous development of noise reduction strategies in hearing aids mainly focused on spectral subtraction (Levitt et al., 1993, Elberling et al., 1993, Alcantara et al., 2003, Dahlquist et al., 2005) or Wiener filtering algorithms (Levitt et al., 1993). Moreover, most noise reduction algorithms were originally developed to improve speech perception for normal hearing subjects and were later adopted for hearing aid users. Due to the hearing loss factors in

HI listeners, algorithms that are optimal for NH listeners might not be optimal for HI listeners.

When developing noise reduction algorithms for HI listeners, the hearing loss factors should be taken into account and compensated for when possible. As described in the previous chapters, for people with sensorineural hearing loss, hearing loss factors include threshold elevation, loudness recruitment, reduced frequency selectivity and reduced temporal resolution. Automatic gain control can compensate for threshold elevation and loudness recruitment, but there are currently no appropriate solutions to reduced frequency selectivity and reduced temporal resolution. Some researchers have attempted to compensate for reduced frequency selectivity with spectral sharpening but have not measured any intelligibility improvements (Baer et al., 1993). A possible solution to the speech in noise problem to compensate for reduced frequency selectivity is to extract and preserve key speech information while at the same time reducing the overall noise. This way, essential speech information will not be blurred or smeared by broadened auditory filters in HI listeners.

In this chapter, we propose a noise reduction strategy based on the principle of sparse coding shrinkage (SCS). It assumes a super-Gaussian (sparse) distribution of the principal components in clean speech and SCS is performed on the principal components. SCS was first proposed by (Hyvärinen, 1999), then applied to image noise reduction (Hyvärinen et al., 1998) and later applied to speech enhancement (Potamitis et al., 2001a, Sang et al., 2011a, Zou et al., 2008). Sparse coding has shown significant improvement in cochlear implant users (Li and Lutman, 2008) and this implies that there may be potential benefits of SCS in hearing aid users. One previous research study (Jesper and Richard, 2007) also assumed that the principal components of the speech are super-Gaussian but their calculation is complicated. The shrinkage function in our SCS is simplified by an approximation method.

This SCS algorithm is compared with a Wiener filtering approach, of which the code is provided by Timo Gerkmann. This algorithm was chosen as it is a competitive state-of-the-art noise reduction algorithm (Breithaupt et al., 2008, Gerkmann and Martin, 2009, Gerkmann and Hendriks, 2012). Wiener filtering approaches can reach optimal performance when the speech and noise are both in Gaussian distribution. However, previous research has shown that speech components are usually not in Gaussian distribution. As SCS has been developed to estimate the speech components with assumption of super-Gaussian distribution, we hypothesize that SCS might perform better than CS-WF especially for hearing impaired listeners who require less-but-key information from noisy speech.

This SCS was also compared with unprocessed speech which can show the baseline performance in a noisy environment without any algorithms applied. Previous research demonstrated that noise reduction algorithms might reduce speech intelligibility for HI listeners (Dahlquist et al., 2005). The comparison with unprocessed speech will be used to investigate whether there is any benefit of noise reduction algorithms for HI listeners. Babble noise and speech shaped noise were chosen as the additive noise, due to their similar average long term spectrum when compared with the speech signal. This makes the noise reduction performance challenging.

The perceptual consequences of noise reduction algorithms are demonstrated in ‘normal’ and ‘impaired’ spectrograms for NH listeners and HI listeners respectively. The ‘normal’ spectrograms are normal short-time Fourier Transform (STFT) domain speech amplitude representations. The ‘impaired’ spectrograms are speech with the hearing loss simulation (HLS) model added and then shown in spectrograms. If the HLS model is realistic to some degree, the ‘impaired’ spectrogram becomes an easy way to show the effects of signal processing strategies on HI listeners visually.

Previous research has shown that objective criteria such as segmental SNR or spectral distortion were uncorrelated with either speech quality or speech intelligibility (Ma et al., 2009). In this chapter, we will evaluate the noise reduction algorithms with several objective criteria (PESQ, fwsegSNR, CSII, NCM) which have been validated to reflect speech quality or intelligibility in normal hearing (NH) subjects (Hu and Loizou, 2008, Ma et al., 2009). We will also evaluate the noise reduction algorithms with an objective measure HASQI for hearing aids users (Kates and Arehart, 2010) to predict speech quality in HI listeners.

The structure of this chapter is as follows. Firstly, the concept and principle of sparse coding shrinkage is introduced. Secondly, the implementation of the sparse coding shrinkage principle for speech enhancement is proposed. Thirdly, we will introduce a competitive Wiener filtering algorithm which is going to be compared with SCS. Then, we will demonstrate the perceptual consequences of noise reduction algorithms for NH and HI listeners in ‘normal’ and ‘impaired’ spectrograms respectively. After this, comparison among the proposed SCS strategy, the competitive Wiener filtering algorithm and the noisy speech is performed through objective measures. Discussion and conclusions are also given to justify the effects of noise reduction algorithms with objective evaluations.

3.2 Introduction of sparse coding shrinkage

3.2.1 Gaussian distribution and sparse distribution

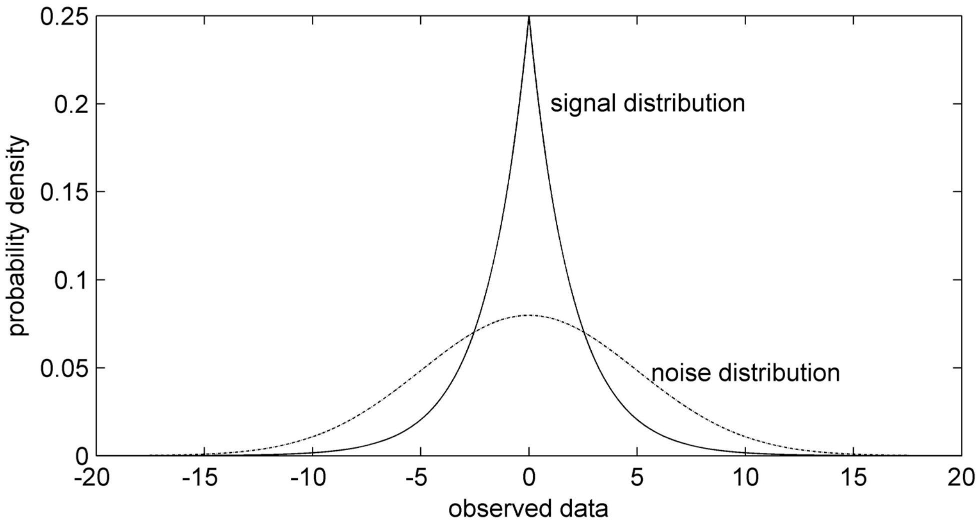


Figure 3.1: Examples of Gaussian distribution and sparse distribution centred on zero. Here the signal distribution is sparse (solid line) and the noise distribution is Gaussian (dotted line).

Figure 3.1 depicts examples of Gaussian distribution (dotted line) as a bell shape and sparse distribution (solid line) as a peaked shape. Suppose the clean signal is mixed with the noise and the clean signal needs to be estimated from the observed noisy signal. If the clean signal and the noise are in the same distribution, it is difficult to estimate the clean signal from the noisy signal. However, if the distribution of the signal is sparse (also called super-Gaussian) and the distribution of the noise is Gaussian as shown in Figure 3.1, estimation of the signal is possible due to the large deviation between their distributions. The sparse coding shrinkage principle can be used to estimate the clean signal in the latter situation. Derivation of the sparse coding shrinkage is presented in the following section. Generally, no matter which principle is applied to the noise reduction algorithms, it is always important to differentiate the characteristics between the signal and the noise.

3.2.2 **Principle of Sparse Coding Shrinkage**

This SCS principle is applied to estimate a random variable corrupted in Gaussian noise given sparse distribution of the random variable. Details are also described in (Hyvärinen, 1999). We will propose a noise reduction algorithm for speech enhancement based on the SCS principle in the next section.

We first consider only scalar random variables. s denotes the original non-Gaussian random variable and v the Gaussian noise with zero mean and variance σ^2. Assume that we observe only the random variable y:

$$y = s + v \tag{3.1}$$

The maximum a posterior estimator (MAP) is used to estimate the original variable s:

$$\hat{s} = \arg\max_s p(s|y) = \arg\max_s \frac{p(y|s)p(s)}{p(y)} \tag{3.2}$$

where $\hat{s}$ is the estimated original clean variable s and p denotes probability density.

Since $p(y)$ does not depend on s,

$$\hat{s} = \arg\max_s p_v(y-s)p_s(s) \tag{3.3}$$

where $p_v(y-s)$ is the density of noise evaluated at $y-s$ and $p_s(s)$ is the density of the original signal evaluated at s. As the noise is assumed to be white Gaussian noise, the distribution of noise is shown in the following equation,

$$p_v(y-s) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y-s}{\sigma}\right)^2\right) \tag{3.4}$$

Using $f_s = -\ln p_s$ the negative log density, then

$$\hat{s} = \arg\min_{s} \frac{1}{2\sigma^2}(y-s)^2 + f_s(s) \tag{3.5}$$

Assuming f_s to be strictly convex and differentiable, this equation could be solved using the following expression

$$\hat{s} = g(y) \tag{3.6}$$

which is called 'shrinkage function'. The shrinkage function used in our research is shown in Figure 3.2 with its specific expression in Equation (3.17). The effect of this function is to reduce the absolute value of its argument by a certain amount, which depends on the noise variance. Small arguments are suppressed to zero.

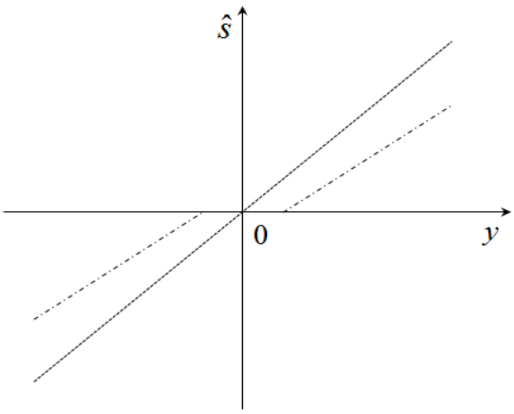


Figure 3.2: The shrinkage function used in our research (dash-dotted line). y is the observed signal and $\hat{s}$ is the estimated clean signal. The effect of this function is to reduce the absolute value of its argument by a certain amount, which depends on the noise variance. Small arguments are suppressed to zero (Hyvärinen, 1999).

3.3 Sparse Coding Shrinkage in Speech

3.3.1 **Implementation**

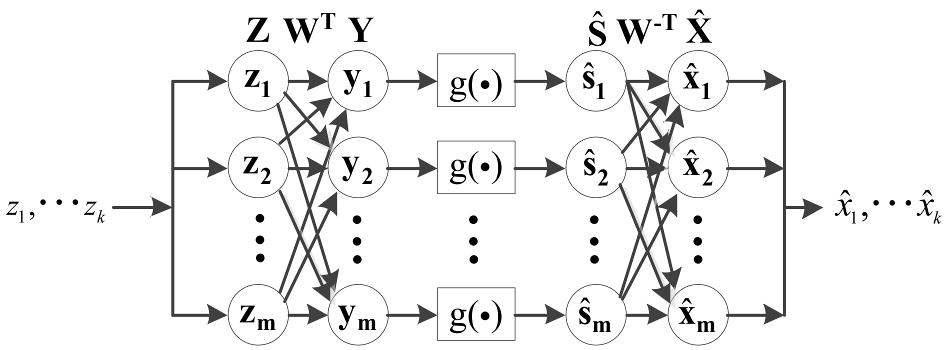


Figure 3.3: The flowchart of sparse coding shrinkage in noisy speech.

Figure 3.3 illustrates the flowchart of conducting the sparse coding shrinkage in noisy speech. This flowchart is implemented on each speech segment with length of k. The observed noisy speech is reconstructed into a noisy speech matrix $\mathbf{Z}$ as Equation (3.7). The noisy speech matrix was transformed into principal components where clean signals are transformed into a sparse distribution and noise is transformed into a Gaussian distribution. The shrinkage function g(·) is applied to suppress the noise in noisy components and estimate the clean components. After that, the inverse transform and reconstruction is calculated to derive the estimated clean speech signals. How to derive $\mathbf{W}$ will be described in Figure 3.4. $\mathbf{W}$ indicates the transpose of $\mathbf{W}^{\mathbf{T}}$ and $\mathbf{W}^{\mathbf{-T}}$ means the inverse of $\mathbf{W}^{\mathbf{T}}$.

The noisy speech signal z is assumed to be produced by corrupting the original speech sequence x with Gaussian noise n:

$$z = x + n \tag{3.7}$$

The noisy speech matrix is constructed by reshaping z as overlapping frames (50% overlap),

$$\mathbf{Z} = \begin{bmatrix} z_1 & z_{m/2+1} & \cdots & z_{m(l-1)/2+1} \\ z_2 & z_{m/2+2} & \cdots & z_{m(l-1)/2+2} \\ \vdots & \vdots & \cdots & \vdots \\ z_m & z_{m/2+m} & \cdots & z_{m(l-1)/2+m} \end{bmatrix} \tag{3.8}$$

where l (l=15) denotes the number of frames and m (m=64) is number of samples in each 4-ms frame at a sampling rate of 16 kHz. Accordingly, the total number of samples in each speech segment k equals $m(l-1)/2+m$ and the duration of each speech segment is 32 ms.

After reshaping, the original noisy speech can be written as

$$\mathbf{Z} = \mathbf{X} + \mathbf{N} \tag{3.9}$$

Noise is first estimated from noisy signal through a noise power estimation method (explained in Section 3.3.3). The estimated noise covariance matrix $\hat{\mathbf{R}}_{\mathbf{n}}$ and the estimated clean speech covariance matrix $\hat{\mathbf{R}}_{\mathbf{x}}$ are derived as noise and speech are assumed as uncorrelated. When the noise and speech both have zero means, the covariance matrices can be calculated through the following equations,

$$\begin{aligned} \mathbf{R_n} &= \mathbf{N} \times \mathbf{N^T} / m \\ \hat{\mathbf{R}}_\mathbf{x} &= \mathbf{X} \times \mathbf{X^T} / m \end{aligned} \tag{3.10}$$

Figure 3.4 shows how to derive eigenvalue matrix $\mathbf{\Lambda_x}$ and eigenvector matrix $\mathbf{W}$ through simultaneous diagonalization of the estimated clean speech and noise covariance matrices (Hu and Loizou, 2003).

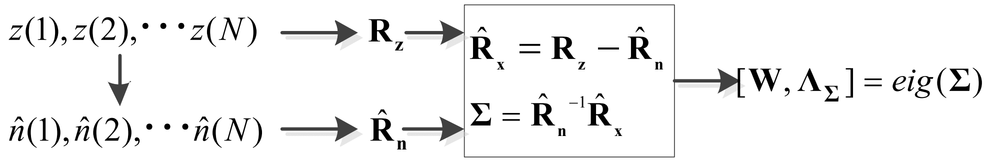


Figure 3.4: Flowchart of simultaneous diagonalization of the estimated speech and noise covariance matrices.

The transformation from noisy speech to principal components is realized with the eigenvector matrix **W** as illustrated in Figure 3.3.

Through the implementation in Figure 3.4, not only eigenvector matrix is derived but also noise is pre-whitened as illustrated in Equation (3.9)

$$\begin{aligned} \mathbf{W^T} \hat{\mathbf{R}}_\mathbf{\Sigma} \mathbf{W} &= \mathbf{\Lambda_\Sigma} \\ \mathbf{W^T} \hat{\mathbf{R}}_\mathbf{n} \mathbf{W} &= \mathbf{I} \end{aligned} \tag{3.11}$$

Transforming the noisy speech matrix to principal components is realized with the matrix **W** as follows,

$$\mathbf{Y} = \mathbf{W^T Z} = \mathbf{W^T X} + \mathbf{W^T N} = \mathbf{S} + \mathbf{V} \tag{3.12}$$

where $\mathbf{Y} = [\mathbf{y_1}; \mathbf{y_2}; \cdots \mathbf{y_m}]$, $\mathbf{S} = [\mathbf{s_1}; \mathbf{s_2}; \cdots \mathbf{s_m}]$, $\mathbf{V} = [\mathbf{v_1}; \mathbf{v_2}; \cdots \mathbf{v_m}]$, and the clean speech components $\mathbf{s_i}$ are in super-Gaussian distribution and noise components $\mathbf{v_i}$ are in Gaussian distribution. Therefore the sparse coding shrinkage function can be applied to each element y_i to estimate the clean element s_i :

$$\hat{s}_i = \mathrm{g}(\mathrm{y}_i) \tag{3.13}$$

where y_i is one element in the component $\mathbf{y_i}$ and s_i is one element in the component $\mathbf{s_i}$. $\hat{\mathbf{S}} = [\hat{\mathbf{s}}_\mathbf{1}; \hat{\mathbf{s}}_\mathbf{2}; \cdots \hat{\mathbf{s}}_\mathbf{m}]$ is the estimated clean speech matrix in the space of principal components. Inverse transformation of the estimated clean matrix yields

$$\hat{\mathbf{X}} = \mathbf{W}^{-\mathbf{T}}\hat{\mathbf{S}} \tag{3.14}$$

Finally, the enhanced speech $\hat{x}$ is reconstructed by reshaping $\hat{\mathbf{X}}$ back into a vector by the overlap and add method (Deller et al., 2000).

Although the above implementation is performed on a short speech segment (32 ms), the processing with longer speech can be realized by dividing it into short segments and using the overlap and add method (Deller et al., 2000).

3.3.2 **Super-Gaussian distribution and shrinkage function**

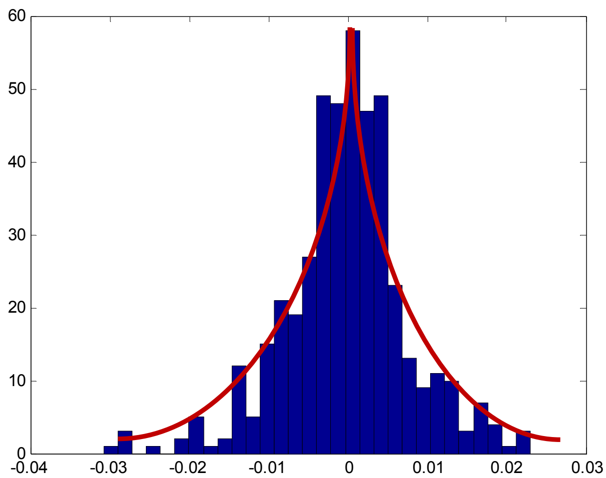


Figure 3.5: An example of a histogram showing the distribution of the coefficients of one principal component vector $\mathbf{s_i}$ in speech. The horizontal axis indicates the amplitude of the principal components. The vertical axis indicates the number of components in each amplitude bin.

Figure 3.5 shows the distribution of one principal component in speech. This shows an example of moderately super-Gaussian distribution. The red line was plotted to emphasize the peaked shape of the distribution which is different from the bell shape of Gaussian distribution.

Sparsity can also be quantified by kurtosis of the signal (Field, 1994). The more sparse, the larger the kurtosis is. The larger sparsity level is also shown as a more peaked distribution shape in Fig. 3.1. Sparsity of each component $\mathbf{s_i}$ in $\mathbf{S}$ can be estimated through normalized kurtosis:

$$K(\mathbf{s_i}) = \frac{1}{n}\sum_{j=1}^{n}\frac{(s_{ij}-\mu)^4}{\sigma^4} - 3 \tag{3.15}$$

where s_{ij} is the j^{th} observation value in $\mathbf{s_i}$, μ is the mean, σ is the standard deviation and K is the measured normalized kurtosis. The measured kurtosis is much larger than zero, the normalized kurtosis of a Gaussian distribution. This means that the distribution of each component $\mathbf{s_i}$ is super-Gaussian. Different super-Gaussian levels have been categorized as moderately super Gaussian, Laplacian and strongly super Gaussian (Hyvärinen, 1999). The distribution of clean speech components was selected as a linear combination of Gaussian and Laplacian distributions (Hyvärinen, 1999):

$$f_s(s_i) = C\exp\left(-as_i^{\,2}/2 - b|s_i|\right) \tag{3.16}$$

where C is an irrelevant scaling constant. Different values of a and b represent different degrees of super-Gaussianity.

Through maximum-a-posterior (MAP) derivation (Hyvärinen, 1999), the shrinkage function corresponding to the distribution of Equation (3.16) is derived as:

$$g(y_i) = \frac{1}{1+\sigma^2 a}\operatorname{sign}(y_i)\max\left(0, |y_i| - b\sigma^2\right) \tag{3.17}$$

where σ^2 is noise variance in each noise component v_i. The above shrinkage function is shown as the dash-dotted line in Figure 3.2.

This shrinkage function is interpolated between the shrinkage function of the Gaussian density and the shrinkage function of the Laplacian density. Specifically, when the distribution of s_i is

Laplacian, a is 0 and b is estimated as $\sqrt{2/E\{s_i^2\}}$; when the distribution of s_i is Gaussian, b is 0 and a is estimated as $1/E\{s_i^2\}$. Therefore it is reasonable to constrain the values of a and b in the intervals $\left[0, 1/E\{s_i^2\}\right]$ and $\left[0, \sqrt{2/E\{s_i^2\}}\right]$, respectively.

To simplify the estimation of a and b in Equation (3.17), we estimate that $a = \mu_1/E\{s_i^2\}, b = \mu_2 \times \sqrt{2/E\{s_i^2\}}$,

where $\mu_{1,2}$ are coefficients to be adjusted according to the distribution of s_i. In our test, μ_1 is set to 1, μ_2 is set to 0.3.

$E\{s_i^2\} = E\{y_i^2\} - \sigma^2$ the speech and noise are assumed to be uncorrelated.

The choice of moderately super Gaussian distributions is justified by the criterion that when $\sqrt{E\{s_i^2\}} f_s(0) < \frac{1}{\sqrt{2}}$, the distribution model can be assumed to be described as Equation (3.16) (Hyvärinen, 1999).

3.3.3 **Estimation of noise covariance matrix**

Generally, the noise covariance matrix needs to be estimated as $\hat{\mathbf{R}}_{\mathbf{n}}$ in Equation (3.11) and Figure 3.4 when a noisy speech signal is given. Here, a state-of-the-art noise estimator proposed by Gerkmann and Hendriks (2012) is adopted to track non-stationary noise. Instead of estimating noise power based on a voice activity detector, this method estimated noise power based on a soft speech presence probability (SPP) with a fixed *priori* SNR. The amplitude of the noise spectrum is therefore the square root of the noise power spectrum at each frequency bin frame by frame. The phase of the noise spectrum is assumed to be the same as that of noisy speech spectrum. The noise spectrum can be obtained by multiplying the amplitude of the noise spectrum with the phase of the noisy spectrum. The time-domain noise waveform is accordingly estimated with the inverse FFT of the estimated noise spectrum. The noise covariance matrix is estimated by calculating the covariance of noise in temporal domain as shown in Figure 3.4.

Another way to estimate the noise covariance matrix is estimated through direct inverse Fourier transform of the noise power spectral density according to Wiener-Khinchin Formula (Deller et al., 2000):

$$\begin{aligned} \Phi_{nn}(e^{jw}) &= \sum_{m=-\infty}^{+\infty} \phi_{nn}[m]e^{-jmw} \\ \phi_{nn}[m] &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \Phi_{nn}(e^{jw})e^{jwm}dw \end{aligned} \tag{3.18}$$

Where $\Phi_{nn}(e^{jw})$ is NPSD and $\phi_{nn}[m]$ is noise auto-correlation coefficient which can be derived through inverse Fourier transform of NPSD. When the mean of noise is zero, noise auto-covariance coefficients equal noise auto-correlation coefficients. If the noise covariance matrix is in the length of M, it could be constructed as a symmetric Toeplitz matrix with the first M values of noise auto-covariance coefficients.

3.3.4 **Introduction of comparison algorithm**

This SCS algorithm was compared with a Wiener filtering approach, of which the code was provided by Timo Gerkmann (Breithaupt et al., 2008, Gerkmann and Martin, 2009, Gerkmann and Hendriks, 2012). The idea of Wiener filtering has been introduced in Section 1.5.2. One characteristic of this Wiener filtering approach is that the *a priori* SNR is estimated by the cepstral smoothing method. We refer to this approach as 'CS-WF' herein. This algorithm was chosen because Wiener filters are used frequently in today's hearing aids and CS-WF is a competitive, state-of-the-art, algorithm (Breithaupt et al., 2008, Gerkmann and Martin, 2009, Gerkmann and Hendriks, 2012). There were two critical techniques in CS-WF. One technique was to estimate the noise power spectral density (NPSD) based on a speech presence probability (SPP), where the *a priori* SNR was a fixed value in estimating the SPP (Gerkmann and Hendriks, 2012). SCS also adopted the same NPSD estimation method as described in section 3.3.3. The other technique involved estimating the *a priori* SNR using temporal cepstrum smoothing with bias compensation (Breithaupt et al., 2008, Gerkmann and Martin, 2009). This algorithm can reduce musical noise and suppress non-stationary noise effectively.

3.4 Objective evaluation

The SCS, CS-WF and the noisy speech (baseline condition) were evaluated through objective criteria. Although the current objective measures are not as reliable as the subjective tests especially for HI subjects, it is quick and cheap to test algorithms with objective measures during the strategy development stage. Subjective evaluations are presented in the next two chapters. Bamford-Kowal-Bench (BKB) (Bench et al., 1979) sentences recorded by a female British speaker are used as speech material. The sentence database comprised 21 lists with 16 sentences in each list and 3 or 4 keywords in each sentence. Speech shaped noise (SSN) and babble noise are used as additive background noise.

3.4.1 **Demonstration through 'normal' and 'impaired' spectrograms**

In this section, we demonstrate the noise reduction effects for NH and HI listeners with speech temporal waveforms and spectrograms. Figure 3.6 and 3.7 show speech in speech shaped noise or babble noise (0 dB input SNR) with and without noise reduction algorithms for NH listeners. Supposing that the hearing impaired listening can be approximated with the additive combination effects of hearing loss distortion and normal hearing listening, stimuli processed with a realistic hearing loss simulation can mimic the perceptual consequence of hearing loss. The spectrograms of speech with the HLS model developed and described in Chapter 2 are named 'impaired spectrograms'. The HLS model assumed the same hearing thresholds as shown in Figure 2.2. Figure 3.8 and 3.9 show speech in speech shaped noise or babble noise (0 dB input SNR) with or without noise reduction algorithms for HI listeners in 'impaired' waveforms and spectrograms. It is of note that all the four spectrograms in each figure share the same colour scale to show the amplitude differences among them.

Figure 3.6 and Figure 3.7 show the 'normal' time domain waveforms and spectrograms of an example BKB sentence "*she drinks from her cup*" under 0 dB input SNR with different noise reduction strategies in the background of SSN and babble noise respectively. Clean speech and noisy speech are also shown. The waveforms and the spectrograms are (a) original speech; (b) noisy speech; (c) noisy speech processed with CS-WF; (d) noisy speech processed with SCS. By comparison through (b) to (d), the noise reduction algorithms show more noise reduction effects in speech shaped noise but less in babble noise. Babble noise is a competitive multi-talker noise and is difficult to reduce due to its non-stationary property. SCS reduced the speech shaped noise most efficiently (shown in Figure 3.6) and CS-WF reduced babble noise most efficiently (shown in Figure 3.7).

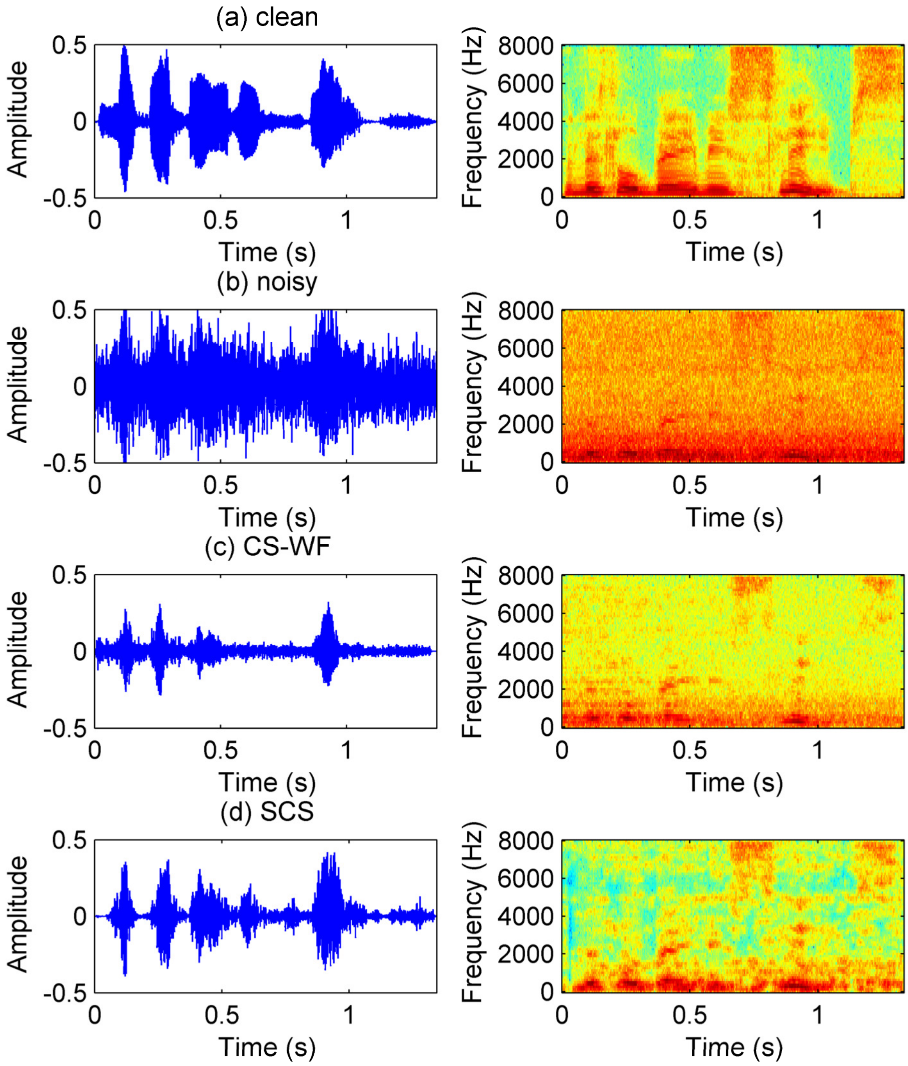


Figure 3.6: The demonstration of the speech in stationary noise with/without noise reduction algorithms. The BKB sentence is “*Little baby sleeps*”. The spectrograms and waveforms are shown for (a) original speech; (b) speech in speech shaped noise (0 dB input SNR); (c) the same noisy speech processed with CS-WF; (d) the noisy speech processed with SCS.

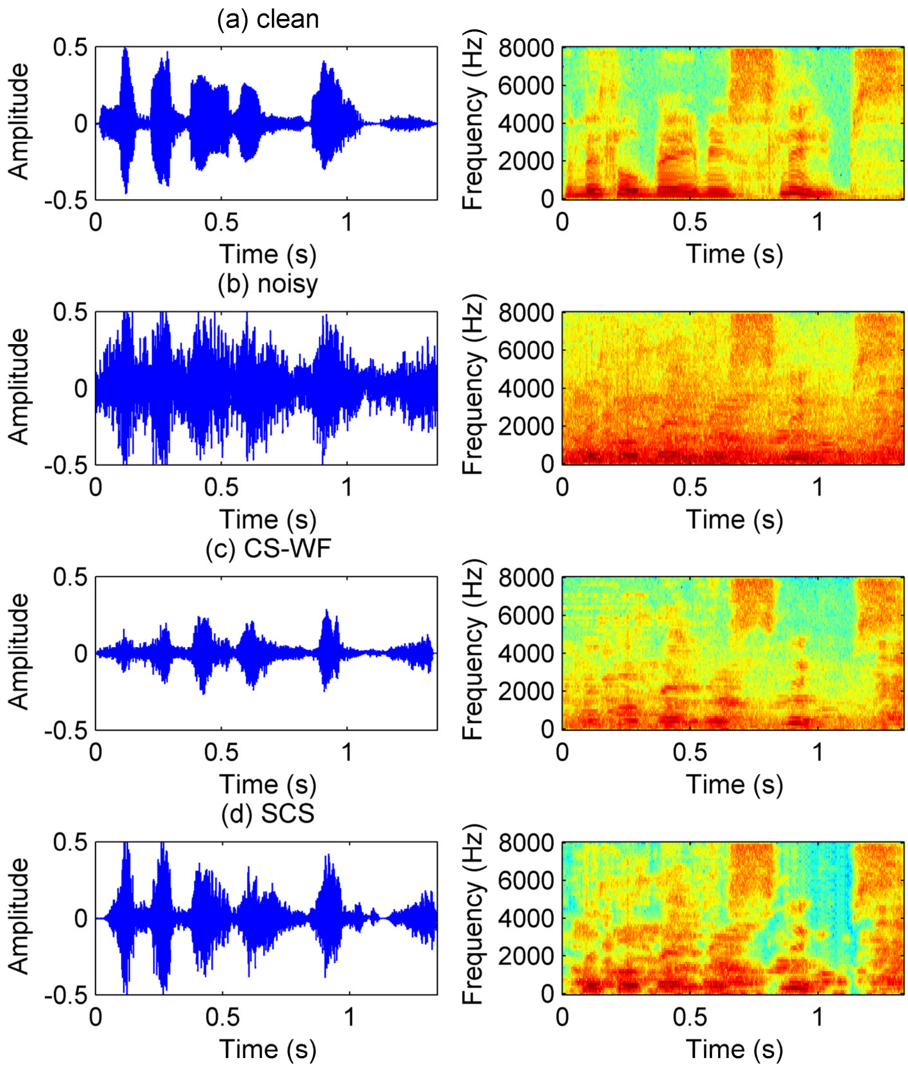


Figure 3.7: The demonstration of the speech in babble noise with/without noise reduction algorithms. The BKB sentence is "*Little baby sleeps*". The spectrograms and waveforms are shown for (a) original speech; (b) speech in babble noise (0 dB input SNR); (c) the same noisy speech processed with CS-WF; (d) the noisy speech processed with SCS.

Figure 3.8 and Figure 3.9 show the 'impaired' time domain waveforms and spectrograms of an example BKB sentence "*she drinks from her cup*" under 0 dB input SNR (in the background of SSN and babble noise respectively) with different noise reduction strategies. The 'impaired' clean speech and noisy speech are also shown. The waveforms and the spectrograms shown are (a) original speech with the HLS model; (b) noisy speech with the HLS model; (c) the same noisy speech processed with CS-WF plus the HLS model; (d) the same noisy speech processed with SCS plus the HLS model. By visual comparison and informal listening among (b) to (d) in each figure, noise has detrimental effects in the hearing loss simulation model. The 'impaired' spectrograms in Figure 3.8 show that the noise reduction effects increase in the order of noisy, CS-WF and SCS in speech shaped noise. This indicates the potential of noise reduction algorithms to reduce noise in the HLS model in speech shaped noise. The 'impaired' spectrograms in Figure 3.9 show that the noise reduction algorithms only slightly reduce noise. This indicates the difficulty of noise reduction algorithms to reduce babble noise in the HLS model.

Through informal listening of the 'impaired' speech (Figure 3.8 and Figure 3.9), there is no difference in speech quality ('comfort', 'clarity') level among (b) to (d); however, the difference in noise loudness among (b) to (d) can be perceived especially in speech shaped noise. In contrast, through the informal listening of the 'normal' speech (Figure 3.6 and Figure 3.7), the difference in both speech quality and noise loudness can be detected among (b) to (d). This suggests that 'impaired' speech is less sensitive to speech distortion by noise reduction algorithms than 'normal' speech.

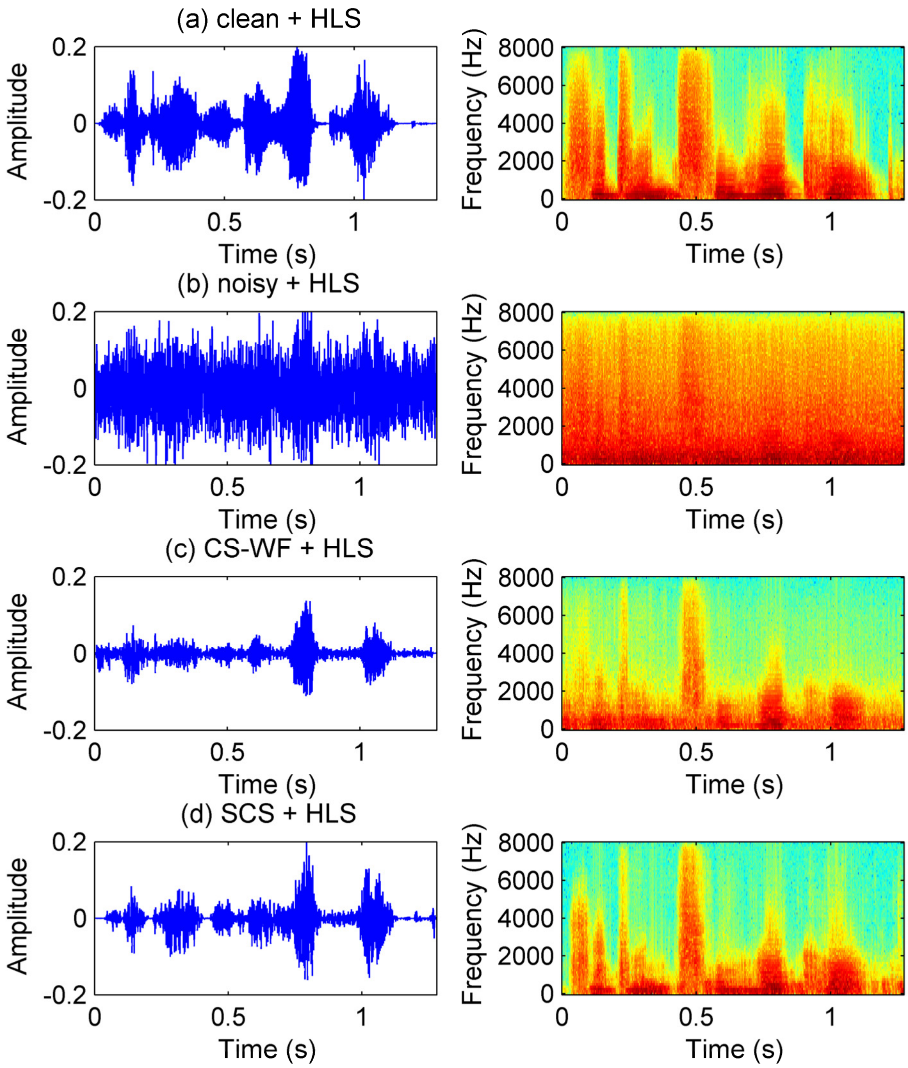


Figure 3.8: The demonstration of the unprocessed/enhanced speech (speech in stationary noise) with the HLS model. The BKB sentence is "*she drinks from her cup"*. The waveforms and spectrograms are shown for (a) original speech with the hearing loss simulation; (b) speech in speech shaped noise (0 dB input SNR) with the hearing loss simulation; (c) the same noisy speech processed with CS-WF and then with the hearing loss simulation; (d) the same noisy speech processed with SCS and then with the hearing loss simulation.

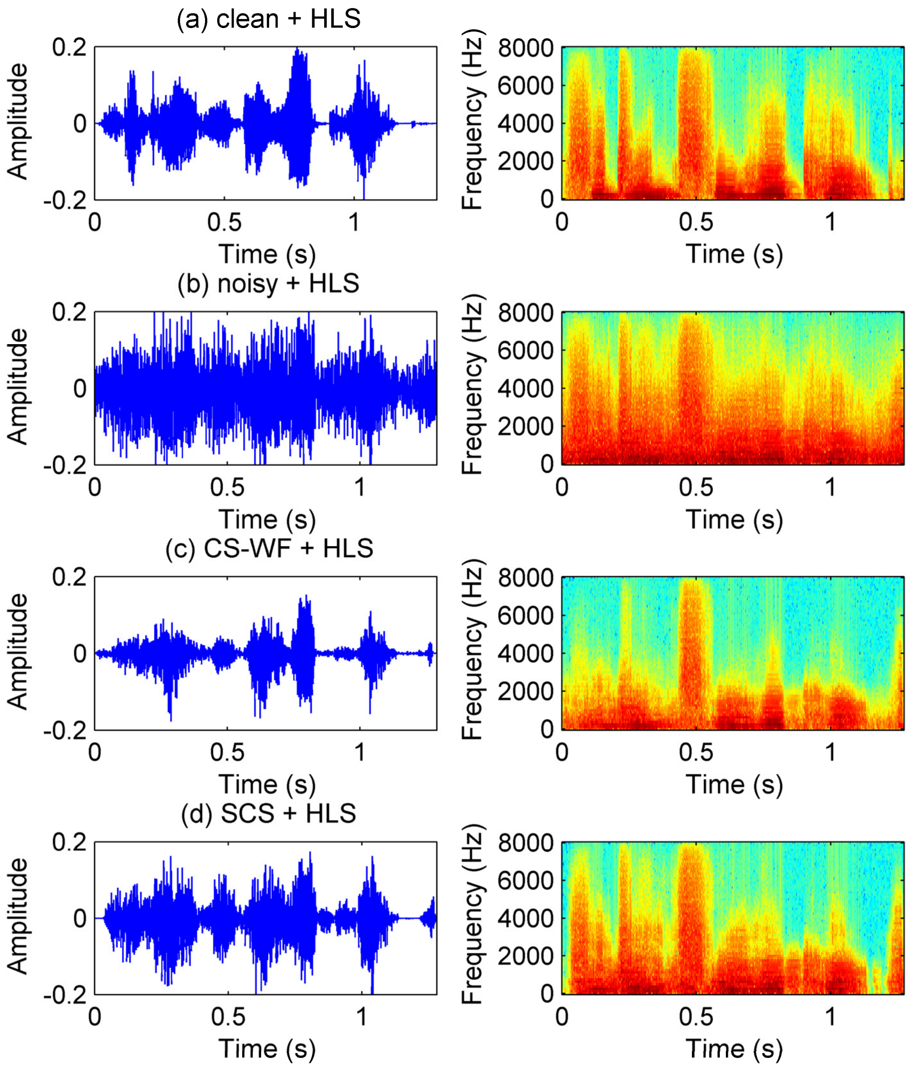


Figure 3.9: The demonstration of the unprocessed/enhanced speech (speech in babble noise) with the HLS model. The BKB sentence is "*she drinks from her cup*". The waveforms and spectrograms are shown for (a) original speech with the hearing loss simulation; (b) speech in babble noise (0 dB input SNR) with the hearing loss simulation; (c) the same noisy speech processed with CS-WF and then with the hearing loss simulation; (d) the same noisy speech processed with SCS and then with the hearing loss simulation.

3.4.2 **Objective quality and intelligibility evaluation**

This section introduces the results of objective measures that were explained in detail in Section 1.6.3, including fwsegSNR, PESQ, HASQI, NCM, STOI. FwsegSNR reflects the physical noise reduction effects in NH listeners. All of the measures were calculated based on 336 BKB sentences (21 lists with 16 sentences in each list). The results presented in Figure 3.10 were the averages from the 336 sentences in each condition. While mixing the speech with noise for different SNRs, the level of speech was kept constant and the level of noise was varied. Another advantage of the five objective measures is that their results don't vary with the amplitude of speech as long as the amplitude of speech is not over clipped. That is, when the reference speech (clean speech) and processed speech are varied with the same magnitude, the result of the objective measure is kept the same.

The most widely used objective quality metric 'segmental SNR' (segSNR) is not used here as it has been proved to be uncorrelated with either subjective quality or intelligibility performance (Yi and Loizou, 2008, Ma et al., 2009). PESQ reflects sound quality and has been recommended by ITU-T to assess speech quality of narrow-band handset telephony and narrow-band speech codec. PESQ has also been validated to reflect subjective quality impression for NH listeners. NCM and STOI are two validated objective intelligibility measures of noise reduction algorithms for NH listeners. The objective measures that have been validated with NH listeners might not reflect HI listening effects, as large deviations occurred between the objective measures for NH listeners and the performance with HI subjects (Pavlovic, 1984).

Although some effort has been made to develop objective measures for HI listeners, their accuracy in predicting effects of noise reduction algorithms for HI listeners are still uncertain. HASQI is a speech quality metric for HI listeners and it incorporates hearing thresholds as input parameters. We assumed the same moderate hearing loss level (shown in Figure 2.2) in HASQI.

Figure 3.10 (a) shows the fwsegSNR results of noisy, CS-WF and SCS in speech shaped noise (left half panel) and babble noise (right half panel) under 0, 5, 10 dB input SNRs. SCS and CS-WF show the most physical noise reduction effects in speech shaped noise and babble noise respectively. This also is in accordance with the spectrograms shown in Fig. 3-4 and Fig. 3-5 that SCS reduces speech shaped noise the most and CS-WF reduces babble noise the most.

Figure 3.10 (b) shows the PESQ results of noisy, CS-WF and SCS in speech shaped noise (left half panel) and babble noise (right half panel) under 0, 5, 10 dB input SNRs. CS-WF shows the best speech quality in this metric in both speech shaped noise and babble noise. This indicates that although SCS achieved greater noise reduction, speech distortion might be introduced at the same time which is reflected by PESQ.

Figure 3.10 (c) and (d) respectively show the NCM and STOI results of noisy, CS-WF and SCS in speech shaped noise (left half panel) and babble noise (right half panel) under 0, 5, 10 dB input SNRs. Both of these are objective intelligibility measures and have been validated with NH listeners. These results show very slight differences among noisy, CS-WF and SCS in each condition.

Figure 3.10 (e) shows the HASQI results of noisy, CS-WF and SCS in speech shaped noise (left half panel) and babble noise (right half panel) under 0, 5, 10 dB input SNRs. This metric has incorporated the parameters of hearing thresholds with the moderate hearing loss and is expected to reflect speech quality for HI listeners. SCS and CS-WF show the best increase in quality for speech shaped noise and babble noise respectively in this metric. The results of HASQI are similar to fwsegSNR where SCS shows the best performance in stationary noise and CS-WF show the best performance in babble noise. Furthermore, the benefits of noise reduction algorithms in HASQI are relatively larger than in fwsegSNR. HASQI and fwsegSNR reflect effects on quality for HI listeners and NH listeners respectively. This indicates noise reduction algorithms might benefit HI listeners more than NH listeners in speech quality.

It is always difficult to set up a reliable objective intelligibility measure for HI listeners especially when evaluating comparative intelligibility effects between different signal processing strategies. Although several objective intelligibility measures (Hollube and

Kollmeier, 1996) have been developed for HI listeners, their validation with effects of noise reduction algorithms on HI listeners has not yet been proven. Moreover, there are seldom publicly available objective measures for HI listeners. Therefore, no objective intelligibility measure for HI subjects is adopted in this section. Alternatively, we have proposed a methodology to predict intelligibility effects for HI listeners with a hearing loss simulation model and normal listening in Chapter 5.

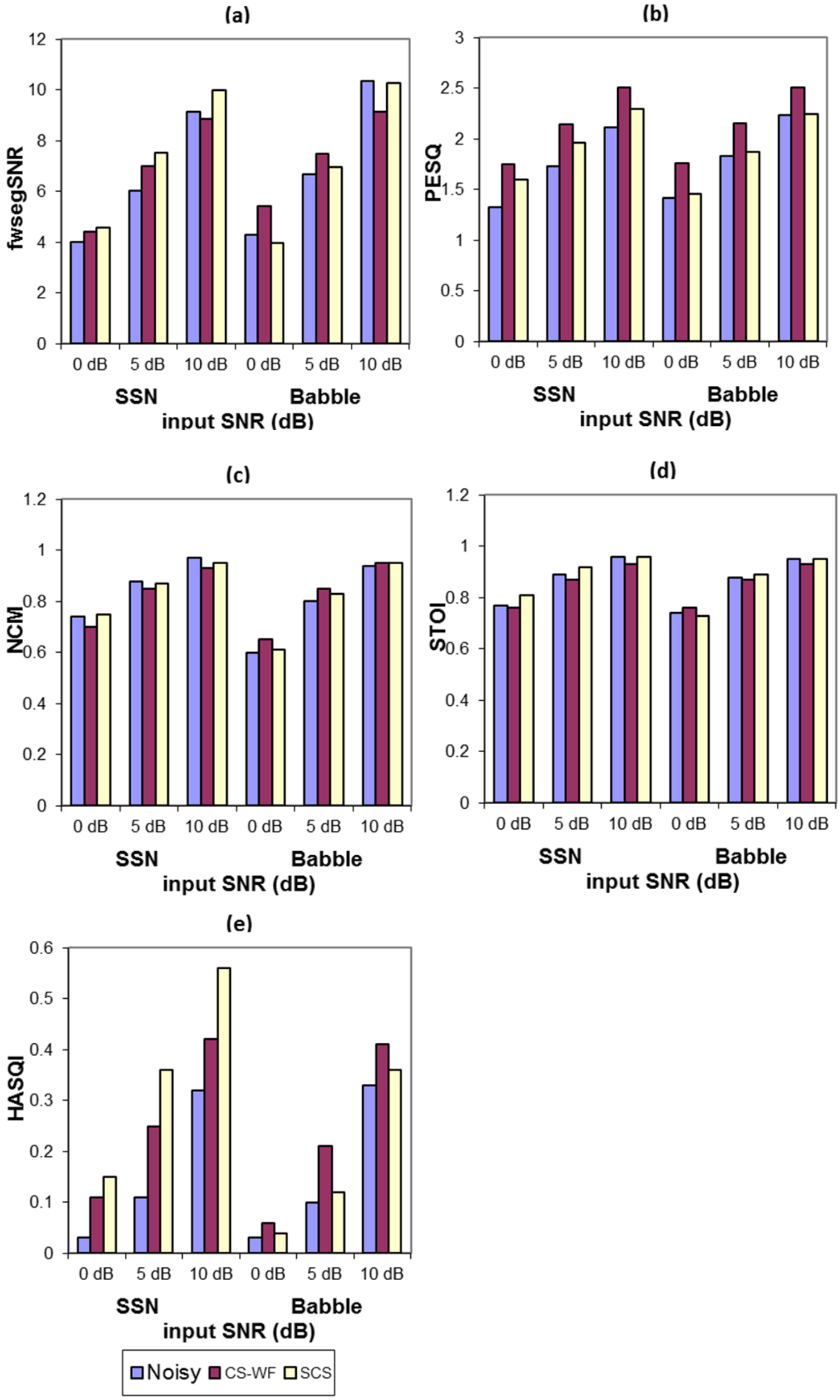


Figure 3.10: Results of objective measures of fwsegSNR, PESQ, NCM, STOI and HASQI. FwsegSNR and PESQ are quality measures for NH listeners; NCM and STOI are intelligibility

measures for NH listeners; HASQI is a quality measure for HI listeners. A more positive value corresponds to better performance in each measure.

3.5 Discussion

3.5.1 **Potential of sparse coding strategies in speech enhancement**

The sparse coding shrinkage principle assumes a super-Gaussian distribution of speech representation coefficients and Gaussian distribution of noise representation coefficients and then derives a gain function (shrinkage function) which is different from Wiener filtering. Sparse coding shrinkage is not limited to principal components (PCA) or independent components (ICA), but could be applied to different speech representation domains. For example, in the STFT-based speech enhancement methods, the traditional Gaussian STFT coefficient assumption (Ephraim and Malah, 1984) has been replaced by Laplacian or gamma assumptions (Martin, 2005) or a generalized super-Gaussian assumption (Lotter and Vary, 2005). Sparse coding shrinkage has also been applied in the spectral envelope domain to realize speech enhancement in CI vocoded speech (Li, 2008, Li and Lutman, 2008, Hu et al., 2011a, Sang et al., 2011a).

The proposed SCS strategy is different from previous SCS strategies in speech enhancement (Potamitis et al., 2001b, Zou et al., 2008, Sang et al., 2011a, Potamitis et al., 2001a) which usually need a *priori* knowledge of statistics of speech and are therefore not robust in different speech environments. The proposed SCS performed shrinkage function in the domain of principal components with approximate shrinkage parameters without a *priori* knowledge.

Some advanced sparse coding strategies have already been applied to speech enhancement besides the sparse coding shrinkage principle. For example, Sigg applied K-SVD (Aharon et al., 2006) in speech enhancement (Sigg et al., 2010). For Gaussian noise, they derive a dictionary from training speech data; for structured noise (nongaussian), they dictionaries from training speech and training noise separately. Although most of the state-of-the-art sparse coding strategies have the potential to apply to speech enhancement strategies (Aharon et al., 2006, Zhou et al., 2009, Zhou et al., 2011, Mairal et al., 2010), it is critical to apply sparse coding strategies to blind environments without *a priori* knowledge of either speech or noise but with robust performance. One noticeable work by Jafari and Plumbley (2011) is one of the very few sparse coding dictionary learning algorithms that aim to decompose speech signals and reduce noise. However, they only evaluated the noise reduction effects with one objective measure. Subjective tests are necessary to further justify such algorithms.

3.5.2 **Validity of objective measures**

The objective measures in this chapter include two quality measures (fwsegSNR, PESQ) for NH listeners, two intelligibility measures (NCM, STOI) for NH listeners and one quality measure (HASQI) for HI listeners. The comparisons between these objective measures and subjective tests (in Chapter 4) can be performed to validate the effectiveness of objective measures. No objective intelligibility measure for HI listeners is used in our thesis due to the limited availability and reliability of objective intelligibility measures for HI subjects, in addition to the distortion with noise reduction algorithms. Alternatively, we will propose a methodology in Chapter 5 to predict speech intelligibility of noise reduction algorithms for HI listeners with a hearing loss simulation model using NH subjects. In this chapter, we also demonstrated the perceptual consequences of noise reduction algorithms in HI listeners through impaired spectrograms. The 'impaired' spectrograms included the HLS model to illustrate the effects of signal processing strategies on HI subjects visually. If the HLS model is realistic, the demonstration with impaired spectrograms can be a reliable method to assess strategies before subjective tests.

3.5.3 **Comparison between SCS and CS-WF**

The motivation to choose CS-WF as a competitor algorithm instead of choosing the same PCA based method with another shrinkage function is to assess whether the PCA-based sparse coding shrinkage is competitive compared to one of the state-of-the-art algorithms (CS-WF).

This can present an up-to-date knowledge of the performance of SCS. Through the comparison between SCS and CS-WF using objective measures, it was found that SCS shows more noise reduction effects in speech shaped noise and CS-WF shows more noise reduction effects in babble noise. This may be due to SCS reaching optimal performance in the presence of speech with white Gaussian noise and performing worse when the noise has the similar distribution statistic as the speech.

3.6 Conclusions

This chapter proposed a noise reduction algorithm based on the sparse coding shrinkage principle. This is a model-based method with principal component analysis. The sparse strategy is expected to keep less but key information of noisy speech.

This SCS algorithm was compared with a competitive state-of-the-art Wiener filtering approach (CS-WF) and the noisy speech through objective measures in this chapter. Final evaluation of noise reduction algorithms will be decided by the subjective tests in Chapter 4. Objective measures can be compared with subjective tests to justify the reliability of objective measure.

The objective measures in this chapter suggested that:

- The two noise reduction algorithms may bring benefits in speech quality but not in speech intelligibility to NH listeners.
- The two noise reduction algorithms may bring more benefits in speech quality to HI listeners (HASQI) than to NH listeners (fwsegSNR).
- It is uncertain whether the two noise reduction algorithms can improve speech intelligibility for HI listeners given the shortage of relevant intelligibility measures; SCS might improve speech quality more in speech shaped noise and CS-WF might improve speech quality more in babble noise.

Chapter 4 Subjective evaluation of noise reduction algorithms in normal hearing and hearing impaired listeners

This chapter presents a subjective experimental evaluation of noise reduction algorithms in speech intelligibility and quality through normal hearing (NH) and hearing impaired (HI) listeners. Speech intelligibility is evaluated using an adaptive speech recognition test with the result of a speech reception threshold (SRT, in dB) which corresponds to the speech-to-noise ratio required for 79.4% correct recognition. Speech quality is evaluated through a method called Interpolated Paired Comparison Rating (IPCR) with the result of a speech-to-noise ratio gain (SNR-gain in dB). The SNR gain in quality is the difference between the SNRs of the noisy speech and the processed speech that reach the same quality impression. IPCR can sensitively detect speech quality improvement quantitatively and can be compared with SRT or objective SNR gain on the same scale. Speech under two different noise types (babble, speech shaped noise) with and without noise reduction algorithms has been evaluated using NH and HI listeners.

4.1 Introduction

Previous research has shown that most single-channel noise reduction algorithms did not improve speech intelligibility for NH listeners (Hu and Loizou, 2007). One exception is a Bayesian classification algorithm which can optimize noise reduction algorithm with *a priori* knowledge of acoustic environment and thus improve speech intelligibility for NH listeners (Kim and Loizou, 2010). However, the algorithms with *a priori* knowledge of statistics in speech or noise are neither robust nor practical in blind acoustic environments. Generally, most of the practical single-channel noise reduction algorithms cannot improve speech intelligibility for NH listeners. Despite this, single-channel noise reduction strategies have shown significant improvement in speech intelligibility in cochlear implant (CI) users (profoundly HI listeners) (Verschuur et al., 2006, Li, 2008, Li and Lutman, 2008, Yang and Fu, 2005, Hu et al., 2011a, Dawson et al., 2011). The contrast in intelligibility benefits between NH listeners and CI users may reflect subjective performance variation with the hearing loss level. Some may speculate the intelligibility improvement from noise reduction algorithms increases with the hearing loss level. However, the effects of single-channel noise reduction algorithms on intelligibility for hearing aid users are still not clear. Previous research has shown varied intelligibility performance of noise reduction strategies in hearing aid users (Levitt et al., 1993, Elberling et al., 1993, Aharon et al., 2006, Dahlquist et al., 2005). This is consistent as hypothetically listeners with mild to severe hearing losses acquire intelligibility benefits from noise reduction algorithms somewhere in between listeners with normal hearing and profound hearing losses. The variations of noise reduction effects in hearing aid users across previous evaluations are due to various factors, e.g. the hearing loss level as hearing aid users span a large range of hearing loss levels (mild, moderate and severe), the hearing aid experience, the noise reduction algorithm, the noise type, the input SNR, etc. Elberling *et al.* (1993) evaluated three spectral subtraction algorithms and indicated that the algorithms decreased noise level but did not improve speech intelligibility in either NH listeners or HI listeners. Levitt *et al.* (1993) assessed four noise reduction algorithms and showed significant improvements for some HI listeners with single-channel short-term Wiener filtering algorithms. Dahlquist *et al.* (2005) tested a nonlinear spectral subtraction algorithm in four different noise types but found no speech intelligibility improvement in HI listeners. Aharon *et al.* (2006) evaluated an auditory masked threshold noise suppression algorithm in both NH and HI listeners and found significant intelligibility improvement in HI listeners in background of communication channel noise and highway noise. Most recently, Harlander *et al.* (2012) evaluated model-based versus non-parametric monaural noise reduction approaches with HI listeners. They found that none of the algorithms improved speech intelligibility, although the two model-based noise reduction algorithms improved speech quality (Harlander et al., 2012).

Our study differed from previous studies in several ways. First, young subjects who have acquired hearing losses and were wearing hearing aids at an early age are recuited in our study as compared to the older subjects recruited in the investigaion (Dahlquist et al., 2005). Second, our evaluation assessed the sparse coding shrinkage algorithm as well as a competitive Wiener filtering algorithm as compared to previous noise reduction evaluations in hearing aids that were mostly limited to Wiener filtering or spectral subtraction algorithms; our evaluation uses speech shaped noise and babble noise as additive noises which approach closer to realistic noise environments.

In most of the previous subjective intelligibility tests, speech intelligibility is often quantified in terms of percentage of identified words (or syllables) correct. Percentage intelligibility is often measured at fixed input SNRs. Such intelligibility measures are inherently limited by floor or ceiling effects. A more reliable measure of speech intelligibility that is not sensitive to the input SNR is recommended as an adaptive procedure to measure the speech reception threshold (SRT), which is the signal-to-noise ratio at which the participant scored a fixed percentage correct (Dirks *et al.*, 1982). The adaptive method called the up-down procedure has been evaluated to be accurate and efficient (Dirks et al., 1982, Levitt, 1971). Our evaluation follows the three-up-one-down procedure in speech recognition tests to measure the SRT corresponding to 79.4% correct recognition threshold. This is more stringent than other research which uses the one-up-one-down procedure to identify 50% correct recognition threshold.

Although there is an unclear intelligibility improvement when using noise reduction algorithms for hearing aid users, speech quality has been reported to improve with noise reduction strategies for both NH and HI subjects. It is still of interest to evaluate whether the two noise reduction algorithms (CS-WF and SCS) can improve speech quality for NH or HI subjects and whether there is a difference in sound quality impression of noise reduction algorithms between NH and HI subjects.

The most commonly reported subjective techniques of quality evaluation can be broadly classified into two categories: those based on relative preference tasks (Munson and Karlin, 1962, Hecker and Williams, 1966, Combescure et al., 1982) and those based on assigning a numerical value to the quality of the speech stimuli (Marzinzik, 2000, Boymans et al., 1999, Jamieson et al., 1995). In relative preference tests, listeners are presented with a pair of speech stimuli consisting of the test stimulus and the reference stimulus. Listeners are asked which stimuli they prefer. In the rating tests, such as the MOS test used in telephone applications, listeners are asked to rate the quality of the stimuli on a numerical scale, typically, a five-point scale with 1 indicating poor quality and 5 indicating excellent quality. As individual listeners have different perceptual criteria of rating quality, there will be a large variance in rating tests.

A method called Interpolated Paired Comparison Rating (IPCR) (Dahlquist et al., 2005) has been developed to measure subjective SNR gain in sound quality, which is the difference in signal-to-noise ratios between processed and unprocessed speech that give equal subjective sound quality impression. One direct way to measure such SNR-gain is to compare adaptively between enhanced stimuli with a fixed SNR and unprocessed item with a variable SNR. However, it is too time-consuming (Dahlquist et al., 2005). IPCR is performed by comparison between processed stimuli with fixed SNR and unprocessed stimuli with only two different SNRs. A rating value between -10 and 10 is given as subjective impression feedback in each comparison between enhanced stimuli and unprocessed stimuli. Interpolation or extrapolation was used to find the SNR-gain of subjective equality of quality impression. The advantage of IPCR is that it can evaluate speech quality quantitatively and efficiently. This evaluation method IPCR was developed in LISCOM, an EU project within the Telematics Application Programme and later validated by tests with 30 hearing-impaired listeners each at two laboratories: KTH in Sweden and ISVR in the UK (Dahlquist et al., 2005). Sound quality dimensions include ‘preference’, ‘comfort’, ‘speech clarity’ and ‘background noise’. These four dimensions were evaluated by Dahlquist *et al.* (2005). Due to the time restrictions of evaluating the two algorithms (SCS, WF) with two noise types (babble noise, speech shaped noise) through two groups of listeners (NH and HI), only two subjective quality impression dimensions (‘preference’ and ‘background noise’) are tested in our experiment. ‘Preference’ is rated according to the overall impression of speech quality, including ‘comfort’, ‘speech clarity’ and ‘noise loudness’, etc. ‘Noise loudness’ is rated according to how much they can perceive the background noise, a more positive rating indicates less ‘noise loudness’ in our evaluation.

In this chapter, several research questions are raised: Is there any difference in noise reduction effects between NH and HI listeners? Is there any difference in performance between CS-WF and SCS? Is there any difference in performance under different noises? Can noise reduction strategies benefit HI subjects in speech quality or in speech intelligibility, or both?

4.2 Stimuli

The speech materials were BKB sentences (Bench et al., 1979) which are standard British sentences recorded by a female talker. The corpus consists of 21 lists with 16 sentences in each list. There are three or four key words in each sentence. Two noise types are used, babble noise and speech shaped noise. The speech shaped noise is generated such that the frequency spectrum matched the long term average spectrum of speech being used. The babble noise is a multi-talker babble noise. The speech corrupted with noise is further processed with and without noise reduction strategies to produce stimuli that will be denoted as ‘noisy’, ’CS-WF’ and ‘SCS’ in this chapter. The final presentation of these stimuli will be adjusted to compensate for the hearing threshold elevation of subjects as explained in Section 4.3 and 4.4.

4.3 Participants

Nine NH listeners and nine HI listeners with sensorineural hearing loss participated in this experiment. All subjects were native English speakers. The NH listeners had hearing thresholds at or below 20 dB HL from 250 Hz to 8 kHz, and their ages ranged from 20 to 36. The NH listeners were not informed of the purpose or design of the experiment. Figure 4.1 shows the individual hearing thresholds for the aided ears of 9 HI subjects. The HI listeners all had mild to severe hearing losses and most of them had sloping high frequency hearing losses. All listeners were tested monaurally. All the HI listeners were experienced hearing aid users and their ages ranged from 18 to 30. The tests were performed with their hearing aids taken off, and compensation was applied to each HI subject individually. Specifically, a linear gain prescription was computed from each individual’s audiogram through the NAL-R procedure (Dillon, 2001). The NAL-R procedure is a predetermined formula that produces appropriate frequency-dependent gain response to compensate for threshold elevation and loudness recruitment. Table 4.1 shows the age, tested ear, cause of hearing loss and hearing aid experience of each HI participant. All of them are bilaterally hearing impaired. The experiment has been approved by the ethical committee in University of Southampton.

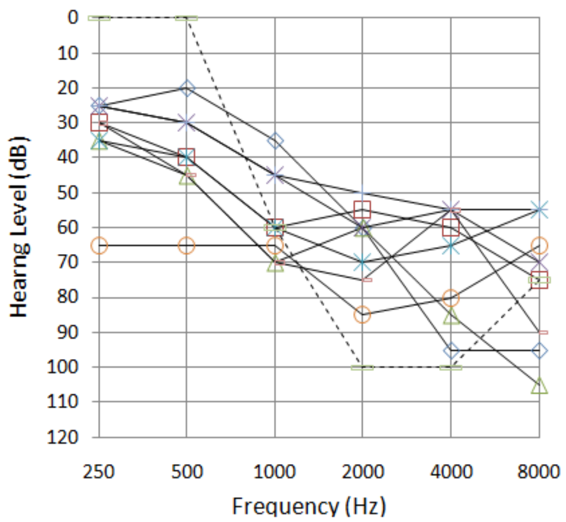


Figure 4.1: Audiograms showing the individual hearing thresholds for the aided ears of HI subjects (N=9).

Table 4.1: Age, tested ear, cause of hearing loss and hearing aid experience of the listeners with hearing losses. All of them are bilateral hearing impaired.

Listener	Age	Gender	Ear	Cause of hearing loss	Hearing aid experience
HI1	20	F	R	meningitis at 2 years old	16 years
HI2	31	F	R	Congenital	31 years
HI3	22	F	R	Congenital	20 years
HI4	18	F	R	Congenital	14 years
HI5	21	F	R	Congenital	18 years
HI6	20	F	R	Tinnitus, noise exposure	4 years
HI7	22	M	L	Congenital	18 years
HI8	20	F	R	Congenital	19 years
HI9	22	M	R	congenital, hereditary	6 years

4.4 Equipment

All listeners were seated in a sound-isolated room and listened to the sounds presented through

Sennheiser HDA 200 headphones presented through a Behringer UCA202 sound card and Creek OBH- 21SE headphone amplifier. The presentation levels of speech were kept at 65 dB SPL for NH listeners and are adjusted individually for each HI listener to their comfortable and audible level.

4.5 Procedure

4.5.1 **Procedure in speech intelligibility tests**

There were a total of six test conditions in this experiment: two noise types (SSN, babble) by three noise reduction conditions (no noise reduction, SCS, WF). No noise reduction condition means noisy speech which shows baseline performance. BKB lists were randomly selected from the corpus for each test condition. Subjects were instructed to repeat as many words as they could after listening to each sentence, and they were not given any feedback during the tests. Practice was given for each subject with one randomly selected condition. The order of the six conditions was balanced among the listeners. A subject test with the six conditions and a training session together took less than 1 hour.

The speech recognition test was performed through a three-up-one-down adaptive procedure as described in (Dahlquist et al., 2005) to find the speech reception threshold (SRT in dB) required for 79.4% correct recognition in each condition. A sentence was deemed to have been recognised correctly when at least two keywords were repeated correctly. Sentence order was controlled so that participants did not receive the same sentence repeatedly. The step size of the procedure was 1 dB. Only one trial was performed in each condition with each subject.

4.5.2 **Procedure in speech quality tests**

There are a total of eight conditions in this experiment: two noise types (SSN, babble) by two noise reduction algorithms (CS-WF, SCS) by two compared SNR conditions (unprocessed speech at 5 dB SNR vs processed speech at 5 dB SNR, unprocessed speech at 10 dB SNR vs processed speech at 5 dB SNR) as shown in Table 4.2. Unprocessed speech indicates noisy speech while processed speech indicates noisy speech processed with CS-WF or SCS. The speech quality tests is long running speech concatenated with BKB sentences.

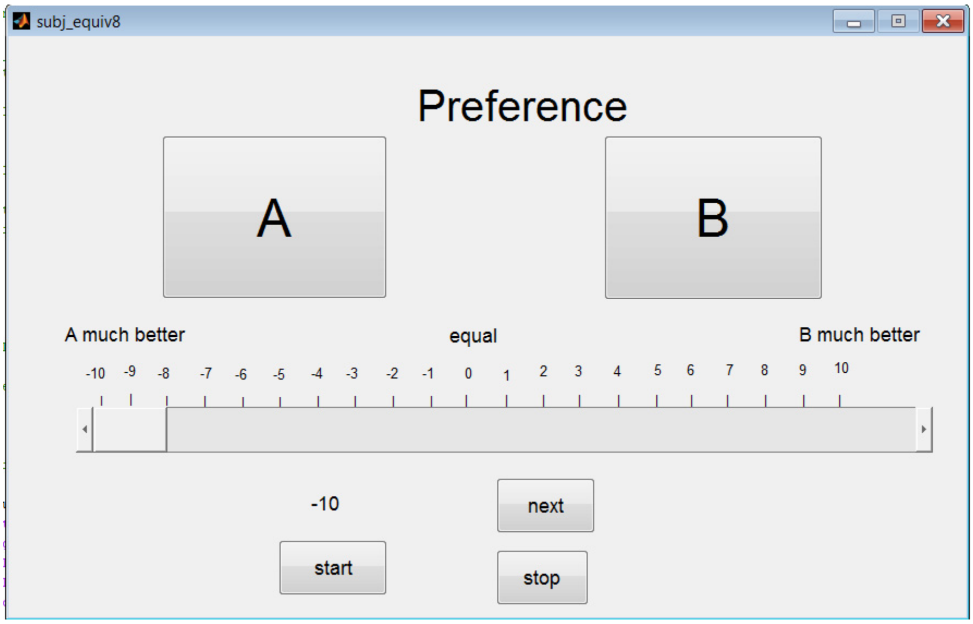


Figure 4.2: An example of MATLAB GUI for paired comparison rating of speech quality. Qualities rated were ‘preference’ and ‘noise loudness’. The buttons A and B refer to two stimuli under comparison and can be switched by clicking. The slider is used for compared quality rating from a subject with the range between −10 and 10.

Figure 4.2 shows MATLAB GUI developed by the author for paired comparison rating. The principle of this interface is adopted from that in (Dahlquist et al., 2005). Participants could switch between processed and unprocessed sound by clicking the button A or B. They were not told which setting corresponded to the processed speech and the assignment was alternated between conditions. They were instructed to keep listening until they reached a final quality impression. Ratings were recorded by using the slider which was quantified by measurement normalized to a ± 10 unit scale. The title ‘preference’ or ‘background noise’ was shown to indicate which quality dimension needs to be rated. The instruction sheet is shown in Appendix C. The remaining part of this section will cite the principle of ICPR (Dahlquist et al., 2005).

Two conditions for each type of noise with each noise reduction algorithm were used: 1) comparison between processed signal at input 5 dB SNR and unprocessed signal at input 5 dB SNR with rating value R_0; 2) comparison between processed signal at input 5 dB SNR and unprocessed signal at input 10 dB SNR with rating value R_5. The ratings of the comparison between processed and unprocessed stimuli were collected for both conditions. The point of subjective equality was obtained either by linear interpolation between these two difference values or by linear extrapolation. Figure 4.3 gives an example of this interpolation as Dahlquist *et al.* (2005). For each subject (index *n*), the SNR gain G_n is obtained through the rating pairs by the following equation,

$$G_n = 5R_{0n} / (R_{0n} - R_{5n})$$

Where R_{0n} indicates rating value in condition 1 and R_{5n} indicates rating value in condition 2 described above. G_n was limited to ± 10 dB, because single extrapolated values may incidentally become very large to exclude any extremes (for example from confusing items in the pair). The measurement of G_n for each noise with each algorithm is repeated four times and the final SNR gain for each subject is the median value from the four SNR gains.

Table 4.2: Combinations of speech-to-noise ratios used for the various types.

Noise type	*Noise reduction algorithm*	*Speech-to-noise ratio (dB) for processed/unprocessed item*
Speech shaped noise	CS-WF	5/5
		5/10
	SCS	5/5
		5/10
Babble noise	CS-WF	5/5
		5/10
	SCS	5/5
		5/10

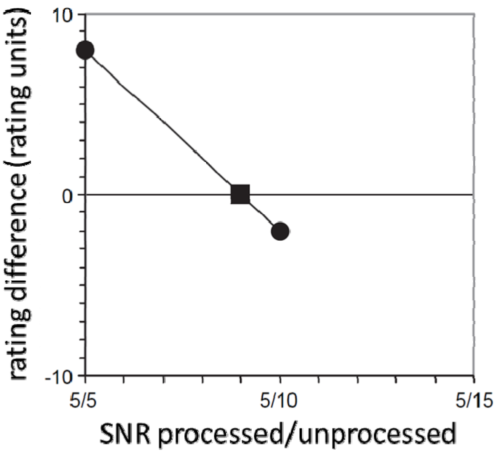


Figure 4.3: The method used to estimate the point of subjective equality (PSE) (filled square) from paired comparison ratings (filled circles), calculated by interpolation or extrapolation (IPCR method). In this example, the pair of SNRs for subjective equality is interpolated to 5/9 dB for processed/unprocessed stimuli (i.e. an SNR-gain of + 4 dB). Example of SNR gain from (Dahlquist et al., 2005).

4.6 Statistical Analysis

Speech reception thresholds (SRTs in dB) are the results from the speech recognition tests and will be processed through analysis of variance (ANOVA). ANOVA was performed on all the SRTs with NH and HI subjects, with within-subject factors of type of processing (noisy speech, CS-WF and SCS) and of type of noise (speech shaped noise and babble noise) and a between-subject factor of subject type (NH and HI). Post hoc tests, the Bonferroni test and/or the Fisher least-significant-difference (LSD) test, were applied where appropriate. The Bonferroni test was used as the most conservative test to verify the hypothesis of significant difference. The Fisher LSD test was used as the most sensitive test to verify the hypothesis of null difference.

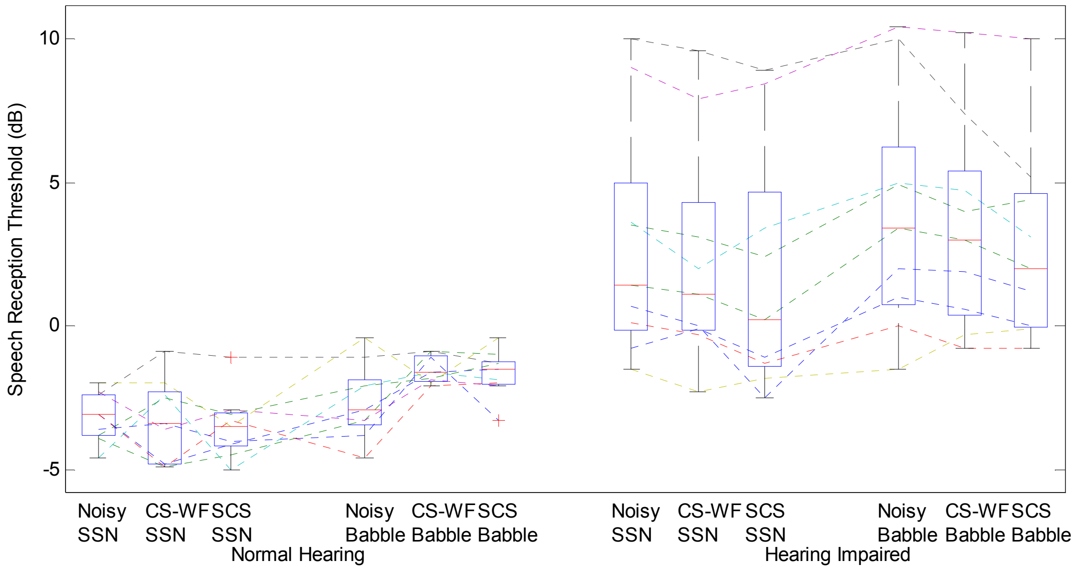
SNR gains (in dB) are the results from speech quality tests (IPCR) for each participant. This measure was based on four replicated paired comparison ratings (one for equal SNR condition and one for unequal SNR condition). Four SNR gains are deduced from the four replicated paired comparison ratings respectively in each condition (one algorithm under one noise type) for each participant. The median SNR gain is deduced from the four replicated SNR gains as described in Section 4.5.2. A Wilcoxon matched-pairs signed ranks test was used to test if the median SNR gain differed significantly ($p<0.05$) from 0 and 5 dB. If the median SNR gain is significantly larger than 0 dB, this indicates subjective speech quality is significantly improved with noise reduction algorithms. If the median SNR gain significantly larger than 5 dB, this indicates subjective speech quality is much better improved with noise reduction algorithms.

Multivariate forward linear regression is performed to set up a prediction model of speech recognition gain with the factors of hearing threshold and SRT with noisy speech (baseline performance). This multivariate linear regression is to check the individual factors that affect the noise reduction performance. The forward regression only extracts the strongest factors that affect the model and excludes the less important factors. The main factors in regression, the correlation coefficients and the significance of correlation are measured through forward stepwise linear regression.

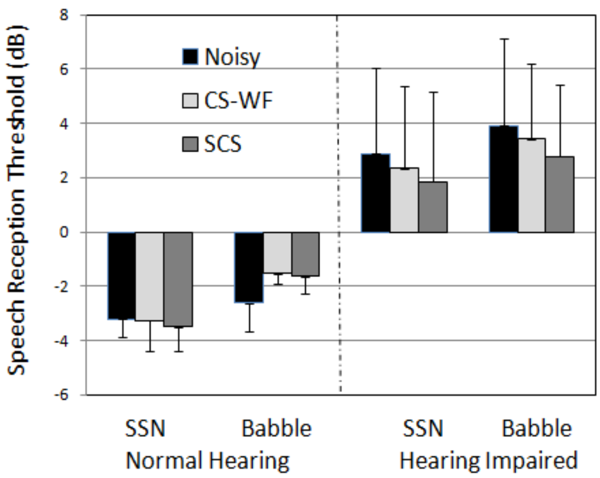
4.7 Results

4.7.1 **Speech recognition results**

Figure 4.4 shows the speech recognition performance of all participants with the results of SRT in all six test conditions: SSN-Noisy, SSN-CS-WF, SSN-SCS, Babble-Noisy, Babble-CS-WF, Babble-SCS, for both 9 NH (left) and 9 HI subjects (right). Figure 4.4 (a) shows the spread of SRTs in box plots. Figure 4.4 (b) shows the average SRT in each condition with a 95% confidence range. The motivation to show the intelligibility results of NH and HI listeners in the same figure is mainly to test whether there is any difference between them, especially in the respect of group benefits from noise reduction algorithms. Although the motivation of the whole thesis is to develop and evaluate noise reduction algorithms for HI listeners, adding tests with NH listeners can indicate the difference in noise reduction effects between NH and HI listeners and accordingly imply how to modify the noise reduction algorithms for HI listeners that were originally developed for NH listeners.



(a)



(b)

Figure 4.4: SRTs for different conditions in 9 NH and 9 HI listeners. SSN: speech shaped noise; Noisy: noisy speech without noise reduction algorithms; CS-WF: the comparison algorithm; SCS: sparse coding shrinkage. (a) Boxplots of SRTs. On each box, the central mark is the median; the edges of the box are the 25th and 75th percentiles; the whiskers extend to the most extreme measured data not considered outliers. A Ladder diagram is also plotted showing individual performance. (b) Mean SRTs with error bars indicate the 95% confidence intervals of the means. A more negative SRT corresponds to better performance.

The spread of the SRTs in Figure 4.4 (a) illustrates that the HI subjects show correspondingly large inter-subject variability compared to NH subjects. We assume that this is due to individual auditory deficits as well as individual experience with hearing aids.

For NH listeners, as shown in left panel of Figure 4.4 (b) and the second column of Table 4.3, the mean SRT for noisy speech with SSN is −3.2 dB, for CS-WF with SSN is −3.3 dB, for SCS with SSN is −3.5 dB, for noisy speech with babble noise is −2.6 dB, for CS-WF with babble noise is −1.54 dB and for SCS with babble noise is −1.6 dB. The results were found to be normally distributed by the Shapiro-Wilk test. A two-way repeated ANOVA shows that for NH subjects, the effect of noise reduction algorithm is not significant [$F(2,16)=2.4$, $p>0.05$], but the effect of noise type is significant for NH subjects [$F(1,8)=27.7$, $p<0.05$]. There is no interaction between noise type and noise reduction algorithm [$F(2,16)=2.4$, $p>0.05$]. The non-significant effect of algorithm indicates that these noise reduction algorithms do not benefit NH subjects in speech intelligibility.

For HI subjects, as shown in right panel of Figure 4-4 (b) and the second column of Table 4.4, the mean SRT for noisy speech with SSN is 2.9 dB, for CS-WF with SSN is 2.3 dB, for SCS with SSN is 1.8 dB, for noisy speech with babble noise is 3.9 dB, for CS-WF with babble noise is 3.4 dB and for SCS with babble noise is 2.8 dB. The results were found to be normally distributed by the Shapiro-wilk test. A two-way repeated ANOVA was also performed to detect within-subject noise type effect and algorithm effect across HI subjects. For HI listeners, both the noise reduction algorithm and the noise type have significant effects [$F(2,16)=9.4$, $p<0.05$, and $F(1,8)=5.5$, $p<0.05$, respectively]. There is no interaction between noise reduction algorithm and noise type [$F(2,16)=0.04$, $p>0.05$]. The significant effect of the algorithm indicates noise reduction algorithms can significantly improve speech intelligibility for HI subjects.

A three-way repeated ANOVA where both NH and HI subjects were included showed that the main effects of subject type, noise type were significant [$F(1,16)=17.767$, $p<0.05$, and $F(1,16)= 22.589$, $p<0.05$, respectively], but the main effect of the noise reduction algorithm is not significant [$F(2,32)=3.085$, $p>0.05$]. There is a significant interaction effect between subject type and noise reduction algorithm [$F(2, 32)=9.210$, $p<0.05$]. There are no significant interaction effects between subject type and noise type or between noise type and noise reduction algorithm [$F(1,16)= 0.55$, $p>0.05$, $F(2,32)=1.423$, $p>0.05$, respectively]. There is no three-way interaction between subject type, noise reduction algorithm and noise type [$F(2,32)=1.561$, $p>0.05$]. These results indicate that the performance of noise reduction algorithms depend on the hearing loss level and the noise type.

Fisher LSD post hoc tests were performed to detect the difference in performance between any pair of the six conditions across NH subjects (Table 4.3) and HI subjects (Table 4.4) separately. The numbers in the brackets are to give each condition an identification number. The numbers without brackets in the second column present the average SRT of each condition. Significant differences ($p < 0.05$) are shown in boldface. For NH listeners, the noise reduction algorithms barely improve speech intelligibility in speech shaped noise (compare (1) and (2), (1) and (3), in Table 4.3) but significantly deteriorate speech intelligibility in babble noise (compare (4) and (5), (4) and (6), in Table 4.3). There is no significant intelligibility difference between speech in babble noise and speech in speech shaped noise within NH subjects (compare (1) and (4) in Table 4.3). For HI listeners, the noise reduction algorithms significantly improve speech intelligibility in speech shaped noise (compare (1) and (2), (1) and (3), in Table 4.4) but not significantly in babble noise (compare (4) and (5), (4) and (6), in Table 4.4). There is a significant intelligibility difference between speech in babble noise and speech in speech shaped noise within HI subjects (compare (1) and (4) in Table 4.3). Comparison between SCS and CS-WF in speech shaped noise through paired sample t-test shows that: the power is only 0.2 with the available 9 HI subjects; at least 47 subjects are needed for $p<0.05$ at 80% power to detect a within-subject between-condition difference of 1.0 dB. Comparison between SCS and CS-WF in babble noise through paired sample t-test shows that: the power is 0.37 with the available 9 NH subjects; at least 23 subjects are needed for $p<0.05$ at 80% power to detect a within-subject between-condition difference of 1.0 dB.

Table 4.3: Fisher LSD post hoc significant tests for the interaction of noise reduction algorithm and noise type in the experiment with NH subjects. Significant effects ($p < 0.05$) are given in boldface. The number in each bracket indicates the number of each condition.

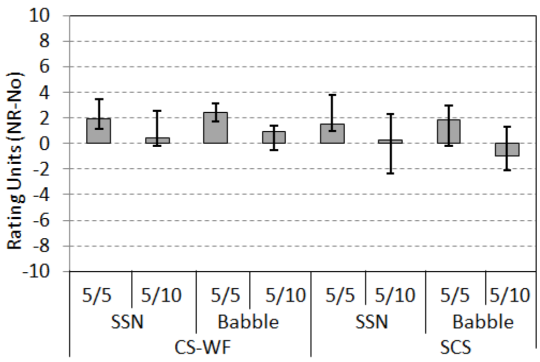
Processing condition	Mean SRT	(1)	(2)	(3)	(4)	(5)
SSN-Noisy	(1) −3.2					
SSN-CS-WF	(2) −3.3	0.896				
SSN-SCS	(3) −3.5	0.317	0.598			
Babble-Noisy	(4) −2.6	0.228	0.050	0.107		
Babble-CS-WF	(5) −1.5	**0.002**	**0.007**	**0.001**	**0.048**	
Babble-SCS	(6) −1.6	**0.002**	**0.011**	**0.003**	**0.018**	0.77

Table 4.4: Fisher LSD post hoc significant tests for the interaction of noise reduction algorithm and noise type in the experiment with HI subjects. Significant effects ($p < 0.05$) are given in boldface. The number in each bracket indicates the number of each condition.

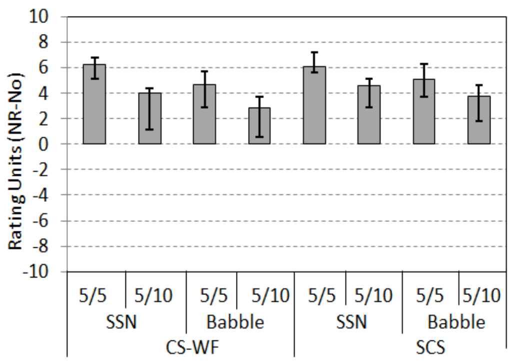
Processing condition	Mean SRT	(1)	(2)	(3)	(4)	(5)
SSN-Noisy	(1) 2.9					
SSN-CS-WF	(2) 2.3	**0.030**				
SSN-SCS	(3)1.8	**0.009**	0.233			
Babble-Noisy	(4)3.9	**0.006**	**0.001**	**0.001**		
Babble-CS-WF	(5) 3.4	0.292	0.074	**0.018**	0.169	
Babble-SCS	(6)2.8	0.865	0.525	0.210	0.073	0.058

4.7.2 **Speech quality results**

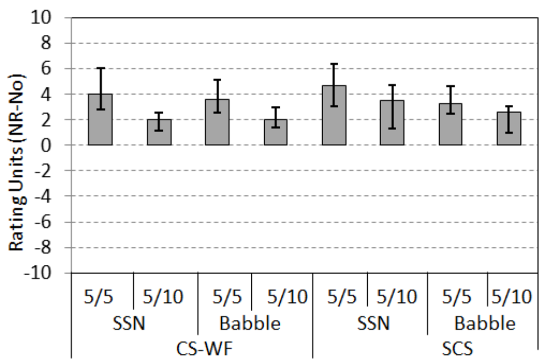
(a) Paired comparison Rating – Preference (NH)



(b) Paired comparison Rating – Preference (HI)



(c) Paired comparison Rating – Noise Loudness (NH)



(d) Paired comparison Rating – Noise Loudness (HI)

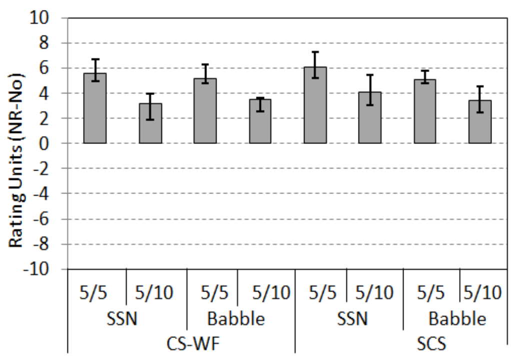


Figure 4.5: Subjective ratings from paired comparison rating (PCR) tests for two sound quality dimensions, two noise types and two noise reduction algorithms with NH and HI listeners separately. Median values of rating difference between processed and unprocessed signals (error bars: inter-quartile range). SSN: speech shaped noise. Labels, such as 5/10, indicate (SNR processed) / (SNR unprocessed) in dB. Left column: NH listeners. Right column: HI listeners. Upper row: ‘Preference’. Lower row: ‘Noise loudness’. Larger values indicate greater preference for, or less noise loudness in processed speech than in unprocessed speech.

Median values of Paired Comparison Ratings for the difference between processed and unprocessed speech are shown in Figure 4.5 for two rating categories (‘preference’ and ‘noise loudness’.) The rating will be given a positive value if a subject prefers the overall quality of the processed speech to the unprocessed speech from overall quality impression (a-b) or if a subject regards the processed speech contains less noise than the unprocessed speech (c-d) as explained earlier. The rating for noise loudness is just the inverse definition of that in Dahlquist *et al.* (2005) where a subject gives a negative rating value if he/she regards the sound in processed speech contains more noise than the sound in unprocessed speech. The inverse definition in our research is to put the ratings of "preference" and "noise loudness" in a more intuitive comparison as shown in Figure 4.5. As 'preference' contains several dimensions (e. g. 'clarity', 'comfort', 'noise loudness') with different perceptual weights for individual subjects, comparison between 'preference' and 'noise loudness' could show the weight of 'noise loudness' in overall subjective rating of 'preference'.

The filled bars represent the median ratings of the differences (processed minus unprocessed) for each noise condition with each noise reduction algorithm in either normal hearing group or hearing impaired group, while error bars indicate the inter-quartile range. Ratings for the various speech-to-noise ratios and noise types are presented in separate bars. Higher bars indicate higher ratings of the processed signal compared to the unprocessed signal on the sound quality dimension, e.g. more 'preference' or less 'noise loudness'. The results are shown in separate panels for the NH group and HI group. All the four panels in Figure 4.5 show positive bars (except one bar in panel (a)) and thus indicate advantages of processed speech over unprocessed speech in either equal SNR conditions or unequal SNR conditions. Comparison between panel (a) and panel (b) indicates NH listeners don't prefer quality of noise reduction algorithms as much as HI listeners do. The comparison between CS-WF and SCS in panel (a) shows that NH subjects prefer the quality of CS-WF rather than SCS. Panel (c) indicates NH listeners can clearly perceive the benefits of noise reduction algorithms in reducing 'noise loudness'. The different levels of quality benefits for NH subjects between 'preference' (a) and 'noise loudness' (c) indicate that NH listeners have considered other quality dimensions (e. g. 'clarity', 'comfort', etc.) besides 'noise loudness' when rating their 'preference' of overall speech quality. Panels (b) and (d) indicate that for HI listeners, both the noise reduction algorithms show obvious benefits in both rating categories 'preference' and 'noise loudness'. The accordance between (b) and (d) implies that the HI listeners place more weight on 'noise loudness' when rating their 'preference' of overall speech quality. The contrast between (a) and (b) suggests that HI subjects are less sensitive to speech distortion than NH subjects. The larger ratings in panel (d) relative to panel (c) indicate that HI subjects are more sensitive to 'noise loudness' than NH subjects. The results in panel (b) and panel (d) are also comparable with the results in Dahlquist *et al.* (2005) where HI subjects gave similar range of quality ratings in 'preference' and 'noise loudness' when tested using a nonlinear spectral subtraction algorithm.

Median speech-to-noise ratio gains (SNR gains) across subjects in the normal hearing group and hearing impaired group are presented in Table 4.5. As explained in Section 4.5.2, the individual SNR gain measures were obtained by linear interpolation of the ratings for two conditions. In one condition, the SNR was the same for the processed and unprocessed stimuli; in the other condition, the SNR for the unprocessed condition was 5 dB higher. All the SNR gains with positive values indicates that noise reduction algorithms can improve speech quality for NH and HI listeners with each noise type under each rating category. A Wilcoxon matched-pairs signed ranks test was used to test if the SNR gains within the NH or HI group are significantly larger than 0 dB or 5 dB ($p<0.05$). ** indicates SNR-gain >5, * indicates SNR-gain >0, both significant at the 5%-level.

This table also shows the different benefits of noise reduction algorithms to NH and HI listeners. In Table 4.5, the value in every even line ('HI') always being larger than or equal to the value above ('NH') indicates that HI listeners acquire larger SNR gain in sound quality than NH listeners with each noise reduction algorithm in each noise under each rating category ('preference' or 'noise loudness'). This shows that noise reduction algorithms are more beneficial to HI listeners than to NH listeners across different noise types and different speech quality criteria. Furthermore, the difference in the category of 'preference' between 'NH' and 'HI' indicates the overall quality with noise reduction algorithms is more acceptable by HI listeners than NH listeners. In Table 4.5, for NH listeners, the SNR gain of 'noise loudness' (the 5th row) is always larger than that of 'preference' in the same condition (the 3rd row). This indicates that for NH subjects, although the reduction in 'noise loudness' with algorithms is recognized, this only partially contributes to the final judgment of 'preference' in overall quality. In Table 4.5, for HI listeners, the SNR gains of ‘noise loudness’ and ‘preference’ are all around 10 dB (the 4th and 6th rows). This also indicates that ‘noise loudness’ is a more weighted factor in determining ‘preference’ of overall speech quality in HI subjects. Noise reduction algorithms can reduce noise at the price of introducing speech distortion which can be sensitively detected by NH subjects but can be tolerated by HI subjects.

Results from subjective SNR gain in ‘preference’ across NH subjects were compared with an objective quality measure fwsegSNR shown in Table 4.6. This is mainly to show whether there were any similar trends between subjective quality tests and objective measures for NH subjects. The results of fwsegSNR of the processed speech were calculated assuming the input SNR was 5 dB (also shown in Figure 3.10 a). Within NH listeners, compared to CS-WF, SCS brought more SNR-gain in speech shaped noise but less SNR-gain in babble in the category of ‘preference’. The objective measure fwsegSNR also showed the similar trend as the subjective SNR gain. Therefore, there was some agreement between these two measures that supports the reliability of fwsegSNR in predicting speech quality for NH subjects.

Table 4.5: Median values for SNR-gain (in dB) for the rating categories “preference” and “noise loudness”. Results are for 9 NH subjects and 9 HI subjects. ** indicates SNR-gain >5, * indicates SNR-gain >0, both significant at the 5%-level.

Rating Category	Hearing Level	CS-WF		SCS	
		ssn	babble	ssn	babble
Preference (dB)	NH	5.9**	7.3**	8.8**	3.1*
	HI	10.0**	9.9**	10.0**	10.0**
Noise Loudness (dB)	NH	9.1**	9.7**	10.0**	10.0**
	HI	9.8**	10.0**	10.0**	10.0**

Table 4.6: Objective and subjective noise reduction effects (dB) for babble noise and speech shaped noise. Frequency-weighted segmental SNR (fwsegSNR) and subjectively estimated with IPCR method for 'preference' criterion from normal hearing group.

Noise reduction effect	SSN		Babble	
	CS-WF	SCS	CS-WF	SCS
Physical (dB)	7.0	7.5	7.5	7.0
Subjective (dB)	5.9	8.8	7.3	3.1

4.7.3 **Speech recognition versus subjective sound quality**

An important characteristic of the IPCR methodology is that it allows for examination of sound quality judgements and subjective speech recognition results on the same quantitative scale. The examination between subjective quality and intelligibility performance provides an intuitive vision as shown in Figure 4.7 whether noise reduction algorithms can be beneficial in either quality or intelligibility within each group (NH or HI). This comparison also can imply overall benefits from noise reduction algorithms for each group.

Figure 4.6 shows individual SNR gains in speech recognition and in quality dimension 'preference' with each noise reduction algorithm (CS-WF and SCS) under each noise type (speech shaped noise and babble noise) within NH or HI group. The speech recognition gain is the SRT of the unprocessed speech minus the SRT of the processed speech. The SNR gain in 'preference' from IPCR tests was explained in Section 4.7.2 and its median values were shown in Table 4.5. Each plot shows individual SNR gain in quality and intelligibility with both CS-WF and SCS strategies under each noise within each group. In each plot, the thick lines divide the plot into four sections. The marks in the upper right section show individual SNR gains positive in both quality and intelligibility; the marks in the upper left section show individual SNR gain positive in quality but negative in intelligibility; the marks in the lower right section show individual SNR gain negative in quality but positive in intelligibility; the marks in the lower left section show individual SNR gain negative in both quality and intelligibility.

Gains are generally smaller in speech recognition than in quality, both measured in dB, in all the four subplots which means both noise reduction algorithms improve speech quality more than speech intelligibility. However, this does not necessarily mean that these two effects are of equal weight contributing to individual overall benefits from algorithms. For NH subjects listening to speech in speech shaped noise shown in (a), the SNR gains in speech recognition are clustered around 0 dB, which reflects no significant difference between processed speech and unprocessed speech as shown in Table 4.3. For NH subjects listening to speech in babble noise shown in (c), the SNR gains in speech recognition are almost all in the left half plot below 0 dB, which reflects a significant intelligibility decrease with processed speech as shown in Table 4.3. For HI subjects listening to speech in speech shaped noise and babble shown in (b) and (d) respectively, almost all the SNR gains in speech recognition are in the right half plot above 0 dB which further verifies that noise reduction algorithms bring more intelligibility benefits to HI subjects than to NH subjects. By inspecting the speech recognition gains with respect to the horizontal axis, the speech recognition gains of CS-WF and SCS scatter in the similar range in each plot which further suggests little significant intelligibility difference between SCS and CS-WF in each noise within each group. However, we should point out that SCS shows slightly greater speech recognition gains than CS-WF within HI subjects. For NH subjects evaluating speech quality in (a) and (c), most of the SNR gains in preference are above 0 dB and nearly half of them are near 10 dB. This indicates noise reduction algorithms generally improve speech quality for NH subjects. However, there are several SNR gains of preference with SCS below 0 dB which indicates some NH individuals perceive negative quality effects from noise reduction algorithms. In (a) and (c), SCS are generally less than with CS-WF in the individual SNR gains of ‘preference’ within NH subjects. SCS introduces some speech distortion while reducing more noise. For HI subjects evaluating speech quality in (b) and (d), all the SNR gains in ‘preference’ are in upper plots above 0 dB and most of them are near 10 dB which indicates noise reduction algorithms improve speech quality more for HI subjects than for NH subjects. This also results in the corresponding difference of median SNR gain in ‘preference’ between NH and HI subjects shown in Table 4.5.

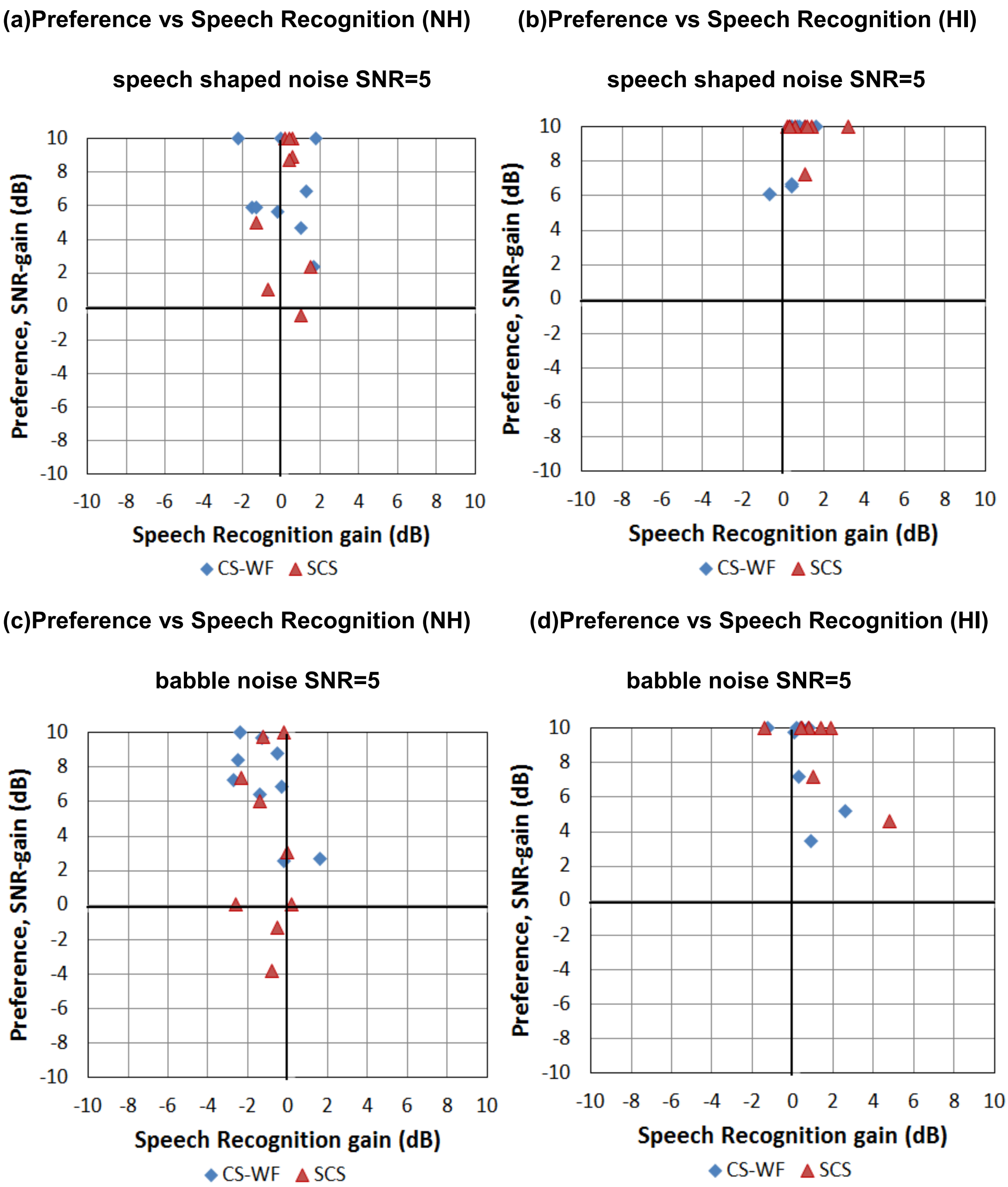


Figure 4.6: Scatter plots showing individual speech-to-noise ratio gains in the quality dimension of 'preference' versus speech recognition gain in different condition with NH or HI listeners. Positive values with respect to horizontal or vertical axis indicate the noise reduction algorithm is beneficial. SNR-gain is limited up to +10 dB.

4.7.4 **Multivariate regression of speech recognition gain**

Single channel noise reduction algorithms seem to still have difficulty to improve speech intelligibility that is quantified as speech recognition gain in the present thesis. Figure 4.4 shows that there are large differences between NH and HI subjects in speech recognition gain from noise reduction algorithms. HI subjects generally acquire higher speech recognition gain than NH subjects. It may be due to the difference between NH and HI subjects. Thus one factor in speech recognition gain may be the hearing loss level. However, it is worth noting that, HI subjects were tested at higher input SNRs, as the SRT of noisy speech was higher in HI listeners than in NH subjects. Therefor the other factor in speech recognition gain may be the SRT of unprocessed noisy speech. To investigate the relationship between speech recognition gain and the above two factors (the hearing loss level, and the SRT of noisy speech). Multivariate regression of speech recognition gain is conducted. As defined in Section 4.7.3, the speech recognition gain is the SRT of the unprocessed speech minus the SRT of the processed speech. A positive speech recognition gain indicates a positive benefit in intelligibility from noise reduction algorithms. The hearing loss level is quantified here by averaging the hearing thresholds at the three frequencies (1, 2, 4 kHz) in each subject. It is assumed the thresholds in the corresponding frequency range affect more in understanding English language (Amos and Humes, 2007). The SRT of noisy speech indicates the baseline performance of unprocessed noisy speech. The higher the SRT with noisy speech, the worse the baseline performance is.

The relationship between each factor and speech recognition gain was primarily investigated. The scatter plots in Figure 4.7 show the relationship between speech recognition gain and average hearing threshold in four conditions. The scatter plots in Figure 4.8 show the relationship between speech recognition gain and baseline performance in four conditions. The four conditions (a-d) in Figure 4.7 and Figure 4.8 are CS-WF in speech shaped noise, SCS in speech shaped noise, CS-WF in babble noise, SCS in babble noise in order. The larger value of R^2 (correlation squared) in each plot shows the higher level of correlation between the factor and the speech recognition gain. Through statistical analysis, the factor of average hearing threshold or baseline performance correlates significantly ($p<0.05$) with speech recognition gain in babble but not in speech shaped noise. However, the fact that Figure 4.8 (c, d) give higher values of R^2 than Figure 4.7 (c, d), suggests the SRT of noisy speech is a more important factor.

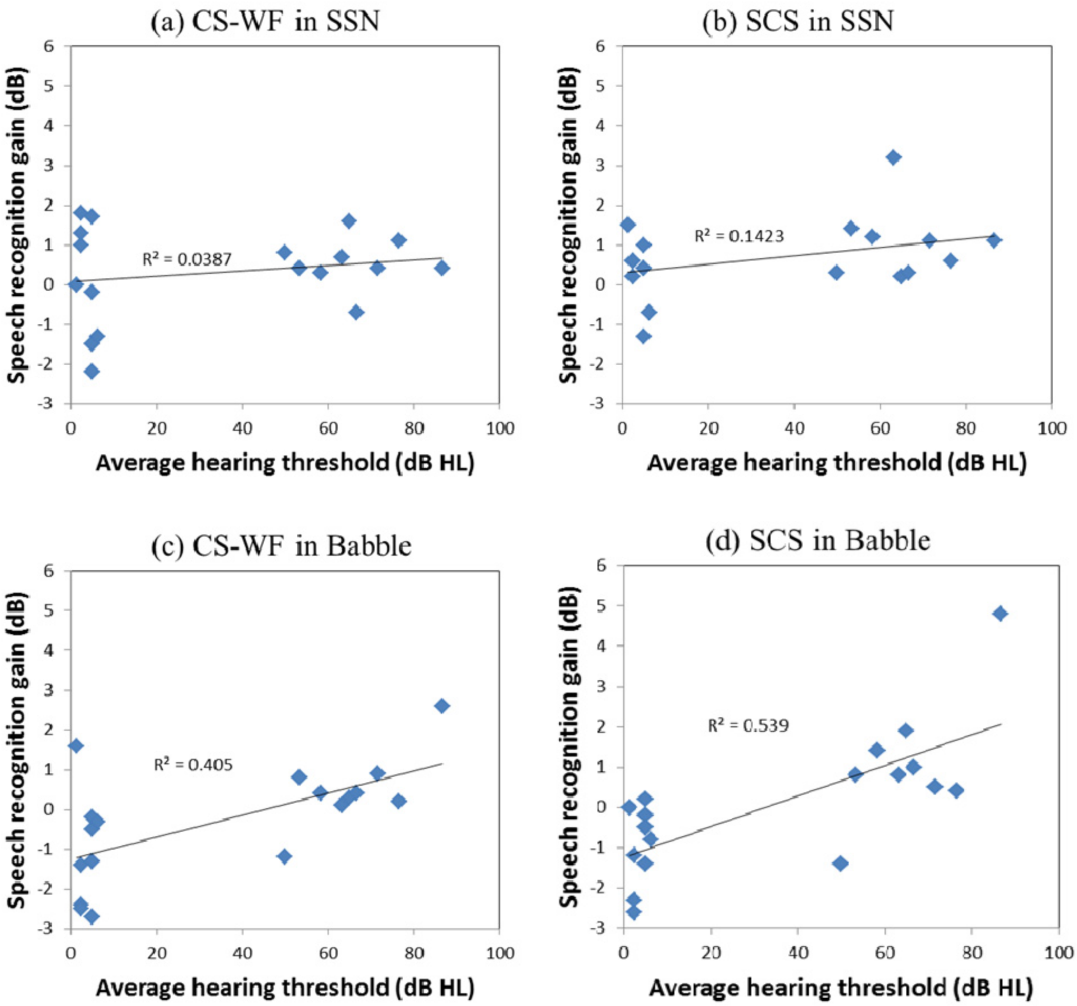


Figure 4.7: Scatter plots showing speech recognition gain versus average hearing threshold in four conditions (SCS-SSN, SCS-Babble, CS-WF-SSN, CS-WF-Babble).

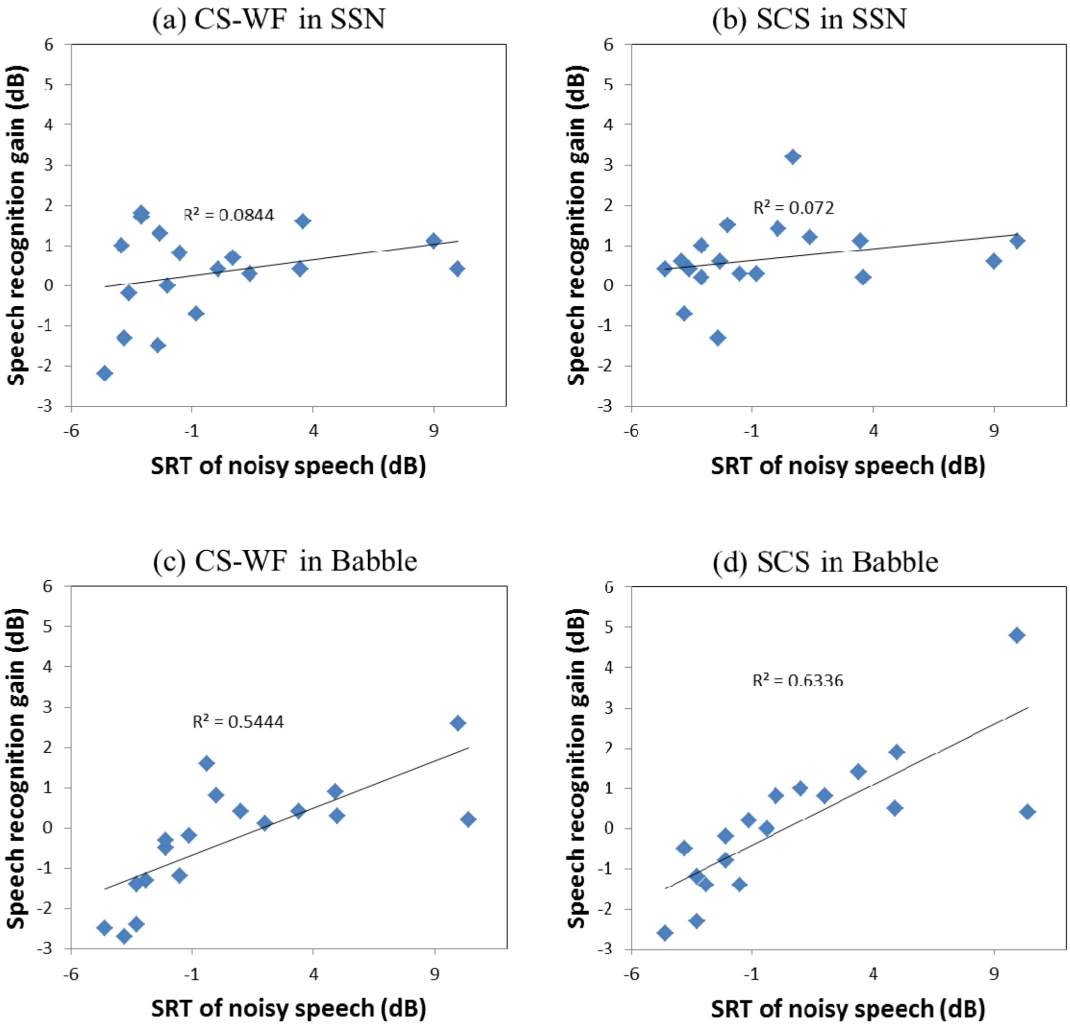


Figure 4.8: Scatter plots showing speech recognition gain versus average SRT of unprocessed noisy speech (baseline performance) in four conditions (SCS-SSN, SCS-Babble, CS-WF-SSN, CS-WF-Babble).

Then, the contributions of the two factors to the speech recognition gain in babble noise were differentiated through multivariate forward stepwise linear regression. The forward stepwise regression can include the most important variable and exclude the less important variable in the linear model. Table 4.7 and 4.8 show the results of multivariate forward stepwise linear regression of the speech recognition gain in babble noise with the algorithms of CS-WF and SCS respectively. In the regression model, the dependent variable is the speech recognition gain, and the two independent variables are average hearing threshold and SRT in noisy speech. Table 4.7 and 4.8 show that the factor of hearing threshold can be excluded compared with the factor of baseline performance in the regression. This suggests that the factor of SRT in noisy speech contributes more to speech recognition gain than the average hearing threshold does. This further indicates that the speech recognition performance is mainly due to the input SNR of noisy speech. This can also clarify the previous speculations that the performance of noise reduction algorithms depends on the hearing loss levels. In fact, as HI subjects were usually presented with speech at higher input SNR and noisy reduction algorithms may perform better with noisy speech at higher SNR, HI subjects seemed to perform better than NH subjects in many pervious evaluations.

Table 4.7: The forward stepwise regression of speech recognition gain with the algorithm of CS-WF in babble noise.

	B	Std Error B	β	Sig.
Included variables				
SRT in Noisy	0.233	0.053	0.738	**0.000**
Excluded variables				
Average HL			0.004	0.991

Note: B is the unstandardized coefficient, β is the standardized coefficient, the value in boldface indicates significance (p<0.001).

Table 4.8: The forward stepwise regression of speech recognition gain with the algorithm of SCS in babble noise.

	B	Std Error B	β	Sig.
Included variables				
SRT in Noisy	0.302	0.057	0.796	**0.000**
Excluded variable				
Average HL			0.189	0.544

Note: B is the unstandardized coefficient, β is the standardized coefficient, the value in boldface indicates significance (p<0.001).

4.8 Discussion

4.8.1 **Choice of noise conditions**

It is worth mentioning that the appropriate choice of noise type is an important methodological issue. Noise that contains mainly low-frequency components is easier to remove from speech compared to noise that shows the similar frequency spectrum shape as speech being used. Dillon and Lovegrove (1993) have indicated that the benefit of previous single-channel noise reduction systems in terms of speech intelligibility is small and the amount of improvement was greatest when the noise spectrum was weighted towards low frequencies. White noise and pink noise are also easy to reduce as they show a different pattern of spectrum from speech. To make the situation of speech in noise more realistic and difficult, speech shaped noise and babble noise were chosen in our study, both of which show similar average spectra as the speech.

4.8.2 **Challenge of babble noise for HI listeners**

HI listeners showed significantly more difficulty in multi-talker babble noise than in speech shaped noise (difference between conditions (1) and (4) in Table 4.4). Babble noise might represent a natural non-stationary background noise for example in social communication situations (‘cocktail party’ effect). This is in accordance with previous investigations that HI subjects performed worse in non-stationary noise (e.g. babble noise, cafeteria noise) than in stationary noise (e.g. speech shaped noise) (Wang et al., 2009). This is in accordance with the difficulty experienced by HI listeners in the ‘cocktail’ party effect. In contrast, NH listeners did not show such significant difficulty in babble noise (difference between conditions (1) and (4) in Table 4.3). NH listeners presumably can catch speech cues that exist in the intervals of babble noise with their normal hearing thresholds, sharp frequency selectivity and temporal resolution.

4.8.3 **Difference between NH and HI subjects in benefits from algorithms**

The speech recognition tests showed that: noise reduction algorithms barely improve speech intelligibility for NH subjects but significantly for HI subjects in speech shaped noise; noise reduction algorithms bring significantly negative intelligibility effects for NH subjects but do not affect intelligibility for HI subjects in babble noise.

The speech quality tests showed that: NH and HI subjects all recognized the noise reduction effects with algorithms; however, HI subjects gave higher ratings of noise reduction algorithms in overall speech quality than NH subjects.

In summary, noise reduction algorithms benefit HI subjects more than NH subjects in both speech intelligibility and speech quality. With hearing loss factors, HI listeners are less sensitive to speech distortion but more sensitive to noise loudness, compared to NH listeners. Noise reduction strategies are usually developed to reduce noise at the price of introducing speech distortion, which might be easily perceived by NH listeners but not by HI listeners. That may be one reason why noise reduction strategies are more beneficial to HI listeners than to NH listeners in speech quality and intelligibility. This is in accordance with (Schijndel et al., 2001) where it was shown that: HI subjects were less sensitive to spectral distortion; when speech is distorted speech intelligibility degrades significantly in NH subjects but not in HI subjects. This is also in accordance to the finding that noise reduction schemes based on the ideal binary mask could benefit HI listeners more than NH listeners (Wang et al., 2009) in speech intelligibility.

The relationship between the speech recognition gain and hearing threshold showed that the noise reduction benefit in speech intelligibility generally increases with the hearing loss level (Figure 4.7). However, the relationship between the SRT of noisy speech and speech recognition gain showed that the input SNR is a more important factor that determines the performance of noise reduction algorithms (Figure 4.8, Table 4.7, Table 4.8). This finding is new as pervious evaluations always thought the difference of performance between NH and CI users might be due to the factor of the hearing loss level. However, Section 4.7.4 dug out a deeper reason why NH users and CI users show different performance from noise reduction algorithms. In fact, they are not tested at the same input SNRs. This indicates that noise reduction algorithms perform better at higher SNRs. It can be further inferred that if NH users can be tested at higher SNRs without reaching ceiling performance, they may also get as much improvement as HI listeners do. This reminds us that we should be careful if we claim the performance of noise reduction algorithms is different between NH and HI listeners.

4.8.4 **Comparison between SCS and CS-WF**

On the whole, there was no significant difference in performance between SCS and CS-WF either within NH subjects or HI subjects. However, as SCS has reduced noise by a larger amount, a bit more distortion can be sensitively detected by NH subjects than can be tolerated by HI subjects. SCS presents sparse stimuli with a larger degree of noise reduction, which is more acceptable to HI subjects who are less sensitive to speech distortion and more sensitive to noise level due to the hearing loss factors. SCS performs slightly better in speech intelligibility than CS-WF within HI subjects. Therefore an algorithm that is not optimal for NH subjects might be optimal for HI subjects. We should be always careful if we directly adopted algorithms that were originally developed for NH subjects into hearing aid or cochlear implant applications.

4.8.5 **Acclimatization effects**

Acclimatization is the process in an individual organism adjusting to a gradual change in environment, allowing it to maintain performance across a range of environmental conditions. Acclimatization effects are important to the individual performance with noise reduction algorithms. Acclimatization effects may be different for different algorithms. Participants in our test did not have much time to adjust to each algorithm. If they had used the algorithm in a wearable hearing aid for several weeks, they might have obtained different results.

4.8.6 **Comparison with previous studies**

Although previous studies have also tested noise reduction algorithms in hearing aid users (Levitt et al., 1993, Elberling et al., 1993, Aharon et al., 2006, Dahlquist et al., 2005), they are not very comparable with our study as different studies used different test types, different speech materials and different noise types. However, our study can be easily compared with the results from Dahlquist *et al.* (2005) because we followed their procedure in testing speech recognition and speech quality with the same speech materials and the similar noise materials.

Firstly, the mean SRTs of unprocessed speech (noisy speech) in our evaluation are compared to the results from Dahlquist *et al.* (2005). It is found that the SRT from HI subjects in our evaluation is 0.4 dB lower in speech shaped noise but 2.9 dB lower in babble noise. The lower SRT corresponds to better speech intelligibility performance. The comparison indicated that our study showed similar performance in speech intelligibility in speech shaped noise but better performance in babble noise compared to Dahlquist's study (2005). The difference might be due to the hearing loss levels and the ages of the subjects. Hearing threshold levels in our research (as shown in Figure 4.1) are approximately in the same range as that shown in Dahlquist *et al.* (2005). Therefore the hearing threshold level is not a factor of SRT difference in noisy speech. Similar hearing loss levels result in similar intelligibility performance in stationary noise. The ages of our participants (average 22 years, range 18-31 years) are much younger than the participants in Dahlquist's study (average 70 years, range 47-78 years). The same hearing threshold does not mean the same speech perception abilities. This might be a factor inducing a difference of SRTs in babble noise between the two groups. There might be at least two reasons why young subjects can perform better than old subjects. Firstly, young subjects usually have acute central processing abilities in speech recognition; secondly, young subjects usually acquire a hearing loss at an early age and have been trained to deal with speech communication in noise with stronger plasticity in the brain. However, reasons cannot be confirmed as we did not compare the equal number of young and old subjects.

Following this, the speech recognition gains are compared between present study and the study by Dahlquist *et al.* (2005). The speech recognition gains are most positive in our study in HI subjects which is different from Dahlquist *et al.* (2005) where most of the HI subjects show negative SRT gains. This might be due to the noise reduction algorithms themselves, the difference in age, or difference in experience with noise reduction algorithms.

The subjective quality results are also compared with Dahlquist's study (2005). Our study showed the similar pattern of paired comparison ratings (Figure 4.5, right column) and showed similar median SNR gains in speech shaped noise but a bit higher median SNR gains in babble noise (Table 4.4).

The comparison between present study and Dahlquist's study validates the reliability of present evaluation and suggests the effect of age on noise reduction performance. However, this effect is not clear unless both young and old HI subjects are recruited. Due to the difficulties in recruiting HI subjects, only 9 HI subjects from the University of Southampton voluntarily participated in our experiment. To make a balanced comparison in the 95% confidence range, 9 NH subjects from university were tested as well.

4.9 Conclusion

The two noise reduction algorithms in speech intelligibility and quality were tested through NH and HI subjects in speech shaped noise and babble noise. The speech intelligibility tests were performed through an adaptive procedure (three-up-one-down) to measure the speech reception threshold (SRT) corresponding to 79.4% correct recognition. The quality tests were performed through the interpolated paired comparison rating (IPCR) to measure the SNR gain in quality impression.

The noise reduction algorithms improve speech quality for both NH and HI subjects; the noise reduction algorithms did not improve speech intelligibility for NH subjects; the noise reduction algorithms significantly improved speech intelligibility for HI subjects in stationary noise but no significant improvement was observed in babble noise; the noise reduction strategies hold more promise to help HI subjects than NH subjects in both intelligibility and quality. Non-stationary noise (babble noise) is more difficult for noise reduction algorithms to suppress compared to stationary noise.

There is no significant difference in speech intelligibility between the two noise reduction algorithms within either NH or HI subjects. However, SCS shows slightly worse speech quality than CS-WF in NH subjects. SCS reduces noise to a larger amount at the price of introducing a bit more speech distortion which can be sensitively detected by some NH subjects but can be tolerated by HI subjects. Both the algorithms might help HI subjects more given more practice. Statistical power will increase with more subjects. The reason that SCS has not shown any advantage over CS-WF in babble noise might be that SCS assumes Gaussian distribution of noise while it is difficult to estimate the power of babble noise and then pre-whiten babble noise appropriately. Although SCS did not show significant advantage over CS-WF, it at least matched the performance of CS-WF which is a competitive state-of-the-art noise reduction algorithm.

Previous studies have already showed that there is difference in benefits from noise reduction algorithms in speech intelligibility between NH and HI listeners. The present study is the first study that pins down the factors behind this difference between NH and HI listeners. This difference is more related with the input SNR rather than the hearing loss level. Noise reduction algorithms might perform better at higher input SNRs. This reminds us to test subjects at appropriate input SNRs when evaluating signal processing strategies. The adaptive speech recognition procedure and the IPCR speech quality procedure are both recommended for algorithm evaluation studies.

Chapter 5 Subjective evaluation of noise reduction effects with the hearing loss simulation model

Although we performed experiments to evaluate noise reduction algorithms in hearing impaired (HI) subjects, it is in general more difficult to recruit HI subjects than to recruit normal hearing (NH) subjects. We assume that a HI subject's speech perception can be approximated by additive combination of a hearing loss simulation (HLS) model and a NH subject's speech perception. If we could predict the noise reduction effects for HI subjects through NH subjects' listening with an appropriate HLS model, the experiment would become much easier. If the HLS model is realistic, the effect of asking the HI subjects to listen to the speech could be approximated by asking the NH subjects listening to the same speech with a HLS model, which simulates the average hearing loss characters in the HI group. Accordingly, we evaluated noise reduction algorithms with the gamma-chirp based HLS model through NH subjects in speech recognition tests. The motivation of this study is to investigate whether the HLS model can predict intelligibility effects of noise reduction algorithms relative to baseline performance (unprocessed speech) for HI listeners. This experiment might give two implications if the experimental results with HLS model and NH subjects show similar results as with HI subjects. On one hand, it can further support the reliability of the HLS model in simulating impaired auditory processing. On the other hand, it can imply whether noise reduction algorithms show benefits in speech intelligibility for HI listeners. The two implications interact with each other and can also suggest whether the auditory filter based HLS model could be used to evaluate speech intelligibility effects of noise reduction algorithms for HI listeners, which is easier, compared to recruitment of HI listeners.

5.1 Introduction

Although several attempts (Hollube and Kollmeier, 1996, Taal et al., 2011b, Taal et al., 2011a) have been made to develop objective speech intelligibility measures, most of them have only been validated with NH listeners rather than HI listeners. Some research has already shown that some methods provide a very simple and very good prediction of speech intelligibility for NH listeners in a variety of listening situations (French and Steinberg, 1947, Kryter, 1962, Pavlovic, 1987). However, large deviations occur with severely HI listeners between the predicted speech intelligibility and the performance of the subjects (Pavlovic, 1984). Even for some objective intelligibility measures that take account of hearing loss factors (Kates and Arehart, 2005, Hollube and Kollmeier, 1996), their performance has only been validated with clean speech or noisy speech rather than with noise reduction algorithms. In other words, a reliable objective speech intelligibility measure to predict effects of noise reduction algorithms for HI listeners has not been developed.

Approaches to modelling speech intelligibility with impaired auditory factors in the literature can be divided into two classes: the first class describes speech intelligibility in terms of unmasked portions of the speech spectrum and estimates a rating value, such as articulation index (AI) or speech transmission index (STI). The speech intelligibility index (SII) concept for estimating intelligibility has been extended for the coherence speech intelligibility index (CSII) to evaluate broadband peak-clipping and centre-clipping distortion, with the coherence between the input and output signals used to estimate noise and distortion effects (Kates and Arehart, 2005). The second class attempts to model speech processing in the impaired auditory system. This can be more advanced in simulating impaired auditory mechanisms and impaired speech perception. Hollube and Kollmeier (1996) developed a speech intelligibility prediction model for hearing impaired listeners by simulating the impaired auditory mechanisms. Their model set parameters of the individual listener's hearing thresholds and measured temporal and spectral resolution. Hollube and Kollmeier (1996) adapted AI (articulation index) (French and Steinberg, 1947, Kryter, 1962, Pavlovic, 1987) and STI (speech transmission index) to predict intelligibility effects for HI listeners. All these objective measures show similar power of prediction accuracy in speech intelligibility effects for HI listeners. However, these were validated with noisy speech rather than with noise reduction algorithms which could introduce different speech distortions.

The accuracy of current objective intelligibility measures to assess noise reduction algorithms for HI listeners is still uncertain. The use of a HLS model is an alternative approach to predict noise reduction effects for HI listeners using NH listeners (Hu et al., 2011b). If the HLS model is a realistic simulation at impaired auditory processing in respect of intelligibility, the intelligibility performance of NH listeners listening to speech processed with the HLS model can approach the performance of HI listeners when listening to the same speech without the HLS model. As NH listeners are easier to recruit, the evaluation by combing the HLS model with NH listeners costs much less in time and effort.

There are several characteristics of this physiologically inspired HLS model. It can simulate hearing threshold elevation, loudness recruitment and reduced frequency selectivity based on a compressive gamma-chirp filter bank. Compared to the gammatome filter bank, the gamma-chirp filter bank adds a level-dependent asymmetric correction to the basic gammatone channel frequency response, thereby providing a more accurate approximation to the auditory frequency response (Irino and Patterson, 2001, Patterson et al., 2003, Irino and Patterson, 2006). Specific details of hearing threshold elevation, loudness recruitment and reduced frequency selectivity have been described in Chapter 2. As Drullman *et al.* (1994) suggested that reduced temporal resolution doesn't reduce speech intelligibility for most hearing-impaired people; we did not take account of reduced temporal resolution in our HLS model.

The effects of noise reduction algorithms might be affected by other impairment compensation strategies in a hearing aid, which could not be detected if a noise reduction algorithm is evaluated in isolation. Therefore appropriate impaired compensation strategies in a hearing aid are better to be combined with noise reduction strategies to mimic combined signal processing strategy effects in a hearing aid. A well-known compensation strategy, called NAL-R procedure (Dillon, 2001), is combined with noise reduction algorithms to test effects in the HLS model with NH listeners. NAL-R is a spectral gain prescription to produce amplification based on an individual audiogram to compensate for threshold elevation. The parameters in the NAL-R procedure are set according to the assumed hearing thresholds in the HLS model.

The structure of this chapter is as follows: Section 5.2 introduces how to produce the stimuli, with noise reduction algorithms, the NAL-R procedure and the HLS model. Sections 5.3 to 5.7 describe the participants, equipment, procedure and statistical analysis for the experiment respectively which are similar to Sections 4.3 to 4.7. The NH subjects who participated in the experiment described in Chapter 4 also participated in the experiment described in this chapter. Section 5.7 will present the results in intelligibility from HLS model with NH listeners and will also compare with results through HI listeners (Chapter 4). Our motivation is to evaluate whether the HLS model can predict the significance of difference between enhanced speech and unprocessed speech for HI subjects if there exists any. Therefore current comparison only compares whether there is according improvement with noise reduction algorithms between HI subjects and the HLS model. That is, the difference in SRT between enhanced speech and processed speech rather than the correlation in SRT of the same condition (enhanced or processed speech) between HI subjects and HLS model will be analysed. The comparison can further validate the accuracy of the HLS model in respect of intelligibility effects of noise reduction algorithms for HI listeners. Section 5.8 discusses the validity and limitation of the HLS model and its potential in setting up a purely objective intelligibility measure for HI listeners without subjects' participation.

5.2 Stimuli

The speech materials are BKB sentences (Bench et al., 1979) which are standard British sentences recorded by a female talker, consisting of 21 lists with 16 sentences in each list. There are three or four key words in each sentence. Two noise types are used, babble noise and speech shaped noise. The speech materials and noise materials are the same as in Section 4.2.

Figure 5.1 shows the procedure to produce the stimuli with the HLS model for NH listeners. The noisy sentences were first processed with or without noise reduction algorithms and then with the NAL-R procedure. After the NAL-R compensation, the speech was processed through the HLS model before finally presented monaurally to one ear of NH listeners through headphones. The processing stages mimic the noise reduction and impaired compensation strategies in a hearing aid and the impaired speech perception in a HI listener.

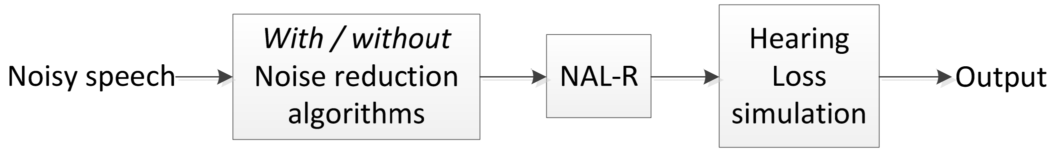


Figure 5.1: Flowchart of how to produce the stimuli with the HLS model for NH listeners,

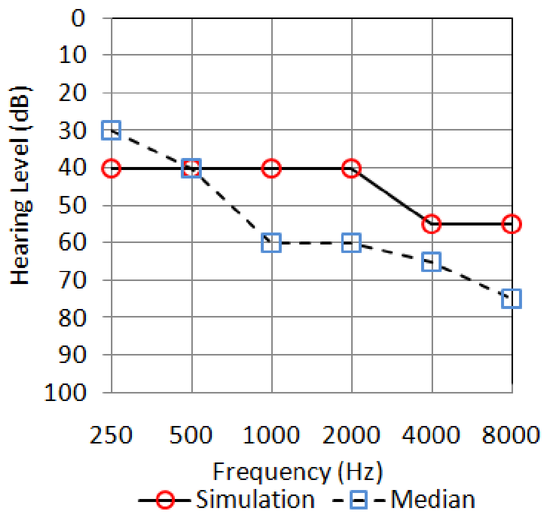


Figure 5.2: The median hearing thresholds (blue rectangle) of the 9 HI subjects in the experiment in Chapter 4 and the assumed hearing thresholds (red circle) used in the HLS model.

Figure 5.2 shows the median hearing thresholds (blue rectangle) of the 9 HI subjects in the experiment (Chapter 4) and the assumed hearing thresholds (red circle) used in the HLS model. There is a difference between the assumed hearing thresholds and the median hearing thresholds of the 9 HI participants. As there was a large variation in the hearing thresholds across the 9 HI subjects, it was not necessary to make the assumed hearing thresholds to be exactly the same as the median measured hearing thresholds. The HLS model in Hu *et al.* (2011b) has been validated in predicting speech intelligibility of speech corrupted with babble noise or speech shaped noise. The same HLS model with the same assumed hearing threshold as in Hu *et al.* (2011b) can be further validated in this chapter to predict speech intelligibility with noise reduction algorithms. The variability in audiogram will cause some variability in the SRT in the HI group. The matched-pair design would eliminate this source of variability and can be conducted in future research.

5.2.1 **Objective demonstration of stimuli**

Figure 5.3 and Figure 5.4 show the time domain waveforms and spectrograms of an example BKB sentence "*she drinks from the cup*" under 0 dB input SNR (in the speech shaped noise and babble noise respectively) processed with different noise reduction strategies and then with the NAL-R procedure and the HLS model. Clean speech and noisy speech processed with the NAL-R procedure and the HLS model are also shown. The waveforms and spectrograms are (a) original speech with the NAL-R procedure and the HLS model; (b) noisy speech with the NAL-R procedure and the HLS model; (c) the same noisy speech processed with CS-WF and then processed with NAL-R procedure and the HLS model; (d) the same noisy speech processed with SCS and then processed with NAL-R procedure and the HLS model. Compared to Figure 3.8 and Figure 3.9, Figure 5.3 and Figure 5.4 add the effects of NAL-R procedure respectively to simulate the combined strategy processing in hearing aids. By visual comparison and listening in figures (b) to (d), it can be seen that the noise in the noisy speech (Figure 5.3 (b)) is further smeared with the HLS model and noise reduction strategies (Figure 5.3 (c-d)) can reduce the noise smearing effects by the HLS model. In speech shaped noise (Figure 5.3), the noise reduction effects increase in the order of the noisy speech, CS-WF and SCS. In babble noise (Figure 5.4), there is no such distinct difference between any pair of noisy speech, CS-WF and SCS, as Figure 5.3. Through informal listening, the difference in speech quality ('clarity', 'comfort') among noisy speech, CS-WF and SCS can be hardly detected. However, the difference in 'noise loudness' can be obviously perceived. This corresponds to the experimental results with the HI subjects (Chapter 4) who could not sensitively detect speech distortion due to hearing loss factors (e. g. reduced frequency resolution), but could sensitively detect noise loudness.

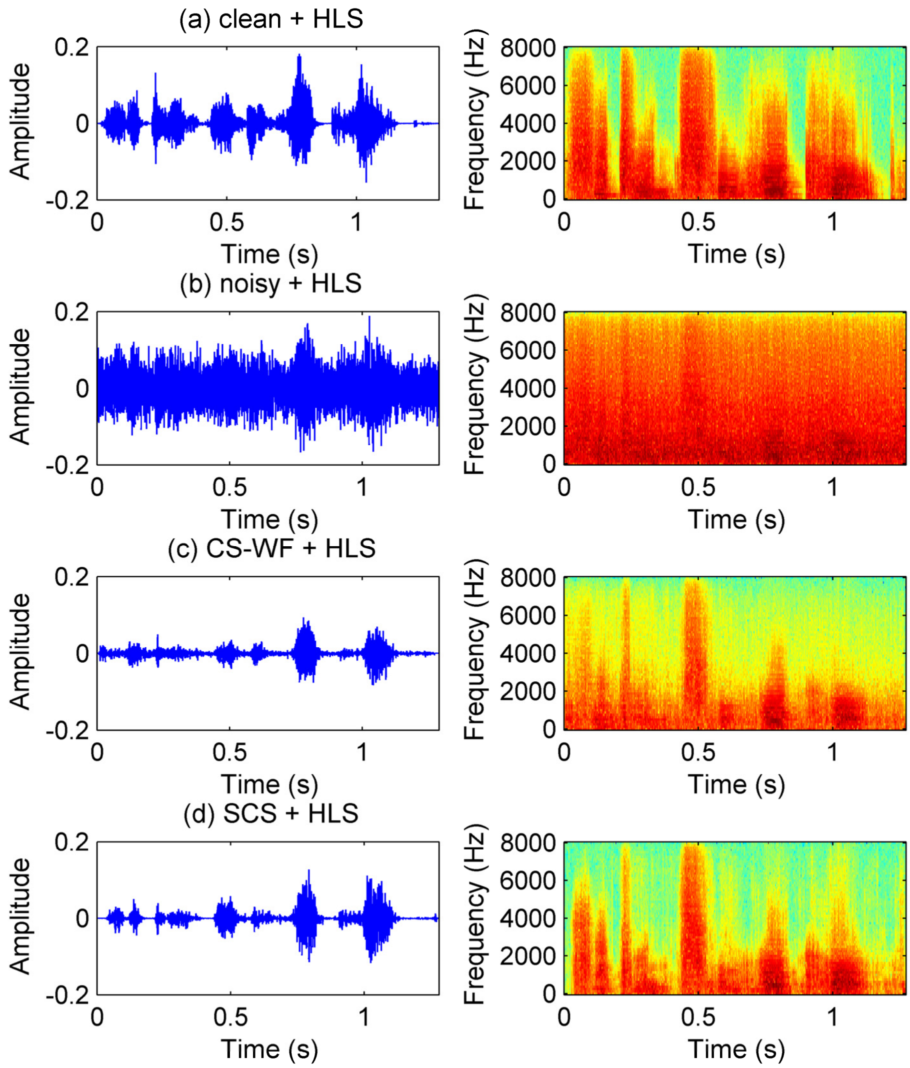


Figure 5.3: Demonstration of unprocessed/processed speech in stationary noise with the HLS model. The BKB sentence is "*she drinks from the cup".* The spectrograms and waveforms are shown for (a) original speech with NAL-R procedure and the HLS model; (b) speech in speech shaped noise (0 dB input SNR) with NAL-R procedure and the HLS model; (c) the same noisy speech processed with CS-WF and then with NAL-R procedure and the hearing loss simulation; (d) the same noisy speech processed with SCS and then with NAL-R procedure and the hearing loss simulation.

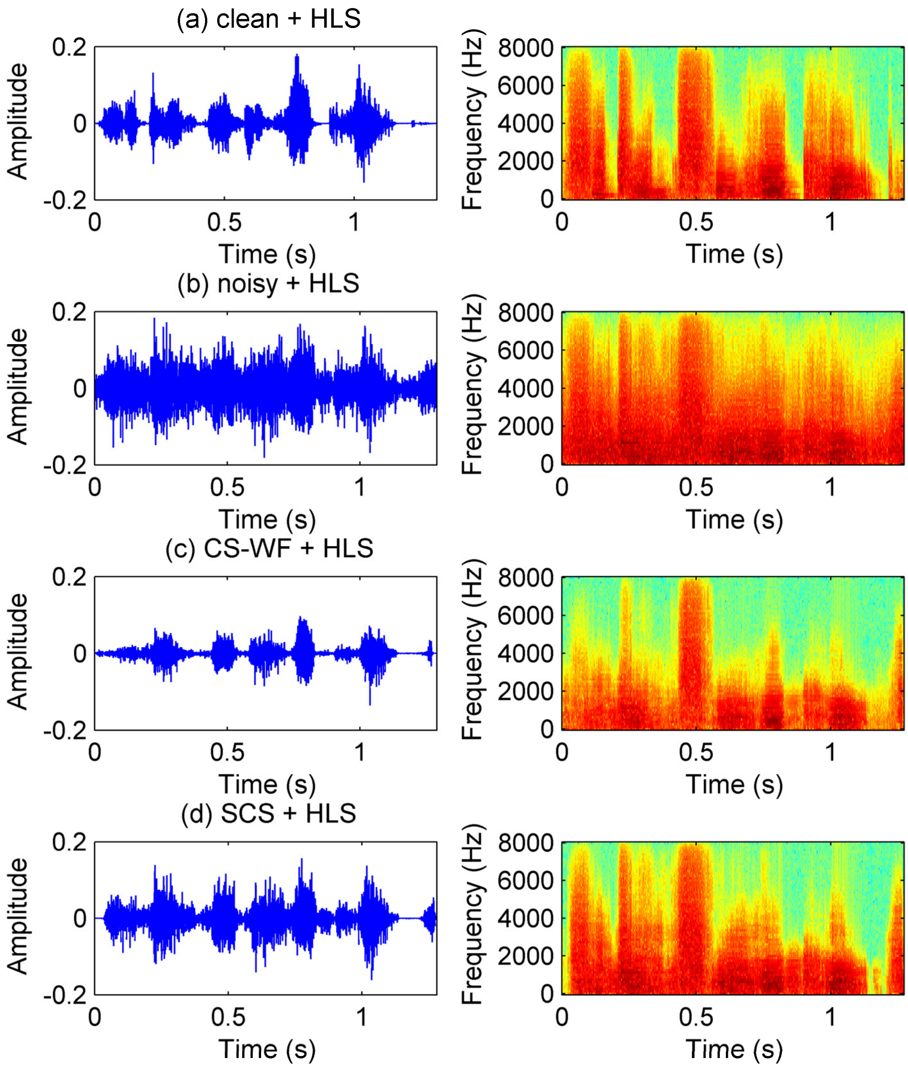


Figure 5.4: Demonstration of unprocessed/processed speech in babble noise with the HLS model. The BKB sentence is "*she drinks from the cup*". The spectrograms and waveforms are shown for (a) original speech with NAL-R procedure and the hearing loss simulation; (b) speech in babble noise (0 dB input SNR) with NAL-R procedure and the HLS model; (c) the same noisy speech processed with CS-WF and then with NAL-R procedure and the HLS model; (d) the same noisy speech processed with SCS and then with NAL-R procedure and the hearing loss simulation.

5.3 Participants

Nine NH listeners who are the same NH participants in experiment in Chapter 4 participated in this experiment. All subjects were native English speakers. The NH listeners had hearing thresholds at or below 20 dB HL from 250 Hz to 8 kHz (confirmed by PTA), and their ages ranged from 20 to 36.

5.4 Equipment

All NH listeners were seated in a sound-isolated room and listened to the sounds presented through Sennheiser HDA 200 headphones presented through a Behringer UCA202 sound card and Creek OBH- 21SE headphone amplifier. The presentation levels of speech were kept at 65 dB SPL.

5.5 Procedure

The intelligibility test is performed through the off-line adaptive speech recognition tests as described in Section 4.5. A total of six conditions were tested: two noise types (SSN, babble) multiplied by three noise reduction conditions (‘noisy’, SCS, WF). All the six conditions are processed with the NAL-R procedure and the HLS model as shown in Figure 5.1. Subjects were instructed to repeat as many words as they could after listening to each sentence with no feedback given during the tests. For familiarization, participants practised the procedure with one randomly selected condition. The order of the six conditions was balanced among the listeners using a latin square. This experiment took each subject no more than one hour.

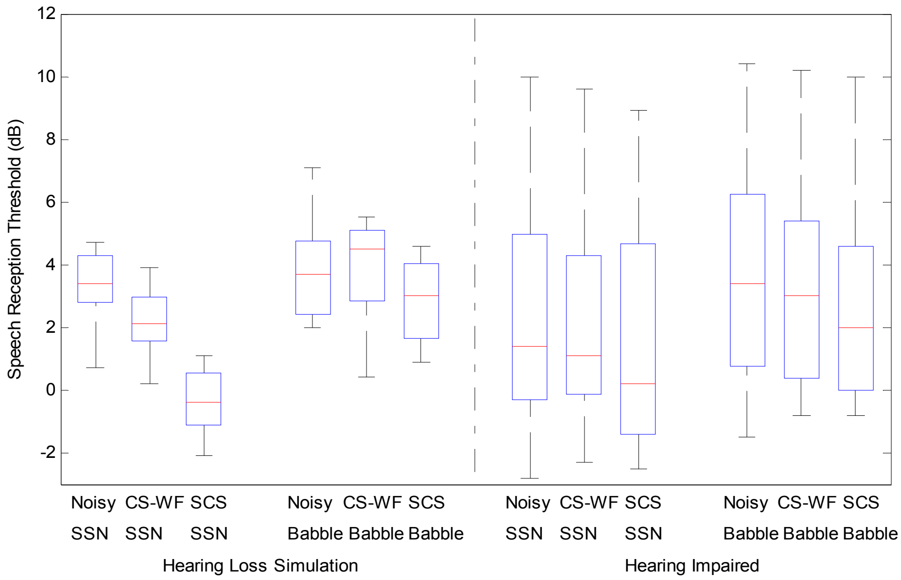
The three-up-one-down adaptive procedure was also used to find the speech-to-noise ratio required for 79.4% correct recognition in each condition (Dahlquist et al., 2005), which is defined as speech reception threshold (SRT) in dB. A sentence was deemed to have been recognised correctly when at least two keywords were repeated correctly. Each sentence has three or four keywords which were fixed in the BKB database. For example, the key words in the sentence “The car engine is running” are “car”, “engine” and “running”.

5.6 Statistical analysis

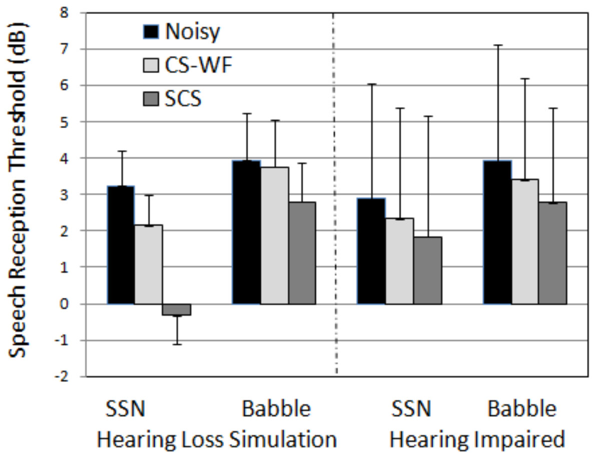
As in Chapter 4, a two-way repeated ANOVA was performed on all of the SRTs across NH subjects. The Fisher least-significant-difference (LSD) test was performed to detect the difference between any pair of the six conditions.

5.7 Results

Figure 5.5 shows the speech recognition performance of all participants in all six test conditions: SSN-Noisy, SSN-CS-WF, SSN-SCS, Babble-Noisy, Babble-CS-WF, Babble-SCS, with the HLS model by 9 NH subjects (left) and with 9 HI subjects (right). Figure 5.5 (a) shows the spread of SRTs in boxplots. Figure 5.5 (b) shows the mean STR with 95% confidence range in each condition. The left panels in Figure 5.5 (a & b) show the results of the test with the combination of HLS model and NH subjects. The right panels in Figure 5.5 (a & b) show the test results with HI subjects (already introduced in Chapter 4, exactly the same as the right panels in Figure 4.4). The test with the HLS model and NH subjects is to mimic the effects of the test with HI subjects. The results of the two tests are presented together to illustrate where both results share any similarity. In Figure 5.5 (a), the SRTs in the right panel show larger variability than in the left panel for each condition. The large inter-subject variability in HI listeners is presumably due to individually different auditory deficits and individual experience with hearing aids. The hearing loss simulation tests simulated the same hearing loss level (shown in Figure 2.3) with all the NH subjects so that this test results in smaller inter-subject variability. In Figure 5.5 (b), by comparing mean SRTs within the left panel and within the right panel, the order of algorithm performance in each noise is similar (SCS > CS-WF > Noisy). It is also to note that SCS performs much better than CS-WF and Noisy speech in speech shaped noise in the HLS simulation test. Further investigation will be performed through statistics.



(a)



(b)

Figure 5.5: SRTs in different conditions with the hearing loss simulation model (left) and with HI subjects (right). SSN: speech shaped noise; Noisy: noisy speech without noise reduction algorithms; WF: Wiener filtering; SCS: sparse coding shrinkage. (a) Boxplots of SRTs. On each box, the central mark is the median; the edges of the box are the 25th and 75th percentiles; the whiskers extend to the most extreme measured data not considered outliers. (b) Mean SRTs with error bars indicate the 95% confidence intervals of the means. A more negative SRT corresponds to better performance.

In the hearing loss simulation test (left panel), the mean SRT for noisy speech with SSN is 3.24 dB, for CS-WF with SSN is 2.14 dB, for SCS with SSN is −0.31 dB, for noisy speech with babble noise is 3.93 dB, for CS-WF with babble noise is 3.74 dB and for SCS with babble noise is 2.8 dB. The results were found to be normally distributed by the Shapiro-wilk test. A two-way repeated ANOVA shows that in the HLS test with NH subjects, the effects of noise reduction algorithm and noise type are both significant [$F(2,16)=28.311$, $p<0.05$, $F(1,8)=27.07$, $p<0.05$]. There is also significant interaction between noise type and noise reduction algorithm [$F(2,16)=9.055$, $p<0.05$]. The Fisher LSD post hoc test was used across NH subjects to detect the difference between any pair of the six conditions (Table 5.1). Significant effects ($p < 0.05$) are given in boldface. In speech shaped noise, there is a significant difference between any pair of the three conditions ((1-3) Table 5.1). This indicates that noise reduction algorithms are significantly better at improving speech intelligibility than unprocessed speech in speech shaped noise in simulation tests which is in accordance with the test by HI subjects in Table 4.4. In Table 5.1, in simulation test, SCS performs significantly better than CS-WF in speech shaped noise which is not the case with HI subjects in Table 4.4. In babble noise, there is no significant effect between any pair of the three conditions ((4-6) Table 5.1). This implies that noise reduction algorithms are not significantly different from unprocessed speech in speech recognition in babble noise with the HLS model which is also in accordance with the test by HI subjects in Table 4.4. On the whole, the HLS model with NH subjects can predict the effects of noise reduction algorithms for HI subjects in speech shaped noise and babble noise. We don’t expect them to be correlated very tightly, as the HLS model only simulates one typical hearing loss level while the HI subjects show hearing loss simulation levels with large inter-subject variance. The significant difference between SCS and CS-WF in speech shaped noise in the simulation test rather than in the HI subjects indicates the HLS model overestimated the effects of SCS.

Table 5.1: Fisher LSD post hoc significant tests for the interaction of noise reduction algorithm and noise type in the HLS experiment with NH subjects. Significant effects ($p < 0.05$) are given in boldface. The number in each bracket is to give each processing condition an identification number.

Processing condition	Mean SRT	(1)	(2)	(3)	(4)	(5)
SSN-Noisy	(1) 3.244					
SSN-CS-WF	(2) 2.144	**0.021**				
SSN-SCS	(3) −0.311	**0.000**	**0.000**			
Babble-Noisy	(4) 3.933	0.102	**0.010**	**0.000**		
Babble-CS-WF	(5) 3.744	0.366	**0.016**	**0.000**	0.683	
Babble-SCS	(6) 2.800	0.400	**0.112**	**0.000**	0.085	0.074

5.8 Discussion

The final evaluation of noise reduction effects in HI listeners is due to subjective evaluation tests through HI subjects. However, it is difficult to recruit HI subjects and performing such tests is expensive. Objective measures are quick and cheap for initial evaluation during the strategy development stage. Current objective measures for HI listeners are rare and are short of validation with different signal processing strategies. In this chapter, we proposed an evaluation methodology to evaluate intelligibility effects of noise reduction algorithms for HI listeners.

This evaluation methodology is based on a physiologically-inspired hearing loss simulation model which has been explained in Chapter 2 and in Hu *et al.* (2011b). The assumption is that consequences of hearing impaired listening can be an additive combination of hearing impaired distortion with normal hearing listening. If we could simulate hearing loss distortion in an appropriate and realistic way, the effects of listening to speech with HI listeners can be approximated by asking NH listeners to listen to the same speech with the hearing loss simulation. Therefore, this proposed evaluation methodology is to perform a subjective speech recognition test by asking NH listeners to listen to speech (noisy, CS-WF, SCS) processed with the HLS model. The same physiologically-inspired hearing loss simulation model was validated with noisy speech (speech corrupted with babble noise or speech shaped noise) (Hu et al., 2011b), but not with noise reduction algorithms before this experiment, in respect of speech intelligibility.

Both the test with HLS model and the test with HI subjects showed that there is a significant effect of noise reduction algorithm in speech shaped noise, but no significant effect of noise reduction algorithm in babble noise. However, the effects of difference in Table 5.1 and Table 4.5 suggest that there might be some significant advantage of SCS in babble noise for both HLS model and HI subjects (p is near 0.05). The effect can be clear if more subjects are recruited. Currently because the powers of the results with HLS model and HI subjects are low, it is not clear if there is any trend. The range of mean SRTs between the HLS model and HI subjects are similar. Therefore, the HLS model can generally predict the noise reduction effects for HI subjects in speech intelligibility. However, several differences still exist between the two tests. One difference is that the test with the HLS model showed much less inter-subject variance than the test with HI subjects. All NH listeners listened to speech with the same hearing loss simulation level and thus show less diverse results compared to HI listeners, who showed variant hearing loss levels. Another difference is that SCS can perform significantly better than CS-WF in speech shaped noise with HLS model but not with HI subjects.

Through further investigation of the parameters in the HLS model, a correction parameter was overestimated. This correction is the difference between the calculated level (dB) in MATLAB representations and the estimated hearing level (dB) of the same signal. The MATLAB representation quantities are normalized in amplitude within the range [-1 1]. The correction parameter can be called the calibration parameter which calibrates the hearing level of the signal from its MATLAB representation quantity. Figure 5.6 shows the distortion by the threshold elevation and loudness recruitment of a signal in the quantity of hearing level (dB). Before HLS processing, the calculated signal level in MATLAB is added with the correction as the input hearing level L_{in}. After processing in Figure 5.6, the output hearing level L_{out} is deduced with the same correction as the output signal level in MATLAB. When the correction is overestimated, the input hearing level L_{in} is increased to L_{in}'. Accordingly, the difference d between the output hearing level and the input hearing level is underestimated as d'. Figure 5.6 shows the consequence of overestimating the input hearing level of envelope with the same function as shown in Figure 2.3. Therefore, when the output hearing level is corrected back to the output signal level in MATLAB, the output signal is actually deduced by d' rather than d in MATLAB. That is, the output signal is overestimated by (d-d').

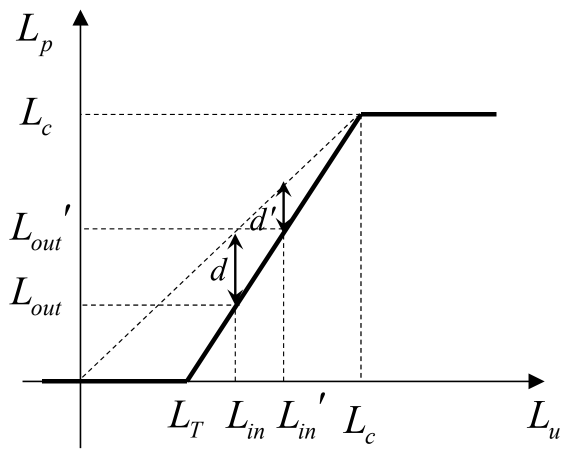


Figure 5.6: Illustration of effects when overestimating the input hearing level L_{in}.

The exact correction parameter in each filter band should be calculated as follows. We calibrated the difference between the sound pressure level in TDH 39 headphone and the level calculated in MATLAB as $\vartriangle L_1$. We can also acquire the difference between the sound pressure level in TDH 39 headphone and the corresponding hearing level from ISO 389-1 as $\vartriangle L_2$, also called reference equivalent threshold sound pressure level (RETSPL) (ISO389-1, 2000). The exact correction parameter is $\vartriangle L_1$ minus $\vartriangle L_2$. The correction parameter in each filter band should be different and can be interpolated. However, our experiment roughly set the correction parameter the same across the filters and this value is around 20 dB higher than the actual correction parameter.

The overestimation of output envelopes (Figure 5.6) as well as frequency smearing (Figure 2.1) will amplify the detrimental effects of additive noise where local SNR is lower than 0 dB in the spectrogram. As speech shaped noise is stationary and spreading the whole spectrogram, speech corrupted with speech shaped noise can show more noise detrimental effects (Figure 5.2 (b)). This may explain why this HLS model has more detrimental effects in noisy speech (Figure 5.2 (b)) rather than in clean speech or enhanced speech (Figure 5.2 (a), (c), (d)). As the babble noise is fluctuating with gaps in the spectrogram, the noise detrimental effect with the HLS model in babble noise (Figure 5.3 (b)) is not as much as that in stationary noise (Figure 5.2 (b)).

The HLS model may amplify the smearing effects of remaining noise in the speech. When the correction parameter is appropriate, the HLS model can be more realistic. The estimation of the correction parameter in the HLS model has been explained in Section 2.2. However, the HLS model with the overestimated correction parameter can differentiate obviously the noise reduction effects between the algorithms. According to the subjective performance in Table 5.1, we can infer that SCS reduced much more speech shaped noise but similar amounts of babble noise compared to CS-WF (compare conditions (2) and (3) or conditions (5) and (6) in Table 5.1).

We still believe the HLS model holds promise to predict effects of noise reduction algorithms relative to unprocessed speech in speech intelligibility for HI subjects to some degree. The HLS model can be made more realistic with appropriate correction parameters.

5.9 Conclusions

Our research is the first to validate the hearing loss simulation model in evaluating effects of noise reduction algorithms for HI listeners. Our study has shown that some of the effects of the noise reduction algorithms seen in HI listeners are reproduced, at least qualitatively, by using the HLS model with NH listeners. If the correction parameters are adjusted, the HLS model might be more realistic. There are several advantages of this evaluation methodology. Firstly, it takes less time and effort to recruit NH listeners than HI listeners. Secondly, by asking NH listeners to listen to speech with the same hearing loss simulation level, this experiment can show intelligibility results with less variance and thus show some factor effect with more power.

Further research can be conducted to test whether there is any correlation in SRT between HI subjects and the HLS model in each condition. This experiment needs a number of HI subjects with similar hearing loss thresholds and the same number of NH subjects with the same hearing loss thresholds as HI subjects.

Most existent intelligibility measures for HI listeners are only validated with clean speech or noisy speech rather than with noise reduction algorithms. A future perspective is to set up a purely objective speech intelligibility measure that includes the auditory filter based HLS model.

Chapter 6 General discussion

The evaluation results in Chapter 4 provide answers to a number of important questions about the benefits of single-channel noise reduction algorithms in hearing aids. One confirmed answer is that single-channel noise reduction algorithms could help hearing impaired listeners more than normal hearing listeners. However, there is no simple way to determine the magnitude of the benefits to hearing aid users. Specific factors include the hearing loss level, the input SNR, the noise reduction algorithm itself, the effect of noise type, etc. Specific benefits include the domain of intelligibility or quality. This chapter will discuss the factors that determine the performance and how these affect performance.

In terms of evaluation methodology for HI subjects, tests on real users are difficult and more time consuming than objective measures. However, currently, there are very few validated objective measures to evaluate noise reduction algorithms for HI listeners. Chapter 2 & 5 respectively proposed and validated an evaluation methodology that is a combination of subjective tests on NH subjects and a novel HLS model. Assuming that the HLS model is realistic, this evaluation avoided the difficulty of recruiting HI subjects and predicted reliable subjective results for a specific hearing loss level. Development of the HLS model in Chapter 2 was prioritised before algorithm development in Chapter 3, as understanding hearing loss factors through the HLS model can have implications to development of noise reduction algorithm for HI listeners. Algorithm testing through NH and HI subjects in Chapter 4 was followed by validation of the HLS model in Chapter 5, as the test with the HLS model is currently a new and additional evaluation test that is not as reliable as the tests with HI subjects (Chapter 4).

The discussion in this chapter first considers the factors and issues that are related to the performance of noise reduction algorithms for hearing aid users. This is followed by a discussion of the effectiveness and limitations of the HLS model as an evaluation tool for HI listeners. Then, potential translation of this research to hearing aids is discussed. Finally, limitations of the study and suggestions for future work are given.

6.1 Related factors and effects

6.1.1 The effect of hearing loss level

The two noise reduction algorithms (CS-WF and SCS) evaluated in our study showed comparatively more benefits to HI listeners than to NH listeners, both in speech intelligibility and speech quality. This is in accordance with previous evaluation that most single-channel noise reduction algorithms could not improve speech intelligibility for NH listeners, but could show significant intelligibility improvement for CI users who are profoundly hearing impaired listeners. This indicates that the benefits of noise reduction algorithms might vary with the hearing loss level. The hearing threshold is one basic characteristic of the hearing loss level. The quantitative linear relationship between the speech recognition performance and the hearing threshold across the 18 subjects in Figure 4.7 also supported this hypothesis. However, when including the hearing threshold and SRT of noisy speech both into the regression model of speech recognition gain, the contribution of the hearing threshold is excluded (Table 4.7, Table 4.8). Therefore, SRT of noisy speech is a more critical factor that determines the performance of noise reduction algorithms. This will be explained in the next section.

The study showed that the two noise reduction algorithms improved speech quality more for HI listeners than for NH listeners. HI listeners, especially with severe or profound hearing losses, are very sensitive to noise loudness, but less sensitive to speech distortion compared to NH listeners. This is in accordance with previous research that spectral distortion in speech can be easily detected by NH subjects, but not by HI subjects (Schijndel et al., 2001). Single-channel noise reduction algorithms usually reduce noise by a larger amount at the expense of greater speech distortion. Therefore, single-channel noise reduction algorithms can use more aggressive noise reduction methods such as SCS to suit the listening properties of the

HI listeners.

6.1.2 **The effect of input SNR**

Table 4.7 and Table 4.8 show that the SRT of noisy speech is a more important factor in the linear regression of the speech recognition gain compared to the hearing loss level. The SRT of noisy speech indicates the level of input SNR in individual evaluation. This reflects that input SNR is a more important factor than the hearing loss level in determining the noise reduction performance. Noise reduction algorithms perform better at higher SNRs. Noise estimation is better at higher input SNRs where it is easier to differentiate speech and noise segments. If NH subjects can be tested at higher input SNRs using difficult speech, they might also get as much speech recognition gain as HI subjects did. However, usually NH subjects can reach ceiling performance at higher input SNRs that do not need noise reduction algorithms. The variation of speech recognition gain with input SNR of noisy speech is more obvious in babble noise than in speech shaped noise. This reflects the difficulty of reducing babble noise at low input SNRs.

This also suggests that the choice of input SNR during evaluation of noise reduction algorithms affects the results, which is especially important for non-adaptive speech recognition tests that choose fixed input SNRs (usually 0, 5, 10 dB).

6.1.3 **The effect of noise reduction algorithm**

The evaluation in Chapter 4 showed no significant difference between the two noise reduction algorithms for either NH subjects or HI subjects. This indicated that the performance of the two noise reduction algorithms were similar in level given the limited number of subjects. However, with the HLS model shown in Chapter 5, SCS showed significantly better performance than CS-WF in speech shaped noise. The HLS model may amplify the benefits of noise reduction effects. Due to the limited power of this study, it is still not certain whether

SCS performs better than CS-WF in some type of noise.

6.1.4 **The effect of noise type**

Speech shaped noise and multi-talker babble noise, both of which show a similar average long term spectrum as the speech, were adopted as additive noise in the evaluation. Speech shaped noise is stationary and babble noise is fluctuating. The two noise reduction algorithms showed better performance for stationary noise than for non-stationary noise in both NH and HI subjects. For NH listeners, noise reduction algorithms did not affect speech intelligibility in speech shaped noise but deteriorated speech intelligibility in babble noise. For HI listeners, noise reduction algorithms showed significant intelligibility improvements in speech shaped noise but not in babble noise. The effect of noise type was also shown in CI users. Yang and Fu (2005) found that the same spectral subtraction algorithm worked much better in speech shaped noise than in babble noise in CI users. Due to the rapid varying characteristics of babble noise, the noise power estimation methods can not accurately track and estimate the noise power in each temporal frame. Babble noise, which shows similar fluctuating properties as the target speech, can be easily misidentified as speech resulting in weak noise reduction effects.

6.1.5 **Benefits in intelligibility or quality**

Intelligibility and quality are two properties typically tested for noise reduction algorithms. These two noise reduction algorithms bring more benefits in quality than in intelligibility for HI subjects (Figure 4.6). This is in accordance with previous findings that noise reduction algorithms can improve speech quality but not necessarily improve speech intelligibility (Harlander et al., 2012). Noise reduction algorithms have reached limits in terms of intelligibility, especially for those who show mild-to-moderate hearing losses and are tested at the limit of their performance (typically at input SNR below 5 dB). However, it is worth noting that quality is evaluated at more positive input SNRs (5 dB, 10 dB) than intelligibility (Figure 4.5). It must be recognized that the benefit of quality and intelligibility are measured at different input SNRs, which means that the algorithms are performing differently in the two types of comparison. This difference in SNR may be the main reason why quality benefits are easier to demonstrate.

6.2 The evaluation methodology with the HLS model

It has been assumed that consequences of hearing impaired listening can be a combination of hearing impaired distortion and normal hearing listening, and that the hearing impaired distortion can be simulated with the HLS model. If this assumption is correct, evaluation of noise reduction algorithms for HI listeners can be predicted by evaluation with a HLS model through NH listeners. Chapter 5 tested this assumption, and showed that the HLS model can predict some intelligibility effects of noise reduction algorithms relative to baseline performance for HI subjects. The advantage of the physiologically inspired hearing loss simulation model is that it adopts the nonlinear and asymmetric auditory filter bank, which can potentially approximate the nonlinear auditory filtering in the basilar membrane.

NH listeners all listen to speech that is processed with the same HLS level while HI listeners show considerable differences in hearing loss levels. Therefore, the results obtained using the HLS model showed less variation than results with HI subjects. The corresponding performance across NH listeners with the HLS model can predict the result for a specific hearing loss level more easily. However, the HLS model overestimated the correction parameter and thus overestimated the output hearing level of the noisy signal (Figure 5.6). This resulted in amplified detrimental effects of additive noise where local SNR is lower than 0 dB in the spectrogram (Section 5.8). That is, the HLS model amplified the benefits of noise reduction effects by exaggerating the detrimental effects of remaining noise in speech. With appropriate choice of the correction parameter in the HLS model, which calibrates the difference between the hearing level and the MATLAB quantities (dB) of the same signal, the HLS model may be more realistic in predicting noise reduction performance for HI subjects. This HLS model also has the potential to be added into a purely objective intelligibility metric for HI listeners.

6.3 Limitations of the study and future work

6.3.1 **Practical issues in hearing aids**

Essentially all commercial hearing aids are equipped with a spatial filtering algorithm (beam-forming) and a single-channel noise reduction algorithm except in-the-ear hearing aids that mostly rely only on single-channel noise reduction algorithms. Final tests of single-channel noise reduction algorithms in hearing aids should be combined with a spatial filtering algorithm.

In the present test, the noise reduction algorithms were combined with the NAL-R formula to mimic the threshold compensation in hearing aids. However, NAL-R can only compensate sound with linear amplification. If there is a large range in amplitude across the sound, soft sounds may not be amplified enough to be audible, while other sounds may become painfully loud. Therefore, advanced hearing aids usually are equipped with automatic gain control that can better mimic the compensation strategies in hearing aids. Therefore in laboratory testing it is better to add noise reduction algorithms with dynamic range compression to test performance in HI subjects.

6.3.2 **Subjects, test materials and test types**

This study only tested nine HI subjects due to the difficulty of recruiting hearing aid users. To make a balanced comparison, this study also tested nine NH subjects. Due to the limited power of the study with only 18 subjects, the effects of noise reduction algorithms may not be thoroughly unearthed. Although some effect has already been shown, statistical power would increase using more subjects.

Input SNR seems to be a very important parameter. Further studies can be conducted to pin down this effect more precisely. For example, using more difficult speech materials to change the input SNR within the same subject; or using more severely hearing impaired subjects to present them with noisy speech at higher input SNRs.

Young HI subjects may perform better than older subjects due to plasticity in speech adaptation (Robinson, 1998). Therefore, an equal number of young and older subjects should be recruited in future evaluation to clarify the effect of age.

As well as the babble noise and speech shaped noise, more noise types can be added to simulate different scenarios. It is worth noting that there is a corpus that collected noise from a real family living room[4]. This would mimic real life background noise more realistically.

Words or syllables can also be used to evaluate the effect of noise reduction algorithms. For example, the vowel-consonant-vowel (VCV) test can help to explore whether the noise reduction algorithm help to understand vowels or consonants specifically.

As the algorithms might show different acclimatisation effects, the performance of algorithms in the experiment might not represent the long-term benefits of algorithms. Clinical trials of algorithms embedded in a wearable device are worthy in future research. This also demands the feasibility and real-time implementation of algorithms in a hearing aid.

6.3.3 **Future research in sparse coding for speech enhancement**

This study applied the sparse coding shrinkage principle to speech enhancement for HI subjects. No clear difference in performance of SCS and CS-WF was seen. However, the method of extracting key information from noisy speech is not limited to the proposed strategy. Previous noise reduction algorithms that assume super-Gaussian distribution of speech were also developed to extract key speech information. Furthermore, with the development of

[4] http://spandh.dcs.shef.ac.uk/projects/chime/challenge.html

various sparse coding strategies in the machine learning society, there are more opportunities to extract key speech information for hearing aid users through different ways. While training to extract statistical features of speech or noise might improve the algorithms, this is not very practical due to variation in conversational speech and environmental noise. The proposed sparse coding shrinkage method in speech enhancement only approximately estimated the sparse parameters, so it is better if more accurate sparse parameters can be estimated online without *a priori* knowledge of speech or noise. Performing sparse coding in the spectral envelop domain (Sang et al., 2011b) is also promising, as the envelop carries plenty of information for speech intelligibility, which has already been shown in the cochlear implant processing.

Previous machine learning algorithms concentrate on text or image signals and have seldom considered speech signals. There is one state-of-the-art sparse coding dictionary learning algorithm that sets out to find sparse atoms for speech signals (Jafari and Plumbley, 2011). Future work can be conducted to develop speech enhancement algorithms with such sparse coding algorithms that focus on sparse decomposition of speech signals in the background of different noises.

Sparse coding has been studied in the neural science (Olshausen and Field, 1996, Rieke et al., 1999, Dayan and Abbott, 2001). If we regard our brain as an intelligibility system with sparse coding mechanisms, the findings from neurophysiology can inspire improvements in sparse coding strategies in the signal processing community.

6.3.4 **Future research in the HLS model**

There is still a minor difference between the test with the HLS model and the test with the HI subjects. It is found that due to the overestimation of the correction parameter in the HLS model, the noise smearing effects are exaggerated in the HLS model. The correction parameter is the difference in dB between the signal level calculated in MATLAB and the corresponding hearing level. In other words, the advantage of algorithms with stronger noise reduction effects is overestimated. As SCS had more noise reduction effects than CS-WF and unprocessed speech in stationary noise, the HLS model predicted a significant advantage of SCS in speech shaped noise for HI subjects. This model can be improved by adjusting the correction parameter. Current study only evaluated the effects of noise reduction algorithms relative to baseline performance between HI subjects and the HLS model. Further research can be conducted to evaluate the correlation in SRT of each enhanced/ unprocessed condition between HI subjects and the HLS model.

Chapter 7 Conclusion

This work focuses on the benefits of single-channel noise reduction algorithms in hearing aids. This work contains development and evaluation of the sparse coding shrinkage (SCS) noise reduction algorithm. The motivation is to evaluate whether the SCS algorithm might benefit HI subjects more than a state-of-the-art competitive noise reduction algorithm (CS-WF). Subjective evaluations were performed with NH and HI subjects in speech intelligibility and quality. An additional evaluation of effects of noise reduction algorithms for HI subjects was performed using the HLS model and NH subjects. The results of this work are about the evaluation and justification of effects of noise reduction algorithms on HI listeners. In particular, this work supports the following conclusions:

- Although SCS seems to perform better than CS-WF in terms of intelligibility in stationary noise (Section 3.4 & Section 5.8), differences between the algorithms were not large enough to be statistically significant for the overall samples of subjects tested here. A more powerful study is required to determine definitely whether SCS outperforms CS-WF.
- Both the algorithms can improve speech quality which is in accordance with previous evaluations (Section 4.7.2).
- The difference between NH and HI subjects in intelligibility gain depends primarily on the input SNR rather than the hearing loss level although the two variables are correlated. The input SNR was set adaptively according to the individual hearing loss level. The algorithms performed better at higher input SNRs where HI subjects can get benefits but NH subjects have reached ceiling performance (Section 6.1.2).
- Babble noise is still a challenge for the noise reduction algorithms used here, especially at low input SNRs and for listeners with mild-to-moderate hearing losses (Section 6.1.2 & Section 6.1.4).
- The HLS model may predict the effects of noise reduction algorithms relative to baseline performance for HI subjects using NH subjects, and it can probably be

improved by adjusting the correction parameter in the model (Chapter 5).

Appendix A: Publications

Sang, J., Hu, H., Zheng, C., Li, G., Lutman, M. E., and Bleeck, S. "Evaluation of sparse coding shrinkage noise reduction algorithm in normal hearing and hearing impaired listeners," in Eusipco, Bucharest, Romania, 2012.

Sang, J., Hu, H., Li, G., Lutman, M. E., and Bleeck, S. "Supervised sparse coding strategy in hearing aids," in IEEE International Conference on Communication Technology (ICCT), China, 2011. *Best paper award*.

Sang, J., Li, G., Hu, H., Lutman, M. E., and Bleeck, S. "Supervised sparse coding strategy in cochlear implants," in Interspeech, Florence, Italy, 2011.

Sang, J., Hu, H., Winter, I., Wright, M., and Bleeck, S. "The 'Neural Space': a physiologically inspired noise reduction strategy based on fractional derivatives," in IEEE International Symposium on Communications and Information Technologies (ISCIT), China, 2011.

Hu, H., Sang, J., and Lutman, M. E. "Simulation of hearing loss using compressive gammachirp auditory filters," in ICASSP, Prague, Czech Republic, 2011.

Taghia, J., Taghia, J., Mohammadiha, N., Sang, J., Bouse, V., and Martin, R. "An evaluation of noise power spectral density estimation algorithms in adverse acoustic environments," in ICASSP, Prague, Czech Republic, 2011.

Hu, H., Li, G., Chen, L., Sang, J., Wang, S., Lutman, M. E., and Bleeck, S. "Enhanced sparse speech processing strategy for cochlear implants," in Eusipco, Barcelona, Spain, 2011.

Appendix B: Results of subjective tests

Table B. 1: SRT of HI listeners in adaptive speech recognition tests (dB). The lower the SRT, the better the intelligibility of the condition is.

	Speech shaped noise			Babble noise		
	Noisy	CS-WF	SCS	Noisy	CS-WF	SCS
HI1	0.7	0	-2.5	2	1.9	1.2
HI2	3.5	3.1	2.4	4.9	4	4.4
HI3	0.1	-0.3	-1.3	0	-0.8	-0.8
HI4	3.6	2	3.4	5	4.7	3.1
HI5	9	7.9	8.4	10.4	10.2	10
HI6	-1.5	-2.3	-1.8	-1.5	-0.3	-0.1
HI7	10	9.6	8.9	10	7.4	5.2
HI8	-0.8	-0.1	-1.1	1	0.6	0
HI9	1.4	1.1	0.2	3.4	3	2

Table B. 2: SRT of NH listeners in adaptive speech recognition tests (dB). The lower the SRT, the better the intelligibility of the condition is.

	Speech shaped noise			Babble noise		
	Noisy	CS-WF	SCS	Noisy	CS-WF	SCS
NH1	-3.1	-4.8	-4.1	-2.9	-1.6	-1.5
NH 2	-3.9	-4.9	-4.5	-3.3	-0.9	-1
NH 3	-3.1	-4.9	-3.3	-4.6	-2.1	-2
NH 4	-4.6	-2.4	-5	-2.1	-1.6	-1.9
NH 5	-2.3	-3.6	-2.9	-3.3	-1.9	-2.1
NH 6	-2	-2	-3.5	-0.4	-2	-0.4
NH 7	-2.4	-0.9	-1.1	-1.1	-0.9	-1.3
NH 8	-3.6	-3.4	-4	-3.8	-1.1	-3.3
NH 9	-3.8	-2.5	-3.1	-2.1	-1.8	-1.3

Table B. 3: Results of paired comparison rating of 'preference' with HI listeners. The higher the rating, the better the quality of algorithm is. 5/5 means input SNR 5dB for processed speech / 5dB for unprocessed speech. 5/10 means input SNR 5dB for processed speech / 10dB for unprocessed speech.

	CS-WF				SCS			
	SSN		Babble		SSN		Babble	
	5/5	5/10	5/5	5/10	5/5	5/10	5/5	5/10
HI1	6.3	4.8	5.1	2.9	6.0	3.6	5.1	3.8
HI2	1.0	0.5	1.1	-0.3	3.6	2.2	2.9	2.0
HI3	6.9	4.0	3.1	2.0	5.4	5.0	6.0	5.0
HI4	6.8	4.4	4.7	1.2	7.3	6.1	6.7	5.1
HI5	7.9	4.4	5.9	4.4	6.1	4.0	6.1	3.2
HI6	5.1	4.0	4.1	3.8	6.2	5.3	5.0	4.1
HI7	5.2	1.3	2.7	0.0	6.1	2.0	2.7	-0.2
HI8	6.3	1.0	5.5	3.5	7.3	5.0	4.5	1.6
HI9	5.4	3.6	7.1	3.7	7.1	4.6	6.5	4.4

Table B. 4: Results of paired comparison rating of 'preference' with NH listeners. The higher the rating, the better the quality of algorithm is. 5/5 means input SNR 5dB for processed speech / 5dB for unprocessed speech. 5/10 means input SNR 5dB for processed speech / 10dB for unprocessed speech.

	CS-WF				SCS			
	SSN		Babble		SSN		Babble	
	5/5	5/10	5/5	5/10	5/5	5/10	5/5	5/10
NH1	1.9	-2.1	3.4	1.8	1.1	-1.7	3.0	-0.7
NH 2	0.7	0.0	2.2	1.8	1.4	1.1	1.9	0.6
NH 3	3.6	2.4	2.4	1.0	3.9	3.2	0.0	-2.1
NH 4	4.9	3.3	2.5	1.1	6.4	4.6	5.4	3.9
NH 5	3.4	1.1	2.4	0.0	3.8	1.4	2.6	1.9
NH 6	3.3	2.7	3.0	-2.6	2.9	-3.1	3.0	-2.1
NH 7	1.5	0.5	1.0	-1.0	1.5	0.0	0.0	-1.0
NH 8	0.6	-0.3	1.3	0.3	0.6	0.3	-0.4	-1.0
NH 9	2.0	0.3	3.8	1.0	0.8	-3.2	-0.9	-2.3

Table B. 5: Results of paired comparison rating of 'noise loudness' with HI listeners. The higher the rating, the better the quality of algorithm is. 5/5 means input SNR 5dB for processed speech / 5dB for unprocessed speech. 5/10 means input SNR 5dB for processed speech / 10dB for unprocessed speech.

	CS-WF				SCS			
	SSN		Babble		SSN		Babble	
	5/5	5/10	5/5	5/10	5/5	5/10	5/5	5/10
HI1	6.9	4.0	5.1	3.7	6.1	3.9	5.9	4.6
HI2	2.5	1.3	3.5	1.9	4.0	2.4	1.8	1.0
HI3	6.1	4.1	4.9	3.5	6.1	4.1	5.1	3.6
HI4	5.1	1.9	5.3	1.1	5.1	3.4	4.6	2.5
HI5	5.0	2.0	7.1	3.7	5.3	2.7	5.3	2.6
HI6	5.0	3.2	4.7	3.5	5.6	5.1	5.1	4.5
HI7	5.6	3.5	5.2	3.2	7.1	5.0	5.6	3.4
HI8	6.5	2.5	5.5	3.5	7.6	6.1	5.0	3.0
HI9	8.7	6.6	7.3	5.2	8.8	5.9	6.4	5.2

Table B. 6: Results of paired comparison rating of 'noise loudness' with NH listeners. The higher the rating, the better the quality of algorithm is. 5/5 means input SNR 5dB for processed speech / 5dB for unprocessed speech. 5/10 means input SNR 5dB for processed speech / 10dB for unprocessed speech.

	CS-WF				SCS			
	SSN		Babble		SSN		Babble	
	5/5	5/10	5/5	5/10	5/5	5/10	5/5	5/10
NH1	2.6	0.9	3.6	3.0	4.3	1.0	2.8	2.7
NH 2	7.0	4.3	5.0	3.5	6.5	5.5	4.9	3.1
NH 3	4.0	2.7	3.3	1.5	4.7	4.2	3.0	2.0
NH 4	5.1	2.5	2.4	1.2	6.2	5.0	3.9	2.7
NH 5	2.7	1.4	2.7	1.3	3.2	1.7	1.7	0.4
NH 6	4.0	0.8	5.2	2.4	2.2	0.1	2.3	0.3
NH 7	7.5	2.0	6.0	3.0	8.0	4.5	5.0	1.5
NH 8	4.0	1.5	4.0	2.0	5.2	3.5	4.3	3.2
NH 9	3.0	2.4	1.8	1.5	3.0	2.8	3.3	3.1

Table B. 7: Speech recognition gain and SNR gain in quality ‘preference’ within HI subjects in each condition.

	Speech recognition gain				SNR gain in ‘preference’			
	SSN		Babble		SSN		Babble	
	CS-WF	SCS	CS-WF	SCS	CS-WF	SCS	CS-WF	SCS
HI1	0.7	3.2	0.1	0.8	10.0	10.0	9.8	10.0
HI2	0.4	1.1	0.9	0.5	6.5	10.0	3.4	10.0
HI3	0.4	1.4	0.8	0.8	10.0	10.0	10.0	10.0
HI4	1.6	0.2	0.3	1.9	10.0	10.0	7.2	10.0
HI5	1.1	0.6	0.2	0.4	10.0	10.0	10.0	10.0
HI6	0.8	0.3	-1.2	-1.4	10.0	10.0	10.0	10.0
HI7	0.4	1.1	2.6	4.8	6.6	7.3	5.2	4.6
HI8	-0.7	0.3	0.4	1.0	6.1	10.0	10.0	7.2
HI9	0.3	1.2	0.4	1.4	10.0	10.0	10.0	10.0

Table B. 8: Speech recognition gain and SNR gain in quality ‘preference’ within NH subjects in each condition.

	Speech recognition gain				SNR gain in ‘preference’			
	SSN		Babble		SSN		Babble	
	CS-WF	SCS	CS-WF	SCS	CS-WF	SCS	CS-WF	SCS
NH1	1.7	1.0	-1.3	-1.4	2.3	-0.5	9.7	6.0
NH 2	1.0	0.6	-2.4	-2.3	4.7	10.0	10.0	7.4
NH 3	1.8	0.2	-2.5	-2.6	10.0	10.0	8.4	0.1
NH 4	-2.2	0.4	-0.5	-0.2	10.0	10.0	8.8	10.0
NH 5	1.3	0.6	-1.4	-1.2	6.9	9.0	6.4	9.8
NH 6	0.0	1.5	1.6	0.0	10.0	2.4	2.7	3.1
NH 7	-1.5	-1.3	-0.2	0.2	5.9	5.0	2.5	0.1
NH 8	-0.2	0.4	-2.7	-0.5	5.7	8.8	7.3	-1.3
NH 9	-1.3	-0.7	-0.3	-0.8	5.9	1.0	6.9	-3.8

Appendix C: Instruction sheet to test speech quality

Paired comparison rating of speech quality

This experiment asks you to give relative ratings of speech quality through hearing tests using headphones. There are two quality parameters: “Preference” and “Background Noise”. We will explain how to rate the two dimensions separately in the following.

1) Preference.

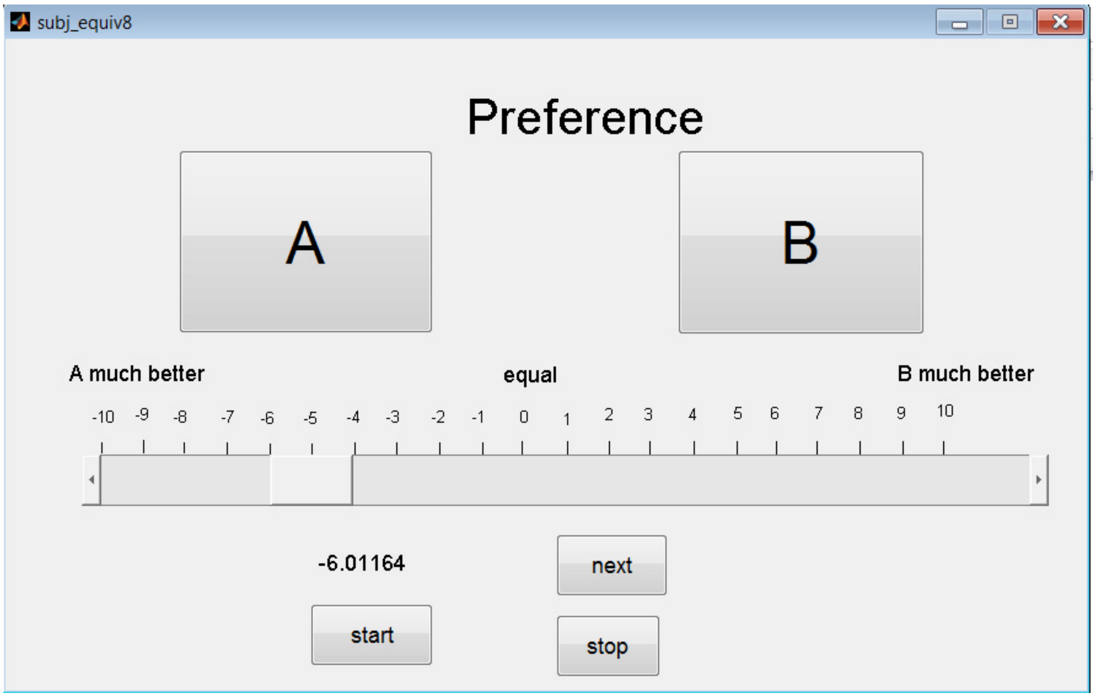


Figure C. 1: User Interface for quality comparison rating of ‘preference’. The quality dimension is “Preference” shown as the title.

The software will automatically show the interface as Figure C. 1on the screen. Please follow the instructions below.

a) First, you will hear the information voice “Preference” that reminds you that the evaluation parameter is preference.

b) You can listen to whichever sentence by clicking button A or B. Please click the button A and button B in turn until you are sure of the quality difference between the two.

c) Please use the slider to give the comparative quality rating between A and B. For example, if you think A is better than B, the slider can be placed in the left side of bar (<0); if you think A is much better than B, the slider can be shifted leftwards (<<0); if you perceives that A equals B in quality, the slider can be placed to the middle of the bar (0). The left side of the slider indicates the rating value that appears below the bar, as shown in Figure C. 1.

d) After you give the rating value using the slider, please click "next" to give another rating with the same procedures as explained in b) and c).

e) If you want a break at any time, please click the button "stop". You can continue the procedure by clicking the button A or B again.

f) Please keep on clicking and give ratings until the title of "Preference" on the screen is replaced by "Finished!".

2) Background Noise

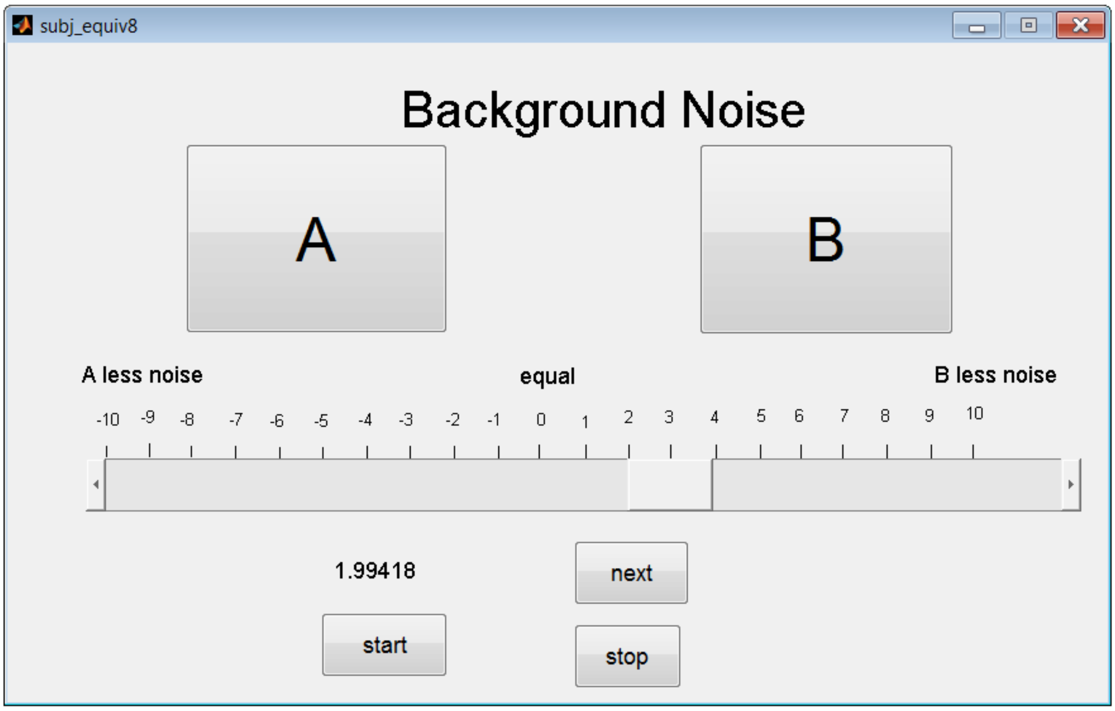


Figure C. 2: User Interface for quality comparison rating of "noise loudness". The quality dimension is "Background Noise" shown as the title.

The software will automatically show Figure C. 2 on the screen. Please follow the instructions below.

a) First, you will hear the information voice “Background Noise” which reminds you that the evaluation parameter is the loudness of background noise.

b) You can listen to whichever sentence by clicking button A or B. Please click the button A and button B in turn until you are sure of the quality difference between the two.

c) Please use the slider to give the comparative quality rating between A and B. For example, if you think A contains less noise than B, the slider can be placed in the left side of bar (<0); if you think A contains much less noise than B, the slider can be shifted leftwards (<<0); if you perceives that A equals B in background noise, the slider can be placed to the middle of the bar (0). The scale on top of the slider . The left side of the slider indicates the rating value that appears below the bar as shown in Figure C. 2.

d) After you give the rating value using the slider, please click “next” to give another rating with the same procedures as explained in b) and c).

e) If you want a break at anytime, please click the button “stop”. You can continue the procedure by clicking the button A or B again.

f) Please keep on clicking and give ratings until the title of “Background Noise” is replaced by “Finished!”.

References

AHARON, M., ELAD, M. & BRUCKSTEIN, A. 2006. K -SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. IEEE T. Signal Proces., 54, 4311-4322.

ALCANTARA, J. I., DOOLEY, G. J., BLAMEY, P. J. & SELIGMAN, P. M. 1994. Preliminary evaluation of a formant enhancement algorithm on the perception of speech in noise for normally hearing listeners. Audiology, 33, 15-27.

ALCANTARA, J. I., MOORE, B. C., KUHNEL, V. & LAUNER, S. 2003. Evaluation of the noise reduction system in a commercial digital hearing aids. Int. J. Audiol., 42, 34-42.

AMOS, N. E. & HUMES, L. E. 2007. Contribution of high frequencies to speech recognition in quiet and noise in listeners with varying degrees of high-frequency sensorineural hearing loss. J. Speech Lang. Hear. R., 50, 819-834.

ANSI 1997. Methods for calculation of the speech intelligibility index. New York: American National Standards Institute.

AREHART, K. H., HANSEN, J. H. L., GALLANT, S. & KALSTEIN, L. 2003. Evaluation of an auditory masked threshold noise suppression algorithm in normal-hearing and hearing-impaired listeners. Speech Commun., 40, 575-592.

AREHART, K. H., KATES, J., ANDERSON, M. & HARVEY, L. 2007. Effects of noise and distortion on speech quality judgments in normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am., 122, 1150-1164.

BAER, T., MOORE, B. C. & GATEHOUSE, S. 1993. Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality and response times. J. Rehabil. Res. Dev., 30, 49-72.

BAER, T. & MOORE, B. C. J. 1993. Effects of spectral smearing on the intelligibility of sentences in noise. J. Acoust. Soc. Am., 94, 1229-1241.

BAER, T. & MOORE, B. C. J. 1994. Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. J. Acoust. Soc. Am., 95, 2277-2280.

BEERENDS, J. G., ENEMAN, K., HUBER, R., KREBBER, J. & LUTS, H. Year.

Speech Quality Measurement for the Hearing Impaired on the Basis of PESQ. In: Audio Eng. Soc. Convention 124, 2008 Amsterdam.

*BEERENDS, J. G. & STEMERDINK, J. A. 1994. A perceptual speech-quality measure based on a psychoacoustic sound representation. J. Audio Eng. Soc., 42***,** *115-123.*

*BENCH, J., KOWAL, A. & BAMFORD, J. 1979. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. Brit. J. Audiol., 13***,** *102-112.*

*BOLL, S. 1979. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. ASSP, 27***,** *113-120.*

*BOYMANS, M., SCHONEVELD, P. & VERSCHUURE, J. 1999. Clinical evaluation of a full-digital in-the-ear hearing instrument. Audiology, 38***,** *99-108.*

*BRAMSLOW, L. 2004. An objective estimate of the perceived quality of reproduced sound in normal and impaired hearing. Acta Acust., 90***,** *1007-1018.*

BRAMSLOW, L. Year. Validation of objective sound quality models for hearing aids. In: Int. Hear. Aid Conf., 2008.

BREITHAUPT, C., GERKMANN, T. & MARTIN, R. Year. A novel a priori SNR estimation approach based on selective cepstro-temporal smoothing. In: IEEE ICASSP 2008. , March 31 2008-April 4 2008 2008. 4897-4900.

BSA 2011. Recommended Procedure: Pure Tone air and bone conduction threshold audiometry with and without masking and determination of uncomfortable loudness levels. The British Society of Audiology.

COMBESCURE, P., LE GUYADER, A. & GILLOIRE, A. Year. Quality evaluation of 32 kbit/s coded speech by means of degradation category ratings. In: IEEE ICASSP '82., May 1982 1982. 988-991.

*DAHLQUIST, M., LUTMAN, M. E., WOOD, S. & LEIJON, A. 2005. Methodology for quantifying perceptual effects from noise suppression systems. Int. J. Audiol., 44***,** *721-732.*

*DAU, T., PUSCHEL, D. & KOHLRAUSCH, A. 1996a. A quantitative model of the "effective" signal processing in the auditory system. I. Model structure. J. Acoust. Soc. Am., 99***,** *3615-3622.*

*DAU, T., PUSCHEL, D. & KOHLRAUSCH, A. 1996b. A quantitative model of the "effective" signal processing in the auditory system. II. Simulations and measurements. J. Acoust. Soc. Am., 99***,** *3623-3631.*

*DAWSON, P. W., MAUGER, S. J. & HERSBACH, A. A. 2011. Clinical evaluation of signal-to-noise ratio-based noise reduction in Nucleus Cochlear Implant recipients. Ear. Hear., 32***,** *382-390.*

DAYAN, P. & ABBOTT, L. 2001. Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. , Cambridge, Massachusetts, The MIT Press.

*DE BOER, E. & KUYPER, P. 1968. Triggered Correlation. Biomedical Engineering, IEEE Transactions on, 15***,** *169-179.*

DELLER, J. R., HANSEN, J. & PROAKIS, J. G. 2000. Discrete-Time Processing of Speech Signals, New York, IEEE Press.

DILLON, H. 2001. Hearing Aids, Thieme, New York.

*DIRKS, D., MORGAN, D. & DUBNO, J. 1982. A procedure for quantifying the effects of noise on speech recognition. J. Speech Hear. Disord., 47***,** *114-123.*

*DRESCHLER, W. A. & PLOMP, R. 1980. Relations between psychophysical data and speech perception for hearing-impaired subjects. I. J. Acoust. Soc. Am., 68***,** *1608-1615.*

*DRESCHLER, W. A. & PLOMP, R. 1985. Relations between psychophysical data and speech perception for hearing-impaired subjects. II. J. Acoust. Soc. Am., 78***,** *1261-1270.*

*DRULLMAN, R., FESTEN, J. M. & PLOMP, R. 1994. Effect of temporal envelope smearing on speech reception. J. Acoust. Soc. Am., 95***,** *1053-1064.*

*DUBNO, J. R. & SCHAEFER, A. B. 1992. Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners. J. Acoust. Soc. Am., 91***,** *2110-2121.*

*ELBERLING, C., LUDVIGSEN, C. & KEIDSER, G. 1993. The design and testing of a noise reduction algorithm based on spectral subtraction. Scand. Audiol. Suppl., 1993***,** *39-49.*

EPHRAIM, Y. & MALAH, D. Year. Speech enhancement using optimal non-linear spectral amplitude estimation. In: ICASSP, 1983. 1118-1121.

*EPHRAIM, Y. & MALAH, D. 1985. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. IEEE T. Acoust. Speech, 33***,** *443-445.*

FABRY, D. A. & VAN TASELL, D. J. 1986. Masked and filtered simulation of hearing *loss: Effects on consonant recognition. J. Speech Hear. Res., 29,* ***,*** *170-178.*

FAIRBANKS, G. 1958. Test of Phonemic Differentiation: The Rhyme Test. J. Acoust. Soc. Am., 30, 596-600.

FIELD, D. J. 1994. What is the goal of sensory coding. Neural Computation, 6, 559-601.

FLETCHER, H. & STEINBERG, J. C. 1930. Articulation Testing Methods. J. Acoust. Soc. Am., 1, 17-21.

FRENCH, N. R. & STEINBERG, J. C. 1947. Factors governing the intelligibility of speech sounds. J. Acoust. Soc. Am., 19, 90–119.

GERKMANN, T. & HENDRIKS, R. C. Year. Noise power estimation based on the probability of speech presence. In: Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on, 16-19 Oct. 2011 2011. 145-148.

GERKMANN, T. & HENDRIKS, R. C. 2012. Unbiased MMSE-Based Noise Power Estimation With Low Complexity and Low Tracking Delay. Audio, Speech, and Language Processing, IEEE Transactions on, 20, 1383-1393.

GERKMANN, T. & MARTIN, R. 2009. On the Statistics of Spectral Amplitudes After Variance Reduction by Temporal Cepstrum Smoothing and Cepstral Nulling. Signal Processing, IEEE Transactions on, 57, 4165-4174.

GLASBERG, B. R. & MOORE, B. C. 1989. Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech. Scand. Audiol. Suppl., 32, 1-25.

GORDON-SALANT, S. 1987. Effects of acoustic modification on consonant recognition by elderly hearing-impaired subjects. J. Acoust. Soc. Am., 81, 1199-1202.

GUELKE, R. W. 1987. Consonant burst enhancement: A possible means to improve intelligibility for the hard of hearing. Journal of Rehabilitation Research and Development, 24, 217-220.

GUSTAFSSON, H., NORDHOLM, S. E. & CLAESSON, I. 2001. Spectral subtraction using reduced delay convolution and adaptive averaging. IEEE Trans. SAP, 9, 799-807.

HANSEN, J. H. L. & CLEMENTS, M. A. 1991. Constrained iterative speech enhancement with application to speech recognition. Signal Processing, IEEE Transactions on, 39, 795-805.

HANSEN, M. & KOLLMEIER, B. Year. Using a quantitative psychoacoustical signal representation for objective speech quality measurement. In: IEEE ICASSP '97, 21-24 Apr 1997 1997. 1387-1390 vol.2.

HANSEN, M. & KOLLMEIER, B. 1999. Continuous assessment of time-varying speech quality. J. Acoust. Soc. Am., 106, 2888-2899.

HARLANDER, N., ROSENKRANZ, T. & HOHMANN, V. 2012. Evaluation of model-based versus non-parametric monaural noise-reduction approaches for hearing aids. Int. J. Audiol., 1-13.

HECKER, M. H. L. & WILLIAMS, C. E. 1966. Choice of Reference Conditions for Speech Preference Tests. J. Acoust. Soc. Am., 39, 946-952.

HOLLUBE, I. & KOLLMEIER, B. 1996. Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model. J. Acoust. Soc. Am., 100, 1703-1716.

HOU, Z. & PAVLOVIC, C. V. 1994. Effects of temporal smearing on temporal resolution, frequency selectivity, and speech intelligibility. J. Acoust. Soc. Am., 96, 1325-1340.

HOUSE, A. S., WILLIAMS, C. E., HECKER, M. H. L. & KRYTER, K. D. 1965. Articulation-Testing Methods: Consonantal Differentiation with a Closed-Response Set. J. Acoust. Soc. Am., 37, 158-166.

HOUTGAST, T. & STEENEKEN, H. J. M. 1973. The modulation transfer function in room acoustics as a predictor of speech intelligibility. Acustica, 28, 66-73.

HOUTGAST, T. & STEENEKEN, H. J. M. 1985. A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. J. Acoust. Soc. Am., 77, 1069–1077.

HOYER, P. Year. Non-negative sparse coding. In: Proc. IEEE Workshop on Neural Networks for Signal Processing, 2002.

HOYER, P. 2004. Non-negative matrix factorization with sparseness constraints. Journal of Machine Learning Research, 5, 1457–1469.

HU, H., LI, G., CHEN, L., SANG, J., WANG, S., LUTMAN, M. E. & BLEECK, S. Year. Enhanced sparse speech processing strategy for cochlear implants. In: Eusipco, 2011a Barcelona, Spain.

HU, H., SANG, J., LUTMAN, M. E. & BLEECK, S. Year. Simulation of hearing loss using compressive gammachirp auditory filters. In: IEEE ICASSP '11, 2011b Prague, Czech Republic. 5428 - 5431

HU, Y. & LOIZOU, P. 2003. A generalized subspace approach for enhancing speech corrupted with colored noise. IEEE Transactions on Speech and Audio Processing, 11, *334-341.*

HU, Y. & LOIZOU, P. C. 2007. A comparative intelligibility study of single-microphone noise reduction algorithms. J. Acoust. Soc. Am., 122, *1777-1786.*

HU, Y. & LOIZOU, P. C. 2008. Evaluation of objective quality measures for speech enhancement. IEEE T. Speech Audi. P., 16, *229-238.*

HUBER, R. & KOLLMEIER, B. 2006. PEMO-Q: A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception. IEEE T. Audio Speech Lang. P., 14, *1902-1911.*

HUBER, R., SCHULTE, M., VORMANN, M. & CHALUPPER, J. Year. Objective measures of speech quality in hearing aids: prediction of listening effort reduction by noise reduction algorithms. In: 2nd Workshop on Speech in Noise, 2010 Amsterdam.

HUMES, L. E., DIRKS, D. D., BELL, T. S. & KINCAID, G. E. 1987. Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers. J. Acoust. Soc. Am., 81, *765-773.*

HYV RINEN, A. 1999. Sparse code shrinkage: Denoising of nonGaussian data by maximum likelihood estimation. Neural Comput., 11, *1739-1768.*

HYV RINEN, A., HOYER, P. & OJA, E. Year. Sparse code shrinkage for image denoising. In: Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence., 4-9 May 1998 1998. 859-864.

IRINO, T. & PATTERSON, R. D. 1997. A time-domain, level-dependent auditory filter: The gammachirp. J. Acoust. Soc. Am., 101, *412-419.*

IRINO, T. & PATTERSON, R. D. 2001. A Compressive Gammachirp Auditory Filter for Both Physiological and Psychophysical Data. J. Acoust. Soc. Am., 109, *2008-2022.*

IRINO, T. & PATTERSON, R. D. 2006. A Dynamic Compressive Gammachirp Auditory Filterbank. IEEE T. Audio Speech Lang. P., 14, *2222-2232.*

ISO389-1 2000. Acoustics - Reference zero for the calibration of audiometric equipment Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones.

ITU 2000. Perceptual evaluation of speech quality (PESQ): An objective method for *end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. ITU-T Recommendation.*

JAE, L. & OPPENHEIM, A. 1978. All-pole modeling of degraded speech. IEEE T. Acoust. Speech, 26, ***,*** *197-210.*

JAFARI, M. G. & PLUMBLEY, M. D. 2011. Fast Dictionary Learning for Sparse Representations of Speech Signals. Selected Topics in Signal Processing, IEEE Journal of, 5, ***,*** *1025-1031.*

JAMIESON, D. G., BRENNAN, R. L. & CORNELISSE, L. E. 1995. Evaluation of a speech enhancement strategy with normal-hearing and hearing impaired listeners. Ear. Hear., 16, 274-286.

JESPER, J. & RICHARD, H. 2007. Improved Subspace-Based Single-Channel Speech Enhancement Using Generalized Super-Gaussian Priors. IEEE T. Audio Speech Lang. P., 15, 862-872.

JI, S., XUE, Y. & CARIN, L. 2008. Bayesian Compressive Sensing. Signal Processing, IEEE Transactions on, 56, 2346-2356.

JOLLIFFE, I. T. 1986. Principal Component Analysis, New York, Springer-Verlag.

KANG, G. S. & FRANSEN, L. J. 1989. Quality improvement of LPC-processed noisy speech by using spectral subtraction. IEEE Trans. ASSP, 37, 939-942.

KATES, J. 1992. On using coherence to measure distortion in hearing aids. J. Acoust. Soc. Am., 91, 2236-2244.

KATES, J. & AREHART, K. H. 2005. Coherence and the speech intelligibility index. J. Acoust. Soc. Am., 117, 2224-2237.

KATES, J. & AREHART, K. H. 2010. The hearing-aid speech quality index (HASQI). J. Audio Eng. Soc., 58, 363-381.

KATES, J. M. & WEISS, M. R. 1996. A comparison of hearing-aid array-processing techniques. J. Acoust. Soc. Am., 99, 3138-3148.

KILLION, M. C., STAAB, W. J. & PREVES, D. A. 1990. Classifying automatic signal processors. Hearing Instruments, 41, 24-26.

KIM, G. & LOIZOU, P. 2010. Improving speech intelligibility in noise using a binary mask that is based on magnitude spectrum constraints. IEEE Signal Proc. let., 17, 1010-1013.

KLATT, D. Year. Prediction of perceived phonetic distance from critical-band spectra:

A first step. In: IEEE ICASSP '82, May 1982 1982. 1278-1281.

KRYTER, K. D. 1962. Methods for the calculation and use of the Articulation Index. J. Acoust. Soc. Am., 34, *1689–1697.*

LARSON, V. D., WILLIAMS, D. W., HENDERSON, W. G., LUETHKE, L. E., BECK, L. B., NOFFSINGER, D., WILSON, R. H., DOBIE, R. A., HASKELL, G. B., BRATT, G. W., SHANKS, J. E., STELMACHOWICZ, P., STUDEBAKER, G. A., BOYSEN, A. E., DONAHUE, A., CANALIS, R., FAUSTI, S. A. & RAPPAPORT, B. Z. 2000. Efficacy of 3 commonly used hearing aid circuits - a crossover trial. J. Amer. Med. Assoc, 284, *1806-1813.*

LEVITT, H. 1971. Transformed Up-Down Methods in Psychoacoustics. J. Acoust. Soc. Am., 49, *467-477.*

LEVITT, H. 1993. Digital hearing aids. In: STUDEBAKER, G. A. & HOCHBERG, I. (eds.) Acoustical Factors Affecting Hearing Aid Performance. Boston: Allyn and Bacon.

LEVITT, H. 2001. Noise reduction in hearing aids: A review. J. Rehabil. Res. Dev., 38, *111-121.*

LEVITT, H., BAKKE, M., KATES, J., NEUMAN, A., SCHWANDER, T. & WEISS, M. 1993. Signal processing for hearing impairment. Scand. Audiol. Suppl., 38, *7-19.*

LI, G. 2008. Speech perception in a sparse domain. PhD thesis, Institute of Sound and Vibration, University of Southampton.

LI, G. & LUTMAN, M. E. Year. Sparse Stimuli For Cochlear Implants. In: Proc. EUSIPCO, 2008 Lausanne, Switzerland.

LOCKWOOD, P. & BOUDY, J. 1992. Experiments with a non-linear spectral subtractor (NSS), Hidden └/Markov models and the projection, for robust speech recognition. Speech Commun., 11, *215-228.*

LOIZOU, P. C. 2007. Speech Enhancement: Theory and Practive. CRC, Talor & Francis.

LUNNER, T., HELLGREN, J., ARLINGER, S. & ELBERLING, C. 1998. Non-linear signal processing in digital hearing aids. Scand. Audiol. Suppl., 27, *40-49.*

LUTFI, R. A. & PATTERSON, R. D. 1984. On the Growth of Masking Asymmetry with Stimulus Intensity. J. Acoust. Soc. Am., 76.

LUTMAN, M. E. 1991. Degradations in frequency and temporal resolution with age *and their impact on speech identification Acta Oto-laryngol, 476, 120-126.*

LUTMAN, M. E., GATEHOUSE, S. & WORTHINGTON, A. G. 1991. Frequency resolution as a function of hearing threshold level and age. J. Acoust. Soc. Am., 89, 320-328.

MA, J., HU, Y. & LOIZOU, P. 2009. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. J. Acoust. Soc. Am., 125, 3387-3405.

MAIRAL, J., BACH, F., PONCE, J. & SAPIRO, G. Year. Supervised dictionary learning. In: In Proc. International Conference on Machine Learning, 2009.

MAIRAL, J., BACH, F., PONCE, J. & SAPIRO, G. 2010. Online Learning for Matrix Factorization and Sparse Coding. Journal of Machine Learning Research, 11, 19-60.

MARZINZIK, M. 2000. Noise reduction schemes for digital hearing aids and their use for the hearing-impaired. PhD, University of Oldenburg.

MILLER, G. A. & NICELY, P. E. 1955. An Analysis of Perceptual Confusions Among Some English Consonants. J. Acoust. Soc. Am., 27, 338-352.

MOHAMMADIHA, N., TAGHIA, J. & LEIJON, A. Year. Single channel speech enhancement using bayesian NMF with recursive temporal updates of prior distributions. In: ICASSP, 2012 Kyoto, Japan. 4561 - 4564

MONTOGOMERY, A. A. & EDGE, R. A. 1988. Evaluation of two speech enhancement techniques to improve intelligibility for hearing-impaired adults. J. Speech Hear. Res., 31, 386-393.

MOORE, B. C. 1996. Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. Ear. Hear., 17, 133-160.

MOORE, B. C. & GLASBERG, B. R. 1988. A comparison of four methods of implementing automatic gain control (AGC) in hearing aids. Brit. J. Audiol., 22, 93-104.

MOORE, B. C. J. (ed.) 2007. Cochlear Hearing Loss: Physiological, Psychological and Technical Issues Wiley-Blackwell.

MOORE, B. C. J. & GLASBERG, B. R. 1993. Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech. J. Acoust. Soc. Am., 94, 2050-2062.

MOORE, B. C. J., PETERS, R. W. & GLASBERG, B. R. 1990. Auditory filter shapes at *low center frequencies. J. Acoust. Soc. Am., 88, 132-140.*

MUNOZ, C. M. A., NELSON, P. B., RUTLEDGE, J. C. & GAGO, A. Year. Frequency lowering processing for listeners with significant hearing loss. In: Proceedings of ICECS '99, 5-8 Sep 1999 1999. 741-744 vol.2.

MUNSON, W. A. & KARLIN, J. E. 1962. Isopreference Method for Evaluating Speech-Transmission Circuits. J. Acoust. Soc. Am., 34, 762-774.

NEJIME, Y. & MOORE, B. C. J. 1997. Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise. J. Acoust. Soc. Am., 102, 603-615.

NILSSON, M., SOLI, S. D. & SULLIVAN, J. A. 1994. Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. J. Acoust. Soc. Am., 95, 1085-1099.

NOLL, P. Year. Adaptive quantization in speech coding systems. In: Int. Zurich Seminar on Digital Communication (IEEE), 1976 Zurich.

OLSHAUSEN, B. A. & FIELD, D. J. 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381, 607-609.

OPPENHEIM, A. & VERGHESE, G. C. 2010. Signals, Systems, and Inference.

PARSA, V. & JAMIESON, D. G. Year. Hearing aid distortion measurement using the auditory distance parameter. In: Audio Eng. Soc. Conv. 111, 2001.

PATTERSON, R. D. 1976. Auditory filter shapes derived with noise stimuli. J. Acoust. Soc. Am., 59, 640-654.

PATTERSON, R. D., ALLERHAND, M. H. & GIGUERE, C. 1995. Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. J. Acoust. Soc. Am., 98, 1890-1894.

PATTERSON, R. D. & MOORE, B. C. 1986. Auditory filters and excitation patterns as representations of frequency resolution. In: MOORE, B. C. (ed.) Frequency Selectivity in Hearing. London: Academic Press.

PATTERSON, R. D., NIMMO-SMITH, I., WEBER, D. L. & MILROY, R. 1982. The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. J. Acoust. Soc. Am., 72, 1788-1803.

PATTERSON, R. D., UNOKI, M. & IRINO, T. 2003. Extending the Domain of Center Frequencies for the Compressive Gammachirp Auditory Filter. J. Acoust. Soc. Am., 114, 1529-1542.

PAVLOVIC, C. V. 1984. Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment. J. Acoust. Soc. Am., 75, 1253–1258.

PAVLOVIC, C. V. 1987. Derivation of primary parameters and procedures for use in speech intelligibility predictions. J. Acoust. Soc. Am., 82, 413–422.

PELLOM, B. L. & HANSEN, J. H. L. 1998. An improved (Auto:I, LSP:T) constrained iterative speech enhancement for colored noise environments. IEEE T. Speech Audi. P., 6, 573-579.

PISARENKO, V. F. 1973. The retrieval of harmonics from a covariance function. Geophysics, J. Roy. Astron. Soc., 33, 347-366.

PLOMP, R. 1986. A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. J. Speech. Hear. Res., 29, 146-154.

PLOMP, R. 1994. Noise, amplification, and compression: Considerations of three main issues in hearing aid design. Ear Hearing, 15, 2-12.

POTAMITIS, I., FAKOTAKIS, N. & KOKKINAKIS, G. Year. Speech enhancement using the sparse code shrinkage technique. In: IEEE ICASSP '01, 2001 2001a. 621-624.

POTAMITIS, I., FAKOTAKIS, N. & KOKKINAKIS, G. Year. Speech enhancement using the sparse code shrinkage technique. In: Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on, 2001 2001b. 621-624 vol.1.

QUACKENBUSH, S., BARNWELL, T. & CLEMENTS, M. (eds.) 1988. Objective Measures of Speech Quality, Englewood Cliffs: NJ: Prentice Hall.

REVOILE, S. G., HOLDEN-PITT, L., PICKETT, J. M. & BRANDT, F. 1986. Speech cue enhancement for the hearing impaired: I. Altered vowel durations for perception of final fricative voicing. J. Speech Hear. Res., 29, 240-255.

RIEKE, F., WARLAND, D., STEVENINCK, R. & BIALEK, W. 1999. Spikes: Exploring the Neural Code, Cambridge, Massachusetts, The MIT Press.

RIX, A. W., BEERENDS, J. G., HOLLIER, M. P. & HEKSTRA, A. P. Year. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs. In: IEEE ICASSP '01, 2001 2001. 749-752.

RIX, A. W. & HOLLIER, M. P. Year. The perceptual analysis measurement system for robust end-to-end speech quality assessment. In: IEEE ICASSP '00, 2000

2000. 1515-1518.

RNID 2010. Analysis of Medical Research Funding into Hearing Loss in the UK.

*ROBINSON, K. 1998. Implications of developmental plasticity for the language acquisition of deaf children with cochlear implants. Int. J. Pediatr. Otorhinolaryngol., 15***,** *71-80.*

*ROSEN, S. & BAKER, R. J. 1994. Characterising auditory filter nonlinearity. Hear. Res., 73***,** *231-243.*

SANG, J., HU, H., LI, G., LUTMAN, M. E. & BLEECK, S. Year. Supervised sparse coding strategy in hearing aids. In: IEEE ICCT 2011 2011a China.

SANG, J., HU, H., WINTER, I., WRIGHT, M. & BLEECK, S. Year. The 'Neural Space': a physiologically inspired noise reduction strategy based on fractional derivatives. In: IEEE Int. Sym. Commun. IT., 2011b China.

SANG, J., HU, H., ZHENG, C., LI, G., LUTMAN, M. E. & BLEECK, S. 2012. Evaluation of a sparse coding shrinkage noise reduction algorithm with normal hearing and hearing impaired listeners. EUSIPCO. Bucharest, Romania.

*SCHIJNDEL, N. H. V., HOUTGAST, T. & FESTEN, J. M. 2001. Effects of degradation of intensity, time, or frequency content on speech intelligibility for normal-hearing and hearing-impaired listeners. J. Acoust. Soc. Am., 110***,** *529-542.*

*SCHMIDT, R. 1986. Multiple emitter location and signal parameter estimation. IEEE T. Antenn. Propag., 34***,** *276-280.*

*SCHUM, D. J. 2003. Noise-reduction circuitry in hearing aids, II: Goals and current strategies. Hearing Journal, 56***,** *32-41.*

SHIELD, B. 2006. Evaluation of the social and economic costs of hearing impairment. A report for Hear-IT AISBL.

SIGG, C. D., DIKK, T. & BUHMANN, J. M. Year. Speech enhancement with sparse coding in learned dictionaries. In: IEEE ICASSP '10, 2010 USA. 4758 - 4761

*TAAL, C. H., HENDRIKS, R. C., HEUSDENS, R. & JENSEN, J. 2011a. An Algorithm for Intelligibility Prediction of Time Frequency Weighted Noisy Speech. IEEE T. Audio Speech Lang. P., 19***,** *2125-2136.*

*TAAL, C. H., HENDRIKS, R. C., HEUSDENS, R. & JENSEN, J. 2011b. An evaluation of objective measures for intelligibility prediction of time-frequency weighted noisy speech. J. Acoust. Soc. Am., 130***,** *3013-3027.*

TOBIN, H. 1997. Practical Hearing Aid Selection And Fitting, Diane Pub Co.

TRIBOLET, J., NOLL, P., MCDERMOTT, B. & CROCHIERE, R. Year. A study of complexity and quality of speech waveform coders. In: ICASSP '78, Apr 1978 1978. 586-590.

VERSCHUUR, C., LUTMAN, M. E. & WAHAT, N. H. A. 2006. Evaluation of a non-linear spectral subtraction noise suppression scheme in cochlear implant users. Cochlear Implants Int., 7, *188-193.*

VOIERS, W. D. 1983. Evaluating processed speech using the diagnostic rhyme test. Speech Technology, 1, *30-39.*

WANG, D., KJEMS, U., PEDERSEN, M. S., BOLDT, J. B. & LUNNER, T. 2009. Speech intelligibility in background noise with ideal binary time-frequency masking. J. Acoust. Soc. Am., 125, *2336-2347.*

WANG, S., SEKEY, A. & GERSHO, A. 1992. An objective measure for predicting subjective quality of speech coders. Selected Areas in Commun, IEEE Journal on, 10, *819-829.*

WATSON, C. S. 1991. Speech-perception aids for hearing-impaired people: Current status and needed research. J. Acoust. Soc. Am., 90, *637-685.*

WEISS, M. & NEUMAN, A. C. 1993. Noise reduction in hearing aids. In: STUDEBAKER, G. A. & HOCHBERG, I. (eds.) Acoustical factors affecting hearing aid performance. Boston: Allyn and Bacon.

WIDROW, B. & LUO, F. 2003. Microphone arrays for hearing aids: An overview. Speech Commun., 39, *139-146.*

WIENER, N. (ed.) 1949. Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications, Cambridge: MA: MIT Press.

YANG, L. & FU, Q. 2005. Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. J. Acoust. Soc. Am., 117, *1001-1004.*

YI, H. & LOIZOU, P. C. 2004. Speech enhancement based on wavelet thresholding the multitaper spectrum. Speech and Audio Processing, IEEE Transactions on, 12, *59-67.*

YI, H. & LOIZOU, P. C. 2008. Evaluation of Objective Quality Measures for Speech Enhancement. Audio, Speech, and Language Processing, IEEE Transactions on, 16, *229-238.*

YOSHIOKA, T. Year. Maximum likelihood approach to speech enhancement for noisy reverberant signals. In: IEEE ICASSP '08, 2008 Las Vegas, USA. 4585 - 4588.

ZENTON, G., KAH-CHYE, T. & TAN, T. G. 1998. Postprocessing method for suppressing musical noise generated by spectral subtraction. IEEE Trans. SAP, 6, 287-292.

ZHOU, M., PAISLEY, J. & CARIN, L. Year. Nonparametric learning of dictionaries for sparse representation of sensor signals. In: IEEE Int.l Workshop, CAMSAP, 13-16 Dec. 2009 2009. 237-240.

ZHOU, M., YANG, H., SAPIRO, G. & DUNSON, D. Year. Covariance-dependent dictionary learning and sparse coding. In: IEEE ICASSP '11, 2011 Prague, Czech. 5824 - 5827.

ZOU, X., JANCOVIC, P., JU, L. & KOKUER, M. 2008. Speech Signal Enhancement Based on MAP Algorithm in the ICA Space. IEEE T. Signal Proces, 56, 1812-1820.

ZUREK, P. M. & DELHORNE, L. A. 1987. Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment. J. Acoust. Soc. Am., 82, 1548-1559.