Evaluation of changes in speech production induced by conventional and level-dependent hearing protectors and noise characteristics

by

Ghazaleh Vaziri

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of
Doctorate of Philosophy in Rehabilitation Sciences

Rehabilitation Sciences Ph.D. Program
Faculty of Health Sciences

Abstract

The use of personal hearing protection devices (HPDs) is often recommended to protect workers' hearing from noise-induced damage when no other means of reducing noise levels at the source is effective. The effects of HPDs on speech communication cannot be neglected in spite of their benefit in reducing the risk of hearing loss. While much research has been directed at speech perception, much less is known on how HPDs affect speech production. The tendency of talkers to raise their vocal effort in noise, known as the Lombard effect, is often disrupted by HPDs due to their occlusion effect and the lower noise at the ears as well as the attenuated feedback from one's own voice.

Three main knowledge gaps are addressed in this thesis. The first gap is to characterize speech produced by talkers with or without HPDs under realistic acoustic conditions while immersed in an external noise field. The second gap is to evaluate more comprehensively speech production under protected and unprotected talker and listener ear conditions in different types of fluctuating and continous noises. The third gap is to assess the alterations in the characteristics of speech produced by talkers wearing level-dependent HPDs set at different transmission gain settings and in comparison with passive HPDs.

This thesis extends methods used to recover Lombard speech elicited in an external noise field. For this purpose, two noise suppression methods, direct waveform subtraction (DWS) and adaptive noise cancellation (ANC), were found to adequately remove noise from speech recorded for SNRs as low as −10 dB. Moreover, this work contributes new knowledge on the effects of conventional passive HPDs on speech production. When talker wears HPD in noise then speech level were found to decrease by up to 9 dB in continuous noises and by 7 dB in fluctuating noises compared to open ears, while speech levels were found to increase by about 5 dB in all noises when the listener wears HPD. Furthermore, changes in pitch and spectral levels were consistent with changes in speech levels. The effects of level-dependent HPD on speech production, depending on the chosen transmission gain setting, revealed that it led to smaller decrease in talkers' speech levels in noise compared to conventional passive HPD. These findings indicate that the level-dependent HPDs may impede communication less than conventional passive HPDs, while providing protection against high levels of noise.

To my parents who have taught me to be a strong woman,

and to my husband.

Acknowledgements

This doctoral thesis as a result of six years of work could not be accomplished but by guidance and support of some wonderful people. I would like to take this opportunity to appreciate them and express my gratitude.

First, I would like to thank deeply my supervisors Prof. Christian Giguère and Prof. Hilmi Dajani. Their guidance and support were invaluable and constructive throughout this challenging journey. It was a great experience for me to work so closely with them who cared so much about my work. Prof. Giguère s passionate supervision, tremendous support, meticulous and detailed review of my work, and his remarkable comments were indeed precious for me helping me find my way as an engineer in the field of audiology, hearing and speech sciences. I am very grateful to Prof. Dajani for his valuable suggestions and comments, support, constructive feedback, which helped me accomplish my thesis.

I am very grateful to the members of my thesis proposal evaluation committee, Prof. Josée Lagacé (President), Prof. Ian Mackay, Prof. Chantal Laroche, and Prof. Jérémie Voix. Their constructive feedback and comments have contributed to the quality of my thesis.

I am very grateful to my friends and labmates, Dr. Nicolas Ellaham and Dr. Flora Nassrallah, for all their support and help. I was fortunate to have Nicolas' technical help as well as his broad knowledge in the field. He was very handful in setting up the experimental environment and was very kind and patient to answer my questions. My thanks goes to Flora who supported me technically and morally over past few years in the Lab.

My sincere thanks go to my best friend Ellie Askari. Her company, sympathy, and support during last few years means a lot to me. I would also like to thank my dear friend

Dr. Navid Tadayon for his support and help. I am also thankful to all my friends in Ottawa whose names I would not be able to list but have truly contributed to this journey through their valuable friendship.

I am forever grateful to my parents Hamzeh Vaziri and Ladan Bakhshi and my brother Shahin. It was their love, encouragement and support that have derived me to get where I am today. They patiently have tolerated thousands miles of separation to let me pursue my passion in life.

Last but not least, I am forever grateful to my wonderful husband Aidin. It was his inspiration and encouragement that have motivated me to begin this long journey. I could not have done it without the support and the confidence he gave me throughout the last six years. He has been with me during all good and/or stressful times and believed in me since the first day.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ANC	Active Noise Cancellation
ANSI	American National Standards Institute
CE	Channel Estimation
Dir.	Directional
DWS	Direct Waveform Subtraction
ESII	Extended Speech Intelligibility Index
EU	European Union
H&H	Hyper & Hypo
HINT	Hearing-In-Noise Test
HPD	Hearing Protection Device
IFFM	International Female Fluctuating Masker
IFnoise	International Female noise
ISTS	International Speech Test Signals
LMS	Least Mean Square
MOS	Mean Opinion Score
NIHL	Noise-Induced Hearing Loss
NR	Noise Reduction
NRR	Noise Reduction Rating
NS	No Speech
Omni.	Omnidirectional
PC	Principe Component
PESQ	Perceptual Evaluation of Speech Quality
RLS	Recursive Least Square
SCoG	Spectral Center of Gravity
SII	Speech Intelligibility Index
SNR	Speech-to-Noise Ratio
SRT	Speech Reception Threshold
SSnoise	Speech-spectrum noise
STI	Speech Transmission Index
SURE	Stein Unbiased Risk
V/UV	Voiced-to-Unvoiced
WPT	Wavelet Packet Transform

Chapter 1

Introduction

1.1 Statement of problem and motivation

Exposure to high level occupational noises is one of the risk factors which jeopardizes well-being and safety of workers in many industrial and military workplaces. In the United States (US), it has been found that around 22 million workers are exposed to daily hazardous occupational noise levels [1]. According to the statistics revealed by the European Union (EU), about 30 % of the workers are exposed to loud noise at least one-fourth of their work time [2, 3]. Several studies conducted in 12 countries in South America, Africa, and Asia also reported that there are high occupational noise levels in many industrial and non-industrial workplaces [2].

It has been found that there is an interaction between the noise exposure and the symptoms of psychological distress, fatigue, and physical problems (e.g. cardiac

malfunction, high blood pressure) in many workers [4]. Moreover, the occupational noise has adverse effects upon concentration on tasks, job performance, and job satisfaction [4]. High noise level also endangers workers and public safety since it impairs the ability of workers to communicate, to detect important and warning signals, and to localize the sounds [5, 6].

Apart from all these effects on health outcomes and auditory function, noise-induced hearing loss (NIHL) is one of the most serious effects arising from exposure to high noise level [2]. In addition to temporary hearing loss, workplace noise can cause permanent hearing loss in the workers. It has been reported that worldwide, 16−24% of hearing impairment is caused by occupational noise [1]. In the US, the Bureau of Labor Statistics revealed that NIHL is one of the most common occupational diseases [5], and accounted for nearly 11% of all occupation-related illnesses in 2004 and 2005 [1, 2, 5]. Recent statistics obtained by Veterans Affairs Canada confirms the high prevalence of NIHL among retired military personnel [7]. The resulting NIHL could exacerbate speech understanding difficulties when communicating in noise or challenging work environments [5].

The most effective solution to prevent NIHL is to eliminate the noise hazards [1, 6]. Although control technologies consisting of engineering and administrative strategies are most desirable to reduce the noise hazards and exposure [8], they may not be practical and possible to implement in many situations [1, 6]. As an example, it is not often economically or technically possible to discontinue the use of a noisy machine, to modify the working process to reduce noise, to isolate the source of noise from workers by using distance or barriers, or to limit the exposure duration in many industrial or military workplaces. Under these circumstances, the only possible approach to protect the workers from the noise-induced damage is to use personal hearing protection devices (HPD) [9, 10]. Several national

and international standards [11-14] specify the procedures for proper testing, selection, fit and use of hearing protectors.

In spite of providing protection against noise hazards, hearing protectors have inevitable effects not only on the auditory perception of speech and other desirable sounds but also on the way speech is produced. It has been found that wearing conventional hearing protectors makes communication more difficult [15] and impairs the perception of workplace sounds and warning signals [16-19]. Level-dependent hearing protection devices are able to enhance situational awareness (e.g. warning signal perception, sound localization, speech communication, detection of distant events) while providing effective protection against loud noise [20, 21]. In contrast to the fixed attenuation of conventional protectors, they have a variable transmission gain which depends on the surrounding noise level. As such, they provide attenuation at high noise levels while they amplify speech and other surrounding sounds in lower levels of noise [20-23]. This is achieved by gain compression circuitry which allows maintaining communications in soft noises while protecting hearing against loud noises [23]. Thus, they can provide new opportunities for workers at risk of noise-induced hearing loss [23, 24]. However, they may have inevitable impacts on speech production resulting from their effects on the user's perception of auditory feedback and surrounding noise,

There is a gap in knowledge regarding the effects of wearing level-dependent hearing protection devices, as well as conventional protectors, on the acoustical characteristics of speech produced by users in noisy environments. The available research has focussed on the effects of conventional passive protectors [15-18, 25-34] and active level-dependent hearing protection devices [17, 18, 24, 31, 33-38] on speech perception, speech recognition, and sound localization. In contrast, there have been few studies about the impact of passive

protectors on the users' vocal output [15, 27, 39-41]. Other limitation of most available studies on speech production is related to the delivery of noise directly to the ears through headphones. This technique makes it difficult to evaluate the real impact of hearing protectors on speech production, especially with earmuffs, and to measure speech output in an actual external noise field.

1.2 Research objectives

The overall goal of this doctoral thesis is to evaluate the effects of wearing passive level-independent and active level-dependent hearing protection devices on speech production in different noise conditions. For this purpose, three main knowledge gaps are addressed in the context of speech communication in noise. The first one relates to the characterization of speech produced by talkers while immersed in an external noise field under different occluded and unoccluded ear conditions. An appropriate speech enhancement method providing significant noise reduction while introducing little distortion to the speech signal is needed in order to extract the Lombard speech from the vocal output recorded in the noise field. Secondly, this work aims to evaluate more comprehensively the characteristics of speech produced in different noise conditions, and so the speech produced by talkers under protected and unprotected ear conditions is investigated in different types of fluctuating and steady-state noise. Thirdly, this work aims to assess the alterations in the characteristics of Lombard speech arising from active level-dependent hearing protectors at different gain settings and in comparison with passive protection.

This thesis aims to extend our understanding of speech production in noise and the effects of using hearing protection devices on speech communication. It also contributes key data towards the development of comprehensive tools or quantitative models to assess the impact of different types of noise, the hearing status of talkers, and hearing protection devices on face-to-face speech communication in real noisy environments.

1.3 Methodology and approach

In order to achieve the above-mentioned objectives, a study involving the technical development of measurement procedures as well as experiments with real subjects was conducted:

- Development and evaluation of measurement procedures for recording speech output in noise from talkers using hearing protectors: Several noise suppression techniques were implemented and evaluated for our purpose of recovering the clean speech in an external noise field with a minimum degree of distortion. A battery of tests and in-laboratory noise measurements were carried out to select the enhancement methods that provide the biggest size of noise reduction over a wide range of SNRs, from a set of available methods.
- Pilot laboratory measurements using an acoustic manikin and a real participant: These data served to evaluate the effects of talkers' body movements on the size of noise reduction by the selected enhancement methods, and to assess the performance of those methods in recovering the Lombard speech in the external noise field.

- In-laboratory speech measurements with multiple talker participants wearing passive protection: The goal was to evaluate the effects of a passive hearing protector (earmuff), while worn by the talker and/or an acoustic manikin serving as the listener, and the effects of the spectral and temporal characteristics of the background noise on the talkers' speech features. In order to assess the impacts of wearing hearing protector by each side of the communication pair, four different protected and unprotected conditions for the pair of talker and listener were also tested in the experiment.
- In-laboratory speech measurements were carried out with multiple talker participants with level-dependent protection: The goal was to evaluate the impact of different gain settings and attenuation of an active level-dependent hearing protection device on the speech produced by the talkers. Two fluctuating and steady-state noises were used to reveal the alterations in the speech produced by protected and/or unprotected talkers in response to the changes in the characteristics of background noises.

1.4 Thesis outline and contributions

The remainder of this thesis is organized as follows:

Chapter 2 provides an overview of speech production in noise by explaining different perspectives proposed by theories of speech. The mechanism responsible for the production of the Lombard effect is explained followed by a review of the literature to address the effects of noise type, noise level, occlusion effect and wearing hearing protectors, on the

Lombard speech. Finally, the practical use of the Lombard effect in different fields is briefly discussed.

Chapter 3 begins with an explanation about the necessity of eliciting the Lombard effect in an external noise field and subsequently the need to use speech enhancement methods to recover the clean speech from the noisy voice recorded in the noise field. The size of noise reduced by several speech enhancement methods is evaluated through experimental tests. The last part of this chapter is an article published in the *Proceedings of Meetings on Acoustics (170th Meeting of the Acoustical Society of America)*, and entitled "A comparison of speech enhancement methods to extract Lombard speech in an external noise field".

Chapter 4, the second article of this thesis submitted to *International Journal of Speech Technology*, investigates the performance of the two selected noise suppression methods, direct waveform subtraction and adaptive noise cancellation, for Lombard speech extraction in a more realistic experimental condition where a real participant's body movements is used to simulate dynamic sound field disturbances rather than static displacements of an acoustic manikin. Moreover, the effects of acquisition configuration (microphone types, distances from the mouth) and noise type on the size of noise reduction will be evaluated.

Chapter 5 is the third article of this thesis prepared for a peer reviewed journal. This study evaluates the effect of a conventional passive hearing protector (earmuff-type) on the level, pitch and the spectral distribution of speech produced in quiet and in different types of continuous and fluctuating noises presented at different levels in an external sound field. Different ear conditions are investigated for talker and listener in terms of hearing protection use through conducting an experiment by 24 participants as talker and an acoustic manikin as a target listener.

Chapter 6 as the fourth article of this thesis for an upcoming submission in a peer reviewed journal contributes new knowledge of the effects of a level-dependent hearing protection device (HPD) on the talkers' speech produced in two noise types (continuous and fluctuating) at three different noise levels in an external sound field. Moreover, different operation modes of the HPD will be compared to the passive mode, representing a passive hearing protector, through an experiment carried out by 24 participants in a noise simulation room. An article entitled as "The effect of wearing conventional and level-dependent hearing protectors on speech production in noise and quiet" was presented in *42th National Hearing Conservation Association Conference (NHCA).*

Finally, Chapter 7 provides a summary of the research findings, contributions, limitations and potential avenues for future work.

Chapter 2

Review of the Literature

2.1 An overview of speech production by different theories

Speech is a major means people use for communication in everyday life [42, 43]. It is an acoustic signal that transmits a linguistic message from the speaker to the listener through the vocal-auditory channel [44]. Speech production is a complicated process which results from the coordination of movements within the individual subsystems of a complex system. The complex system includes respiratory, laryngeal and supralaryngeal subsystems [42, 45].

Different theories have been proposed about the overall nature of speech production [42]. Stage theory hypothesizes that there is a serial level of control to produce speech. In contrast, action or dynamic system theory claims that the speech motors act based on the specific task to be performed. This means that only necessary and related motors respond in order to fulfill a given task. The speech motors are controlled in groups or in synergy to

accomplish a task. Therefore, the control system does not need to prepare and send individual command for each muscle during speech production. Another important characteristic of action theory is the possibility of choosing different coordinative structures for a particular task [42]. The speech controller makes it possible for the speech production system to select different coordinative structure or synergy of muscles to accomplish a specific task based on the biological and environmental situations. In addition to the flexibility in choosing the coordinative structure, the speech production system can also regulate the magnitude of motor responses relative to speaking condition [42, 46]. Therefore, articulatory movements can be modified in response to the changes of speaking rate and stress level. This addresses the fact that a certain sound of speech is produced differently based on the context and speaking condition [42, 46].

In contrast to the two above-mentioned theories (e.g. stage and action theories) focusing on how the speech motor system is controlled during speech production, the Hyper and Hypo (H&H) theory is concerned with the linguistic and communicative functions of speech [42, 43, 46]. This theory is based on intra-speaker variations in speech production. According to the theory, two distinct concepts, hyperspeech & hypospeech (H&H), describe how the speech production system regulates the behaviour of related speech motors. The concept of "hyperspeech" is attributed to the plasticity and output-oriented control of articulatory movements. In contrast, "hypospeech" is defined as economy and system-oriented control of speech motor [43, 46]. Hypospeech demonstrates that the speech motor usually tends to an economical physiological effort [43, 46]. It implies that the control system regulates the speech articulators such that they can move as easily as possible with minimum cost of energy. Thus, the emphasis of hypospeech is on the reduction of energy costs for articulatory movements rather than on the intelligibility and quality of speech [42,

43]. However, there are particular demands for intelligible speech in some challenging communicative settings. Under these circumstances, the output-oriented nature of motor control (hyperspeech) triggers the movements of speech articulators in an effort toward the production of intelligible and high quality output [43, 46]. Thus, according to the Hyper and Hypo (H&H) theory, the communicative needs of the speaker control the production of speech. Speakers are able to control speech production and to adapt their speaking style to different situations through a continuum of hypospeech to hyperspeech or vice versa [42, 46]. The intelligibility of speech can also be modified by the speaker relative to the needs of communicative environment. In an unconstrained usual communicative setting, the default behaviour of motors to produce speech tends to a low-cost form. However, when constraints are imposed by physiological, cognitive, social, and communicative factors, the compensatory feedback is sent to speech motor to modify their movements in favor of intelligible speech [43, 46].

2.2 Speech communication in noise

Speech communication is affected by many factors including age, gender, vocal effort, language proficiency, cognitive skills, hearing loss and/or speech impediments, acoustic transmission properties such as talker-listener distance and orientation, reverberation, noise, availability of nonacoustic cues, and the wearing of devices like headphones, hearing aids, and hearing protectors (See Figure 2.1) [47]. By changing the way speech is produced, transmitted and/or received, each factor can affect the intelligibility and quality of spoken speech, and hence speech perception/recognition.

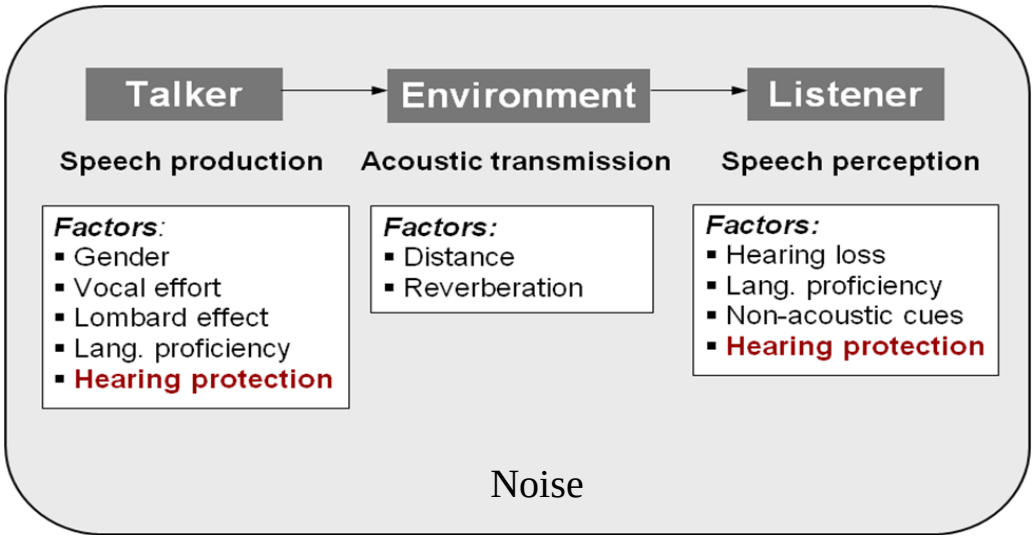


Figure 2.1: *Different Factors affecting speech communication between a talker and a listener (adapted from [47] with permission).*

Among many different factors, noise affects all dimensions of speech communication between a talker and a listener, including speech production, speech perception and the acoustical transmission properties (Figure 2.1). In real life, there are many sources of environmental noises including the speech of other talkers nearby and the sound of machinery which hinder speech communication. The interference of environmental noises with speech communication and auditory perception can cause negative consequences. In manufacturing settings, for example, machinery noise interferes with workers' communication, warning sound perception and the recognition of other important acoustic cues such as audible equipment defects. These interferences not only influence auditory performance and work efficiency, but they also endanger workers and public safety [5, 6, 7]. In addition to the destructive impact of noise on optimal communication, long-term exposure to high levels of noise is one of the most important causes of permanent or temporary hearing loss in workers [6, 8, 48].

By masking the acoustic and linguistic cues available in spoken message, the background noises reduce the intelligibility of speech for target listeners [16, 49]. From a listener's perspective, the masking effect of background noise increases the effective threshold of audibility. The level of a target speech must be elevated to be identified easily in noise since the auditory system needs a neural discharge rate higher than the spontaneous discharge rate to detect that desirable sound. The higher level of the target speech increases the listening speech-to-noise ratio (SNR), often defined as the decibel difference between the speech level and the noise level [50]. As the level of background noise increases, the desirable signal level must also be increased in order to be detected and/or recognized.

2.3 The Lombard effect

Talkers are able to modify the articulatory movements and the acoustic, phonetic, and linguistic content of their speech in order to adapt to the surrounding acoustic environment [51]. The communicative needs of a speaker with respect to target listener are important factors controlling speech articulatory movements. In noisy environments, speech motors drive the movements of speech articulators to produce high-intelligible speech and high-quality vocal output [52, 53]. Intelligibility is often measured as the percentage of words which listeners can understand and identify correctly, while quality is related to the naturalness, clarity and pleasantness of speech [54, 55].

To compensate for the unfavorable effect of background noise, talkers involuntary increase their vocal effort to enhance the audibility of their own voice for the listeners. This phenomenon, referred to as the Lombard effect, was first described by the French

otolaryngologist Etienne Lombard in 1911 [56, 57]. Lombard found a direct relationship between the level of background noise and the level of vocal output: the higher the noise level, the greater vocal intensity [56].

There have been arguments among researchers on how to interpret the cause of the Lombard effect. Some authors believe that the Lombard effect is a physiological reflex, implying that speaker is not aware of raising the voice level in presence of noise. Indeed, it is claimed that the mechanism responsible for the production of Lombard effect is the auditory feedback channel masked by the ambient noise [58]. However, it seems that this assumption is not sufficient to describe the phenomenon. Another assumption is that this effect is driven by the speaker's tendency to overcome the damping effect of noise on speech intelligibility [58]. Thus, the speakers attempting to enhance the audibility of their own voice for the audience by raising the voice level are not completely unaware of the production of the Lombard effect. The importance of the communication aspect in the Lombard phenomenon was supported by stating that the speakers do not change their voice level to communicate better with themselves, but rather with others [59]. It has been observed that speakers have stronger vocal changes in a communicative interaction compared to when they are just reading or speaking in a non-communicative setting [60]. On the other hand, ignoring the Lombard effect as a communicative phenomenon would support the assumption that this effect is only a physiological reflex [61, 62]. In an effort to integrate the different aspects of this phenomenon, it has been hypothesized that the changes of vocal effort in noisy environments, the Lombard effect, is driven by both reflexive and communicative mechanisms [60, 63-65].

Typically, a noise level below 45 dB SPL does not have a significant effect on the level of vocal output. Indeed, the Lombard effect as a mechanism to overcome the destructive

effects of noise typically appears for noise levels over 45-55 dB SPL [66-68]. In response to the increase of noise level, the level of vocal output is raised until about 100 dB SPL, the maximum human vocal effort [69].

The increase of speech level as a function of ambient noise level, referred to as the Lombard slope, has been investigated to be in the range of 0.2 to 1 dB/dB [66, 67, 70]. For example, Korn (1954) reported a Lombard slope of 0.38 dB/dB, meaning a 0.38 dB increase in speech level for every 1 dB increase in the noise level, where 50 speakers read sentences and engaged in conversation [71]. Pickett (1958) reported a Lombard slope of about 0.3 dB/dB in the free-field acoustical condition of an anechoic chamber [72]. Webster and Clumpp (1962) found a Lombard slope of 0.5 dB/dB under conditions of pairs of subjects communicating under various levels of ambient noise [73]. Kryter (1946) found a Lombard slope of 0.33 when the speakers read word lists in noise levels reaching as high as 105 dB SPL [25]. In 2004, Dodd and Whitlock measured the Lombard effect in different masking noises for 18 children reading a book in an anechoic chamber. They showed that for voice level above 53.4 dB, there is a linear relationship between voice level and noise level with the slope of 0.22 dB/dB [74]. In 2004, Sato and Bradley found a slope of 0.82 dB/dB for 28 noisy classrooms [75]. A Lombard slope of 0.69 dB/dB was estimated in an experiment conducted in 10 eating establishments [66].

2.3.1 The characteristics of Lombard speech

From a physiological perspective, the speech produced in noise, called here the Lombard speech, is observed as a hyper-articulation which is reflected by an amplification of the

articulatory movements [76], such as higher amplitude of mouth and jaw, and rigid head movements. The respiratory system and laryngeal mechanism have important contribution to increased vocal intensity [42]. Talkers change their breathing pattern resulting in higher volume and pressure in lung to raise vocal intensity. The stiffness of vocal folds and the glottal shape change when speaking louder, which affect the glottal airflow waveform [76]. All these physiological factors contribute to the increased vocal effort when producing Lombard speech [77, 78].

In addition to the main feature of Lombard speech of an increased vocal output or speech level [78, 79, 80] in noise, many investigations have been carried out to study the complete range of acoustic-phonetic changes associated with Lombard speech produced by normal-hearing talkers. Compared to normal speech, the acoustic-phonetic characteristics of Lombard speech typically include an increase in the fundamental frequency or pitch [63, 81-83, 77-79], a decrease in speaking rate [78], an increase in the relative duration of vowels with respect to consonants [63, 79, 80, 83], a slight decrease in duration of consonants [63], an increase in duration of words [63, 81], a reduction in the percentage of silence in speech [84], an increase in sentence duration [85], a shift of energy from low frequency bands to middle or high frequency bands [79, 80], flattening of the spectral tilt for vowels and semi-vowels [84], an increase in the first [79] and second formant frequencies [63, 64], a reduction in the bandwidth of the first four formants for most phonemes [63, 81], an increase in formant amplitude especially for sonorants [86]. These acoustic-phonetic effects for Lombard speech are listed in Table 2.1. All the articulatory and acoustic-phonetic changes make the Lombard speech to be more intelligible than the conversational speech produced in quiet [79, 87, 88].

Table 2.1: *The variations of acoustic-phonetic characteristics of speech in noise*

Parameter	Effect
Vocal intensity (vocal output Level)	increase
Fundamental frequency or Pitch (F_0)	increase
Speaking rate	decrease
Duration	Vowels: increase Consonants: decrease Words: increase Sentences: increase
The percentage of silence in speech	decrease
Spectral center of gravity	upward shift
Spectral tilt	flattening for vowels and semi-vowels
Formant frequencies	increase in first and second formants
Bandwidth of first four Formants	decrease for most of the vowels, glides, liquids, and nasal
Amplitudes of Formants	increase

Quantitative data on the acoustic-phonetic changes are available from several studies; however, research methods vary widely so that direct comparisons are difficult. Junqua and Anglade (1990) found that the energy of vowels produced in white noise of 85 dB SPL decreased by 17−37 % in the frequency range of 0-500 Hz, while the energy in the frequency range of 4-5 kHz increased for the female speakers. The first formant for vowels, glides, liquids, and nasals increased by 42−113 Hz, and pitch or fundamental frequency (F_0) increased by 82−106 Hz for male speakers [83]. Garnier et al. (2006) investigated the alteration of several acoustic-phonetic features of speech in two white and cocktail party

noises [77]. For white noise, word duration, vocal output level and the fundamental frequency F_0 increased by 11%, 12.9 dB and 55.5 Hz respectively, while for cocktail party noise they increased by 8.2%, 8.6 dB and 65 Hz. Moreover, the increase of spectral energy in the frequency ranges of 1500-3500 Hz and 3500−5500 Hz in white noise was 23 dB and 27 dB, respectively. In cocktail party noise, the increase was 19 dB in both above-mentioned frequency ranges.

Davis et al. (2006) observed an average increase of 11 dB in vocal output level and a higher mean F_0 for all participants in babble and white noises at 80 dB SPL relative to a quiet condition [78]. They also measured the articulatory movements and submitted their data to a Principle Component (PC) analysis to uncover the primary articulatory changes associated with the Lombard effect. The main changes of articulatory movements were reported by the first 6 PCs including PC1: jaw motion (mouth opening), PC2: mouth opening and eyebrow raising, PC3: head translation toward the hearer, PC4: lip protrusion, PC5: mouth opening and eyebrow closure and PC6: rotation. As expected, there was a significant increase in the amplitude of articulatory movements, except for PC3, in noise compared to a quiet condition.

In an experimental study [64], four different communicative conditions including "neither person in noise", "both listener and speaker in noise", "speaker in noise" and "listener in noise" were studied with white noise, if present, presented to subjects at 70 dB SPL. In the "both listener and speaker in noise" condition, the mean vocal level was 7.4 dB higher than that in the "neither person in noise" condition. In the "speaker in noise" and "listener in noise" conditions, the vocal levels were 6.3 dB and 2.2 dB higher compared to the "neither person in noise" condition, respectively. Although the increase in the vocal level by 2.2 dB in the "listener in noise" condition was small, this supports the Lane and Tranel's

(1971) [70] interpretation of the Lombard effect as a communicative phenomenon. In this case, the speaker not situated in noise shows the Lombard effect only by considering the noisy situation of the target listener. The duration of words in the "both listener and speaker in noise", "speaker in noise" and "listener in noise" conditions were 20−30 ms longer than that in "neither person in noise" condition.

2.3.2 The application of Lombard effect in different fields

The Lombard effect has been proposed as a phenomenon that could be integrated in some hearing tests such as for the diagnosis of hearing loss and also for studying the function of audio-vocal channel during speech production [62]. Clinically important, the Lombard effect is used to study speech production, voice disorders and can serve as a treatment method for patients suffering from certain vocal malfunctions [62]. As an example, the evidence indicates that rehabilitation training sessions focusing on the Lombard effect can be beneficial for people with alaryngeal speech, stuttering and Parkinson [39, 61, 89]. It has been shown that the masking effect of noise on speech production can be used as a therapeutic strategy to induce esophageal subjects to raise their voice intensity through the Lombard effect [56]. Likewise, it has been observed that the frequency of stuttering in noisy background is reduced so that interventions based on the Lombard effect could be beneficial for this population to improve fluency [39]. In return, promoting the Lombard effect as therapeutic intervention in the speech clinic could induce hyperfunctional voice disorders so that prior evaluation of the integrity of vocal structures supporting Lombard speech is warranted.

Beyond applications in the medical fields, Lombard speech has also been utilized in the area of telecommunication [62]. Based upon models of the Lombard effect, a speech synthesis system has been proposed to transform natural speech into noise-free Lombard speech [88]. This artificially-generated Lombard speech can then be transmitted over telecommunication channels (i.e. telephone) rather than the original normal speech from the talker. Given that the Lombard speech is more intelligible than normal speech in noise, the system produces more intelligible output over noisy telecommunication channels [90]. Generally speaking, the approaches of enhancing intelligibility based on the Lombard effect would be applicable to many other communication systems such as cell phone, public address system and automatic speech and speaker recognition [91, 92, 93].

Moreover, the size of the Lombard effect can be employed to set the maximum level of noise and to enhance speech intelligibility in architectural acoustic design [62]. By predicting speech intelligibility, models of the Lombard effect can help architects control the effects of reverberation, competing noise and other adverse factors in order to optimize speech communication in public spaces such as eating establishments, dining spaces, and classrooms [66, 74]. This is possible because reverberation time, the interior room materials, the arrangement of furniture, and the location of noise source relative to that of target sound affect the background sound level and hence vocal output and speech intelligibility [94]. More specifically, a novel model based on the Lombard effect was proposed which is able to predict speech and noise levels as well as speech intelligibility in eating establishments as a function of the number of customers [66]. Likewise, a model of teacher's vocal output was provided which estimates the level of speech at 1 metre in noisy free field based on the level of noise in classroom and a Lombard constant to accommodate a range of Lombard slopes [66].

Finally, it has been investigated that the size of changes in F0 and speech level in the presence of noise could be used to estimate the talkers' intended communication distance. The changes in speech production arising from the occlusion of ears by the hearing protectors could also be measures when the level of background noise and the residual noise in the ears are known. Thus, it is hypothesized that a model of speech production in noise as a function of talker-to-listener distance and background noise level can be provided for different occluded ear conditions. Such a model could be used to enhance the communicative experience and interactions among multiple users of wireless communication headsets in some environments, whereby only the intended listeners would receive the message on the basis of the intended communication distance by the talker [41].

2.4 Current topics in speech production in noise

2.4.1 Interaction between the temporal characteristics of noise and speech production

The masking effect of noises, apart from the level, also varies depending on other characteristics, for instance, amplitude modulations or fluctuations, the energy distribution over the frequency spectrum, and the existence of any linguistic information [95]. Thus, the size of noise-induced changes in auditory perception and speech production is affected by the type and characteristics of the background noise [66, 96–98].

Noises can be steady-state or fluctuating in time. As an example, time-varying maskers include speech of competing talkers as well as noise modulated by a sinusoidal or square

wave or by the broadband temporal envelope from single or multiple talkers [99]. Most real noises we encounter in everyday life are often fluctuating [97].

The listeners have better understanding of speech spoken in fluctuating masker compared to that spoken in steady-state noise [97−101]. A speech reception threshold (SRT) improvement of 10−14 dB was reported for listeners exposed to intermittent noise (16 Hz square-modulated noise) compared to stationary noise [97]. In time-varying real noises, speech intelligibility depends on the ability to take advantages of brief "acoustic glimpses" of speech in dips, where the target speech is relatively unmasked due to momentary decrease of fluctuating noise level [97, 99, 102, 103]. This phenomenon is referred to as dip listening or temporal masking release. Dip listening has little importance in steady background noise due to absence of any dips with sufficient duration and amplitude, as a result of which intelligibility highly depends on the higher-level portions of speech [104].

In fluctuating noises, the Lombard effect is hypothesized to be reduced due to less effective masking of these noises [64, 101]. Among the few studies focusing on the talker's end, Cooke and Lu [85] evaluated the effect of N-speaker babble noises on speech production for different values of N ranging from N=1 (single speaker) to N=∞ (speech-shaped noise). It is worth mentioning that as the number of speakers (N) decreases the temporal fluctuations in the babble noise get more pronounced. They measured the noise-induced changes of the acoustic parameters of the produced speech including sentence duration, energy, fundamental frequency, spectral center of gravity, formant frequencies, formant bandwidths, spectral tilt, sentence start time, number and duration of short pauses, and voiced-to-unvoiced (V/UV) energy ratio at the utterance and phoneme levels in the different maskers.

At the utterance level, most parameters showed significant changes in different noise masker conditions compared to a quiet baseline condition [85]. Across the different maskers, energy increased by 3 to 9 dB and mean F_0 rose by 0.6 to 2.5 semitones. Spectral center of gravity had an increase of 20%−38% from the quiet baseline of 870 Hz. The mean sentence duration increased by 2.4%−7.6% relative to the mean duration of 1.64 s found in quiet. Moreover, the pause before speaking which represents the tendency of talkers to avoid overlapping with background masker was raised by 6%−18% from a baseline of 0.55s in the quiet condition. The V/UV energy ratio increased by 2.4 dB from the value of 8.6 dB found in quiet. There was an increase in the duration for most types of speech sounds. The spectral center of gravity also rose by more than 25% for all speech sounds. It was observed that F_1 frequency had an increase by up to 100 Hz for each vowel. Increases in F_2 and F_3 energy for vowels were also obtained in the noise maskers compared to the quiet baseline. Moreover, F_1 bandwidth increased while F_2 and F_3 bandwidths decreased for all the vowels in the noise maskers. They observed that the noise-induced changes of speech parameters relative to the quiet condition increased with the number of talkers in the babble noise. As an example, the modifications of energy, mean F_0 and spectral center of gravity, and the increases in F_2 and F_3 energy were greater for speech-shaped noise compared to competing talker noise. Thus, the results confirm that the effect of fluctuating noise on speech production is smaller than the effect of stationary noise (speech-shaped noise, $N=\infty$), when presented at the same overall level.

MacDonald and Raufer [101] studied the variation and adaptation of speech production (Lombard effect) in response to an amplitude-modulated noise with different modulation rates, delivered at two presentation levels of 60 and 80 dB SPL. The talkers were asked to read a list of 50 fixed grammar five-word sentences and also to speak clearly to the operator.

The level and F_0 of Lombard speech produced in the modulated (i.e. fluctuating) noise were compared with those related to speech produced in quiet and unmodulated (i.e. continuous) background noise. The average speech level and F_0 increased in noises of 60 and 80 dB SPL, either modulated or unmodulated, compared to the quiet condition. This is consistent with the results of other studies regarding the Lombard effect [77, 78, 81, 83]. The increases were smaller in the lower noise level relative to the higher one. Moreover, little changes were observed between the 60 dB SPL modulated and unmodulated noises, while the increases in speech level and F_0 were larger for the 80 dB SPL unmodulated noise compared to that for the 80 dB SPL modulated noise. The latter was consistent with Lu and Cook's findings [85]. This implies that less vocal effort is required to maintain intelligibility in the fluctuating noise relative to that in stationary noise. In addition, based upon the results of energy distribution of speech over time, the talkers reduced the overlap between their utterances and amplitude-modulated noise (with modulation rates of 2 and 4 Hz) by approximately 2% in order to enhance the intelligibility of their vocal output.

In most studies, the acoustical characteristics of speech have been measured where subjects are exposed to steady-state (speech-shaped) noise [64, 77, 78, 83, 91, 105−108] while many natural noises interfering with speech communication in daily life are often temporally fluctuating [97, 109−112]. Thus, the effect of different time-varying masking noises on speech production and perception needs to be evaluated to better recognize the effect of everyday real noises on speech communication. In particular, at the talker's end, few investigations have studied the variations of acoustic-phonetic characteristics of speech in fluctuating noise [85, 101].

2.4.2 Effects of wearing hearing protection devices on speech production

2.4.2.1 Interaction between Lombard and occlusion effect

Use of hearing protectors are advised to prevent noise-induced damage when no other means of reducing noise levels at the source is effective [27, 40, 113]. By blocking the ear canal, hearing protectors can attenuate the level of sound to a more comfortable lower level. However, it has been reported that one important factor hindering workers from consistent use of such protectors is their interference with speech communication and sound perception [27, 29, 32, 33, 37]. The disturbing effect of wearing protectors is not only on speech reception but also modifies the way speech is produced [22, 25, 113, 114]. They interfere with perception and production of speech by affecting the users' auditory feedback of their own voice and the level of surrounding noise heard [22, 25, 39, 114]. Monitoring one's own speech output (auditory feedback) has great importance in adjusting the articulatory movements and in regulating the vocal effort relative to the surrounding acoustical condition, which contribute to maintain speech intelligible for the listeners [40, 41].

Wearing of hearing protectors increases the users' perception of their own voice relative to the auditory feedback received in the open-ear condition. This phenomenon is called as the occlusion effect [115]. Auditory feedback is received through two hearing pathways, air and bone conduction, so that the occlusion effect produced by hearing protectors can affect the balance between these two pathways [41, 116]. Occlusion of the ear canal by an insertion tip or a cap of a hearing protector blocks air conduction through the ear canal while amplifying the bone-conducted vibrations of talker's own voice [115, 117]. Low-frequency

vibratory displacements of the canal walls, concha, or pinna, develop significantly greater pressure at the eardrum in the occluded condition compared to that in open ear condition. The occlusion effect reaches a maximum at low frequencies and reduces with increasing frequency [115]. Thus, the blockage of the ear canal by the protectors causes the talkers to perceive their own voice through bone conduction as boomy and louder in low frequencies [117, 118]. The acoustic reflex (an involuntary contraction of the middle ear stapedial muscles triggered by high-intensity sound stimulus and the talker's vocalization) may also be affected by hearing protectors and further disrupt the self-monitoring of one's own speech. In practice, users of hearing protectors may show a reduction in their Lombard effect because of the increase in their perception of auditory feedback and the decrease in their perception of surrounding noise [39]. The reduction in the vocal output may lead to deterioration in speech intelligibility for nearby listeners [27, 39].

2.4.2.2 Passive hearing protectors

Attenuating speech and noise equally, the conventional passive hearing protectors (i.e. earplugs or earmuffs) commonly used in many occupational settings may impede optimal aural communication of workers and users' awareness of the surrounding acoustic environment [22, 24, 31, 32, 37, 49, 119]. The interference of protectors' fixed attenuation with the hearing acuity can be more challenging for workers immersed in fluctuating noise condition and for who have pre-existing hearing loss [31, 40].

There have been many studies investigating the effect of wearing conventional hearing protectors mostly on speech and auditory perception but only a few on speech production [15, 18, 25, 27, 28, 29, 32, 33, 34, 37, 40, 41, 117]. The effect of protectors on speech

communication depends on the level of noise in which they are worn [25, 27, 40]. Kryter (1946) indicated that intelligibility is increased when earplugs are worn at the high noise level (≥80 dBA), while in lower level of noise the use of earplugs decreases speech intelligibility [25]. It has been found that wearing protectors causes no degradation on the ability of normal-hearing listeners to understand speech in noise levels above 85−90 dBA [25, 27, 40]. It has been observed that there is a benefit of about 10% in speech discrimination in the extreme noise level for this population [28], even though there is an equal amount of attenuation for both speech and noise which keeps the signal-to-noise ratio constant. The benefit appears related to a release from perceptual overload (i.e. auditory filter broadening) when wearing hearing protectors at higher presentation levels [16]. At the low level of noise (<70 dBA), the protectors reduce the speech intelligibility most likely because of the attenuation of the lower level of speech to a level insufficient for recognition.

Talkers wearing the earmuffs and the earplugs lowered their vocal level by 2.7 dB and 4.2 dB, respectively. The intelligibility scores were reduced by 37% to 31% for the listeners [27]. In the situation where both the talkers and listeners wear protectors, the speech intelligibility is reduced because the damping effect on the talkers' vocal level exceeds the benefit of protectors for the listeners [27]. Hormann et al. (1984) investigated the effect of wearing ear protectors (earplugs) by both speaker and listener on verbal communication and the extent of speech intelligibility [15]. They showed that subjects speak an average of 4 dB lower in protected condition. Subjects with ear protectors speak the words approximately 20 percent faster than those with no protectors. Approximately 25 percent shorter pauses are made when wearing the earplugs. There are a 40 percent reduction in the understanding of speech and a considerable delay in the exchange of information in the condition of both talker and listener with protectors. The highest speech intelligibility score is related to the

situation where both talkers and listeners are with open eras, while the lowest score is obtained when the protector is worn by both of them [15]. Tufts and Frank (2003) observed that the overall speech level for open-ear talkers increased by 5 to 13 dB in response to raise in the background noise level from 60 to 100 dB SPL [40]. When talkers wore the earplugs and the earmuffs, there was a drop in the size of overall speech level and the SNRs by 4 to 11 dB relative to the unprotected group. The use of hearing protectors reduces the spectral center of gravity (SCoG: "center of mass" of the spectrum) of speech by 147 to 327 Hz, indicating that the energy of Lombard speech shifted to lower frequencies than in the unprotected condition. Furthermore, the estimation of speech intelligibility index (SII) under conditions where talkers wore earplugs or not showed that the speech information available for listeners decreased in the former condition relative to the latter, particularly in moderate to high level of background noise [40]. Brungart et al. (2012) showed that the speech level of talkers wearing earplugs reduced by 5.8 dB to 10.1 dB depending on the inserted side of plugs into the ear [117]. The use of earplugs impaired the intelligibility of speech predicted at the listeners' ears by 17% to 37% at the high noise level of 80 dBA.

2.4.2.3 Active level-dependent hearing protectors

The advent of level-dependent hearing protection devices (HPD) equipped with active noise cancellation technology and radio communication headsets into the marketplace appeared to be a good alternative for the conventional passive protectors [24]. In contrast to the conventional protectors, level-dependent protectors attenuate high level noise and intense sounds while amplifying speech and other desirable low level sounds. Hence, they are able to simultaneously provide effective protection against noise and to enhance communication

and situational awareness of the immediate work surroundings [21, 22, 49]. In a qualitative assessment study, the noise-exposed workers rated the level-dependent HPDs higher than the passive protectors in terms of their performance in improving communication and situational awareness [22].

Despite the advantages, the inevitable impacts of transmission gain and the associated compression circuitry of advanced level-dependent protector on speech production cannot be ignored. The available research on the advanced level-dependent HPDs is restricted to the studies at the listener' end evaluating the impacts of gain settings and the attenuation on the perception, recognition, detection and the localization of speech and other desirable sounds [17, 18, 23, 31, 33, 34, 35, 36, 37, 47, 120−125].

Based on the review of literature, there is a need to investigate further the consequences of wearing level-dependent hearing protection, as well as conventional protectors, on the speech acoustics produced in a noisy environment (talker's end). Moreover, research is needed to study in more detail the effects of these devices on the production of speech in the noisy workplace taking into account the temporal and spectral characteristics of the noise. Furthermore, more research needs to be conducted to reveal the effect of wearing hearing protectors on speech production in interaction with the listener.

Chapter 3

Speech Enhancement Methods

This chapter imposes challenges regarding the contamination of the recorded Lombard speech by the background noise. Subsequently, different speech enhancement methods are proposed to extract the clean Lombard speech from the noisy voice recording. Tests are conducted to select the methods based on the criteria defined for the most desirable enhancement technique. Finally, an experiment using an acoustic manikin as the talker in the simulation room is conducted to evaluate the performance of the best two selected methods in recovering the clean Lombard speech. Moreover, the effect of body movement simulated by static movement of manikin's head and torso on the speech recovery is investigated.

3.1 The Lombard effect in an external noise field

In most research on the Lombard effect, the noise condition has been generated by the headphones worn over the ears [40, 77, 78, 83, 91, 101, 105, 106, 107, 108]. Although such an experimental approach allows recording the produced speech in an ideal noise-free condition, the vocal level of talkers can be reduced by the occlusion effect of the headphones [117]. Talkers wearing headphones can hear their own voice louder so that the tendency to raise the level of speech could be reduced. Moreover, the noise provided to subjects through headphones is mixed with the auditory feedback from their own voice. Depending on the realism of the auditory feedback simulation and given the possible occlusion effect artefact on one's voice when using headphones, the noise-free Lombard speech recorded under these circumstances may differ from the noisy Lombard speech produced by talkers with open ears in real-world noisy environments. Other limitations of using headphones to elicit the Lombard effect are that it is less practical to use when the effect of wearing hearing devices, such as hearing protectors or hearing aids, is investigated. While head-related transfer functions and other auditory virtual reality processing can be integrated into the headphone noise-delivery system, the close proximity between the headphone unit and the pick-up microphone of in-ear devices such as hearing aids and active earplugs is such that correctly reproducing spatial cues from complex sound fields can become highly problematic due to near field acoustic coupling effects. It is also impractical to deliver of noise through headphones when evaluating earmuff-style hearing protectors.

Therefore, there is a need to study the Lombard effect elicited by an external noise field (e.g. loudspeakers) to take into account the realistic effects of noise on speech. The external noise field where subjects are immersed in noise with open ears is closer to a real-world

situation in which the speech output is modified by both background noise and the Lombard effect. In contrast to the use of headphones to present the noise, the use of loudspeakers allows the talkers to attend to the noise with no effect on the self-monitored auditory feedback [117, 126]. Furthermore, the use of sound field methods for eliciting the Lombard effect is the only possible choice when the effects of hearing protectors are evaluated.

3.2 Enhancement methods

The use of an external noise field, however, poses a challenge because of the contamination of the recorded Lombard speech by the eliciting noise. It becomes necessary to extract the clean Lombard speech from noise by utilizing speech enhancement techniques that provide significant noise reduction while introducing little distortion to the speech signal.

There are different enhancement methods available including spectral subtraction, Wiener filtering and wavelet-based methods, which are known as single-ended methods as they normally use only the microphone signal but are applicable to many practical situations, and adaptive noise cancellation (ANC), channel estimation (CE) and direct waveform subtraction (DWS) methods, which rely on having access to a priori information but are not readily applicable in many practical situations [54, 117, 118, 127, 128].

3.2.1 Spectral subtraction

Spectral subtraction is one of the first algorithms proposed for noise reduction. There are different types of spectral subtraction methods including magnitude spectral subtraction,

power spectral subtraction, spectral subtraction with over subtraction, and multiband spectral subtraction [129].

This method is based on the principle that the clean speech signal can be estimated by subtracting the spectrum of noise signal from the spectrum of noisy speech. Thus, a priori knowledge of the noise is necessary. The noise spectrum can be estimated during silence periods when no speech is present in the noisy speech signal. The noise in the gaps can be a representative of the whole noise contaminating the speech signal, under the assumption that the noise is additive and the spectrum does not change over time. This assumption implies that the noise should be stationary over time or slowly time-varying. Another limitation of spectral subtraction method is that the negative values resulting from the subtraction introduce a new type of noise called "musical noise" [129,130].

3.2.2 Wiener Filtering

A Wiener filter can be used to eliminate the noise in a noisy signal using a priori knowledge of the noise characteristics. A simple classical Wiener filter can be designed spectrally via the estimated ratio of energy of clean speech to that of noisy speech. One example of the transfer function of a Wiener filter is illustrated in Equation 3.1, where $\hat{\mathrm{S}}_k$ and $\mathrm{S_k}$ are the kth component of the vectors representing clean and noisy spectral samples of speech, respectively. $\mathrm{P_{sk}}$ corresponds to the kth parameter of the mean power spectrum of noisy speech (S) and $\mathrm{P_{nk}}$ is the kth parameter of estimated power spectrum of noise. The amount of noise suppression is sometimes controlled by a factor β. In classical case, $\beta = 1$ is chosen yielding a linear filter.

$$\hat{S}_k = \frac{P_{sk}}{P_{sk} + \beta P_{nk}} S_k \tag{3.1}$$

The power spectrum of noise can be estimated over the pause segments of the speech signal which contain only noise. Thus, this filtering method is dependent on the assumption that the noise is stationary [54]. It assumes that the noise extracted from the pause segments is representative of the background noise disturbing the speech signal. Thus, this method would be less accurate and effective for many non-stationary real noises. However, it can be refined by using another microphone to record the interfering noise separately, leading to a to a better estimation of the noise power [54].

3.2.3 Wavelet-based noise suppression method

Donoho and Johnstone (1994) proposed a speech enhancement algorithm based on the wavelet transform. By using the discrete wavelet transform, the noisy speech signal is decomposed into "detail" and "approximation" coefficients. The former relates to the high-frequency components of the speech signal, while the latter relates to its low-frequency components. Wavelet coefficients with large absolute values represent the energy of speech signal and the coefficients with small absolute values are related to noise and fine details of speech. Thus, the small wavelet coefficients of the noisy speech signal falling below a predefined threshold are set to zero in order to suppress the disturbing noise. Then, the inverse wavelet transform is applied to the coefficients in order to reconstruct a cleaner speech signal. The performance of the method depends on the SNR since the dominance of noise at very high noise level conditions (low SNR) causes the noise content with large

absolute values of wavelet coefficients prevents the removal of noise content associated with large absolute values of wavelet coefficients.

There are different threshold selection rule (e.g. universal, minimax, SURE) and thresholding function (e.g. hard, soft, firm, garrote, step-garrote thresholding) [131]. A drawback of wavelet thresholding is speech distortion arising from unwanted removal of speech information components during thresholding process [131].

Bouhara & Rouat (2006) extended the wavelet-based noise suppression technique to allow speech enhancement under conditions of time and frequency-varying noises [132]. Compared to the conventional uniform thresholding explained above, the extended technique with a time-scale adapted threshold for each wavelet's subband yields a better suppression of coefficients related to noise components in the noisy speech signal.

Another wavelet-based algorithm has been proposed in which a reference signal representing the disturbing noise is employed to robustly identify the coefficients arising from noise. This reference noise that is highly correlated with the original contaminating noise is recorded using a second microphone. Then, energies computed for the wavelet coefficients of reference signal are used as a comparator to detect voiced and unvoiced segments of the noisy speech signal. Finally, the coefficients detected as unvoiced are set to zero to eliminate the ones related to the masking noise [48].

3.2.4 Channel estimation (CE) method

In the channel estimation method, the noise contaminating speech is eliminated using the transfer function of the recording room which is estimated when the talker is quiet and still

(Figure 3.1). The transfer function is modeled by recording white noise in the room while the talker is asked to be quiet (See Acquisition and Channel estimation in Figure 3.1). Then, the transfer function is used to estimate the noise contaminating the talker's speech from the noise played in the room while the talker is speaking. Finally, the estimated noise is subtracted from the recorded noisy speech signal (See Site Noise Removal in Figure 3.1). As a necessary condition for obtaining a good result, the acoustic configuration of the recording room must not change at all during recordings. In other words, the talker must sit still while speaking [118]. Moreover, the quality of sound equipment and the distortion in the estimation of the transfer function over low and high frequencies are the main factors affecting the size of noise reduction.

3.2.5 Adaptive noise cancellation (ANC) method

By gaining access to the noise signal through the second channel (microphone), the adaptive noise cancellation (ANC) technique may be more effective for speech enhancement than other filtering techniques in adapting to time-varying noises [133]. It renders the filtering method more appropriate to handle many kinds of non-stationary real noises.

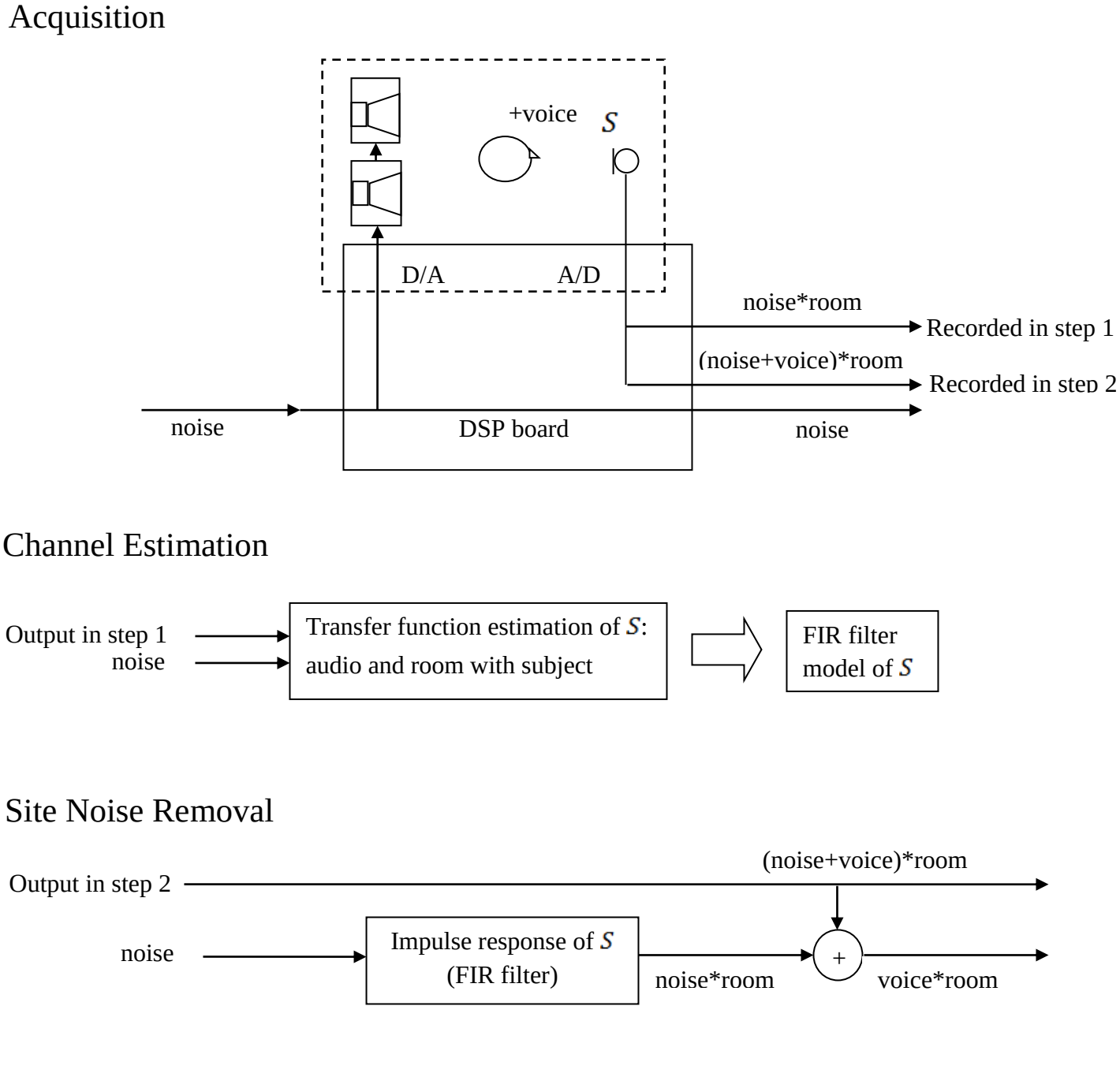


Figure 3.1: *Flowchart of acquisition, channel estimation, and site noise removal used in the Channel Estimation method (adapted from [118] with permission). See description in the text.*

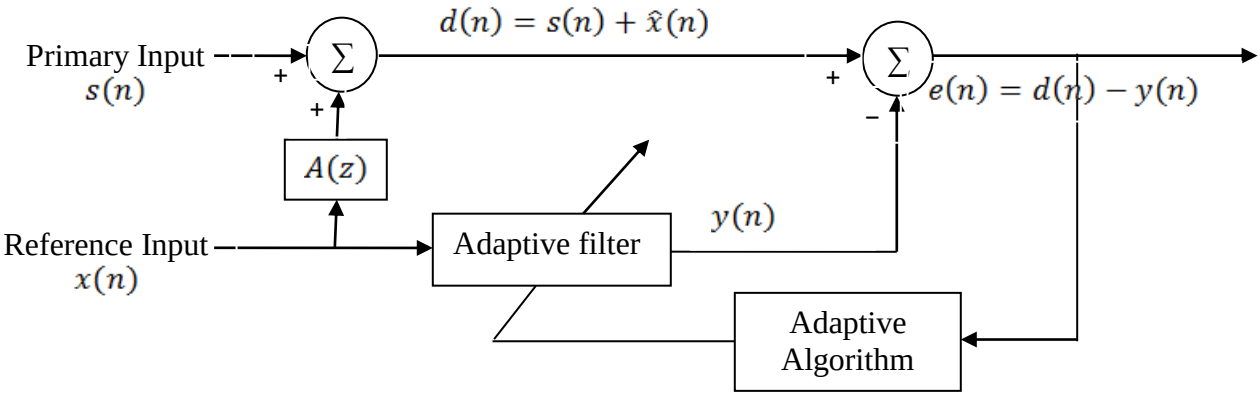


Figure 3.2: *Two-microphone adaptive noise cancellation (adapted from [54] with permission). See description in the text.*

The conventional two-channel adaptive enhancement method uses a second microphone, distinct from the primary microphone, to record the background noise and serve as a reference to enhance the noisy speech [54]. Two microphones should be placed far enough apart such that the reference microphone contains no or little speech, but also close enough so that it resembles as much as possible to the noise affecting the primary signal or target speech [54]. Figure 3.2 shows the block diagram of a two-microphone adaptive noise canceller. Due to the different locations of the two microphones, there is a delay, most likely variable, between the time of recording of the background noise by the reference and the primary microphones. Therefore, a filter is used to match the reference noise to the target noise recorded by the primary microphone [54]. The filter weights are adjusted by an adaptive algorithm. The adaptive process is controlled through a criterion (e.g. the minimization of mean-square error) in order to estimate a noise signal which is a best fit to the noise contained in the primary signal [133]. Then, the filtered reference noise is subtracted from the primary noisy speech signal to suppress the noise.

There are various kinds of adaptive algorithms to estimate the filter coefficients necessary to equalize the reference input, such as the Least Mean Square (LMS), the Normalized LMS (NLMS), and the Recursive Least Square (RLS) algorithms. The performance of the adaptive filter depends on the estimation algorithm so that, for real-time applications, it is important to use an algorithm which has a fast convergence rate and low computational complexity [133].

3.2.6 Direct waveform subtraction (DWS) method

The noise degrading speech signal can be removed using the same waveform of noise recorded separately (frozen noise) [117]. First, the background noise is played and recorded in the room while the subject is asked to be silent and still. Next, the subject produces the target speech in the same frozen noise while the noisy speech is recorded. The speech signal is then enhanced by subtracting the waveform of the reference noise in the first step from the waveform of the noisy speech in the second step. Figure 3.3 illustrates an example of the signal recorded in each step and the recovered clean speech waveform.

The size of noise reduction is highly affected by the time alignment of the noise waveforms from the two recordings. The reference and noisy speech signals must be synchronized and matched before the subtraction. Recording the reference noise in exactly the same conditions as the target noise in the compound signal can increase the amount of noise rejection. Random or voluntary head and body movements can also perturb the noise field around the talker and produce differences between the reference and target noises, which would have a damping effect on the size of noise reduction. Thus, the main disadvantage of this method is that it is only possible in controlled laboratory situations

where the noise can be played back with exactly the same waveform, not in real time applications. There are two main advantages to this method: (1) the size of noise reduction is independent of the speech level and (2) the speech signal is unaffected by this processing.

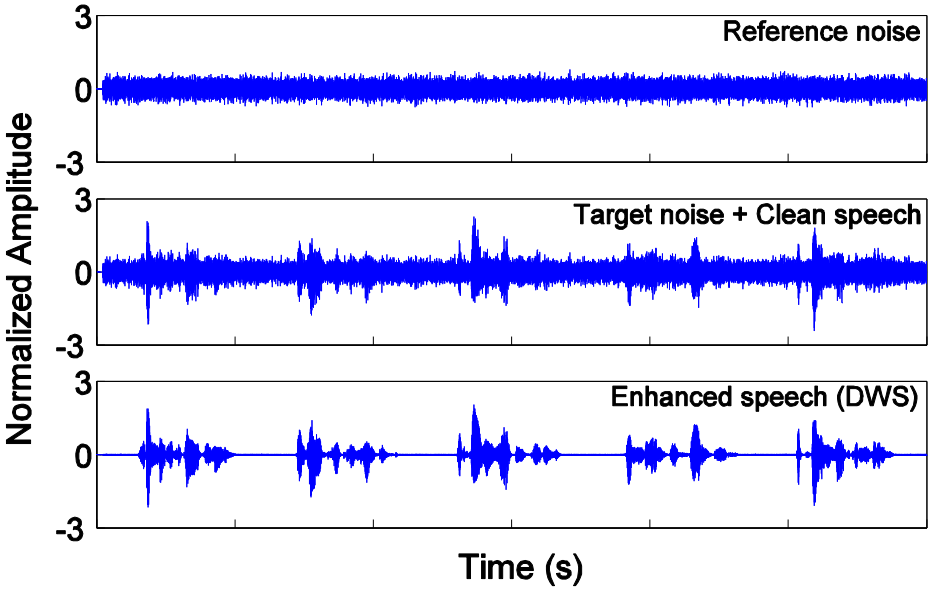


Figure 3.3: *Time-domain waveforms of reference SSnoise, noisy speech and the speech enhanced by direct waveform subtraction method (DWS).*

3.3 Pre-selection of speech enhancement methods

3.3.1 General framework

In the present research, the main task is to extract Lombard speech at a short distance from talkers reading a list of sentences in a noise field. The latter is generated by loudspeakers located around the talker in a test room with dimensions of 4.29 m × 3.65 m × 2.42 m (length × width × height) in the Hearing Research Laboratory at the University of Ottawa, as

shown in Figure 3.4. In this controlled-laboratory methodology, there is full control of the noise so that it can be replayed as needed. An appropriate speech enhancement method needs to be selected to remove as much as the noise as possible from the microphone recordings, while keeping the underlying speech waveform intact.

There are several enhancement methods that can be used to extract the clean Lombard speech from an external noise field, including one-channel methods (e.g. wavelet) as well as two-channel methods (e.g. ANC) as reviewed in Section 3.2.

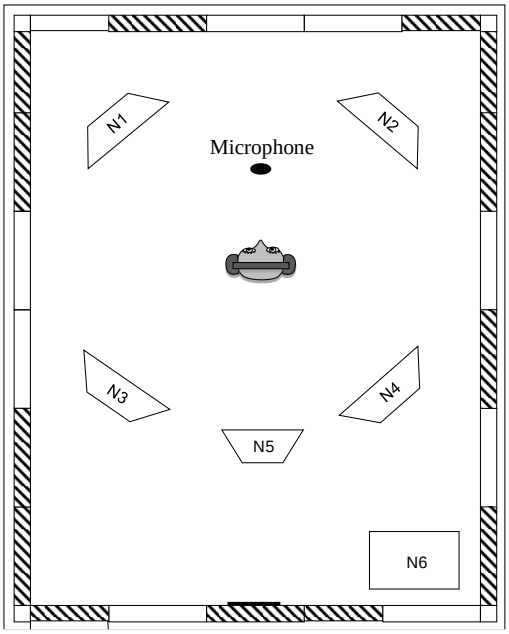


Figure 3.4: *Layout of the simulation room and loudspeaker configuration, N1–N6: Loudspeakers used to generate noise (adapted from [16] with permission).*

In this application, the enhancement methods selected do not need to be real time. However, they must possess a number of desirable features, as listed below:

1. method not too sensitive to the motion of the talker (fixed or dynamic);
2. method not too sensitive to the nonlinear distortion by loudspeakers;

3. method working in a wide range of noises (steady-state and fluctuating) and at different SNRs, both in terms of the NR achieved and the preservation of the speech waveform;
4. method not too sensitive to noise playback and audio recording settings; and
5. method not too sensitive to noise alignment issues (in the case of two-channel methods).

Based on these desirable features and the characteristics of common speech enhancement techniques surveyed in Section 3.2, four enhancement methods out of the six different methods described were preselected for the current application. Due to their inherent limitations (e.g. the appearance of musical noise and the necessary assumption of noise as stationary), spectral subtraction and Wiener filtering methods were discarded before any further measurements. The four methods retained included wavelet-based techniques, channel estimation (CE), adaptive noise cancellation (ANC) and direct waveform subtraction (DWS). Some in-laboratory tests were conducted to assess the performance of the selected methods against the desirable characteristics. The possible inputs available to the enhancement methods under test are the noise waveform as a computer file and/or as an acoustic noise recorded in the room. A noisy speech signal was simulated by adding a computer noise file and/or a noise recorded in a real measurement to the waveform of a clean speech signal selected from the American English Hearing-In-Noise Test (HINT) [131] at different SNRs, referred to as the initial SNR. The SNR is defined as the ratio of clean speech power to the noise power expressed in the logarithmic decibel scale.

3.3.2 Summary of findings by method

3.3.2.1 Wavelet method

The differences among the wavelet-based methods are related to the type of thresholding function, the threshold value, the type of wavelet transform, and the coefficients used to be thresholded. They include following methods:

- Wavelet 1: This method uses a time-scale adapted threshold and soft thresholding algorithm to compress the coefficients of wavelet packet transform (WPT) [132].
- Wavelet 2: This method uses a reference noise signal as a comparator for voiced/unvoiced detection and soft thresholding algorithm to compress the "detail" and "approximation" coefficients of discrete wavelet transform [48].
- Wavelet 3: This method uses SURE (stein unbiased risk) threshold selection method and soft thresholding algorithm to compress the coefficients of discrete wavelet transform in frames of noisy speech signal [134].
- Wavelet 4: This method is the same as Wavelet 3 except that it uses garrote thresholding algorithm [131].
- Wavelet 5: This method is the same as Wavelet 3 except that it uses step-garrote thresholding algorithm [131].
- Wavelet 6: This method is the same as Wavelet 3 except that it uses a new continuous differentiable threshold function proposed by Jia et al. [135].
- Wavelet 7: This method is the same as Wavelet 5 except that it uses the universal threshold value [131].

- Wavelet 8: This method uses the universal threshold and the continuous differentiable threshold function [131] to compress the subbands of wavelet packet decomposition.

By virtue of being a one-channel enhancement method, the wavelet approach has the advantage of possessing the desirable features 1, 2, 4 and 5 listed in Section 3.3.1. However, a major drawback of wavelet-based methods is that they may affect the fine structure and temporal envelope of the speech waveform in addition to removing the noise. As an example, Figures 3.5 and 3.6 illustrate the waveforms of noisy speech (simulated by adding a computer white noise to the waveform of clean speech), clean, and enhanced speech by methods of Wavelet 1 and Wavelet 3 at the initial SNR of −10 dB and 0 dB, respectively. When comparing the second (clean speech) and the third panels (enhanced speech) in Figures 3.5 and 3.6, it is obvious that the wavelet-based method also removes some components of speech information apart from the noise. This problem is such that it is not suitable to our purposes because it seems to cut the low energy portion of the speech waveform at SNRs well within the range projected for noisy Lombard speech.

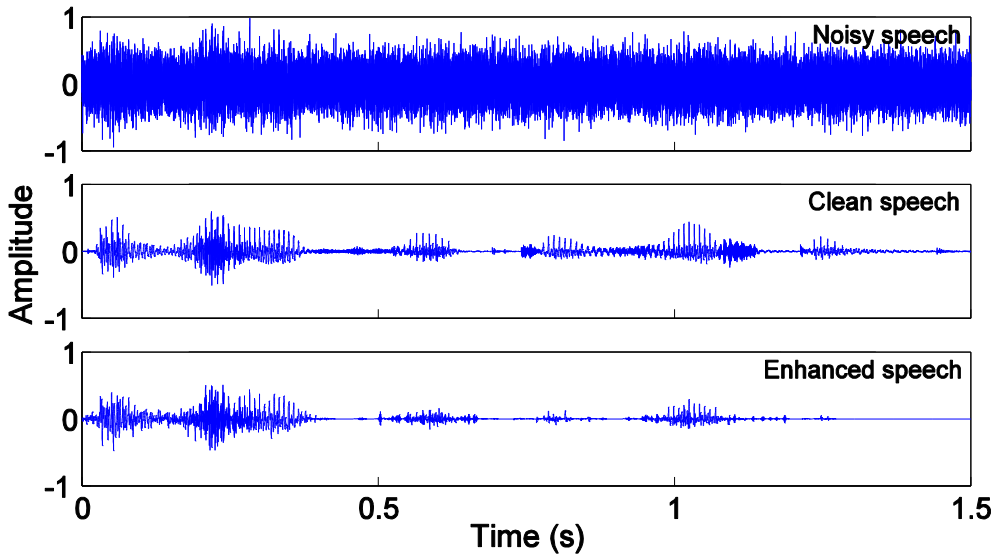


Figure 3.5: *The waveforms of noisy speech (SNR=−10 dB), clean speech, and enhanced speech by Wavelet 1 method. White noise is used.*

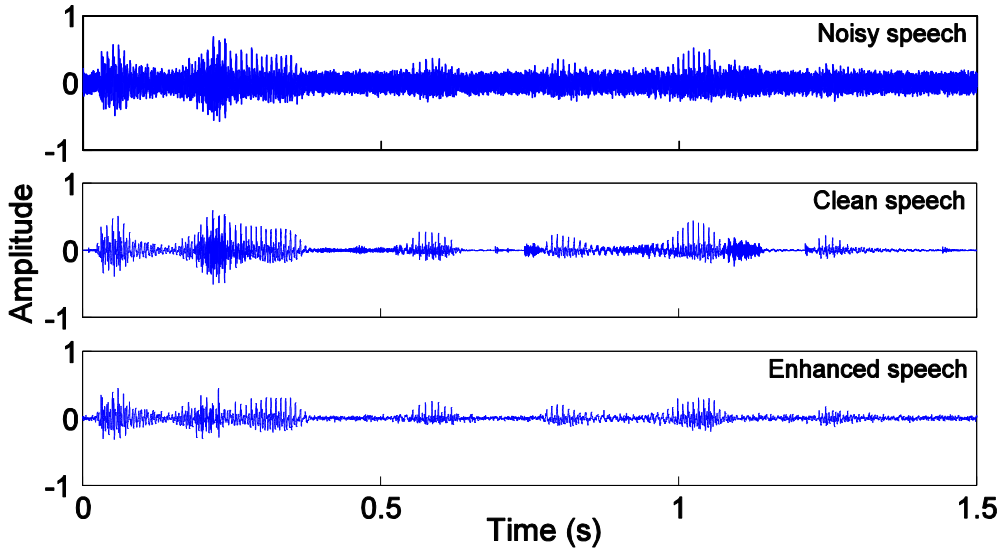


Figure 3.6: *The waveforms of noisy speech (SNR=0 dB), clean speech, and enhanced speech by Wavelet 3 method. White noise is used.*

3.3.2.2 DWS and ANC methods

The DWS and ANC methods are relatively insensitive to nonlinear distortion by loudspeakers (feature 2 listed in Section 3.3.1) since all inputs are acoustical from the room recordings. The ANC method is more immune to motion of the talker and misalignment (features 1 and 5 listed in Section 3.3.1) than DWS, as long as the speech in the second channel does not disturb the computation of the adaptive filter. Some tests were conducted to evaluate the effects arising from the motion of talker's body and the misalignment of noise waveforms on the DWS and ANC methods. In order to create noisy speech, the clean speech waveform was contaminated by noises recorded in the room rather than by computer noise files. This methodology makes it possible to have noisy speech more similar to that produced in real noisy environment. Moreover, the effects of changes in the noise samples, arising from any alteration in the acoustics of the room, on the size of noise reduction can be taken into account.

3.3.2.2.1 Playing and recording the noises with two different sets of equipment

In this section, the test is performed using the noise files recorded by two different sets of equipment; 1) sound card playing the noises in the room and 2) the sound recorder capturing the noises in the room. The noises were recorded in the noise simulation room (Figure 3.4) while an ANSI S3.36 acoustic manikin (B&K 4128) was placed in the room to simulate the sound reflection and diffraction effects around a talker immersed in noise.

The noise stimulus played in the room is a concatenated sequence consisting of six different noises of white, pink, brown, SSnoise, riveter, and helicopter. Each 12-second noise was separated by a 2-second pause from the subsequent noise. The noise file was played from a desktop computer and LynxOne sound card (sampling rate of 24 kHz) through six loudspeakers located around the manikin in the simulation room (Figure 3.4). To calibrate the individual noises, the level was set by changing the volume of amplifier (Yamaha RX-V1000) connected to the PC and then a sound level meter (B&K Type 2235 with omnidirectional microphone B&K Type 4176) was used to measure the level of noises at the centre location of the manikin's head. All individual noises were calibrated to 70 dBC.

Then, the omnidirectional microphone and sound level meter were located in front of the manikin's mouth and a sound recorder (Olympus LS10) was used to record the noises in the room at the sampling rate of 48 kHz. The reference position of the recording microphone is cm in front of the manikin’s mouth. Several other fixed positions and orientations of manikin's head and torso with respect to the recording microphone were tested to evaluate the effect of talker movement on the recording of the noise and the performance of noise reduction techniques. The experimental conditions for the manikin positioning and the label for each condition are listed in Table 3.1.

The noise generation (LynxOne sound card) and recording (Olympus LS10) equipment are independent in this experiment. There is no sampling rate synchronization between the two sets of equipment.

In order to create a noisy speech signal, the noise recorded in different measurements (e.g. M2a, M2b, M3a, M3b, and M3c) was added to the waveform of clean speech signal selected form HINT while noise M1 was used as reference noise. In preliminary tests, computer noise samples rather than noise recorded in the room (M1) were used as reference

noise. It was observed that the adaptive filter could not estimate the noise contained in noisy speech from the computer noise samples likely because of sampling rate mismatch between playback and record sound cards.

3.3.2.2.2 Results

Since the noises were played and recorded using two different sets of sound equipment, a MATLAB function was used to synchronize the noise measurements. It is essential for the DWS method that the reference noise be synchronized with the target noise contained in the noisy speech signal. The effect of misalignment on the size of noise reduction is evaluated by manually misaligning the noise waveforms after synchronization.

Analysis of the results showed that the DWS and ANC methods could provide a good size of noise reduction (NR) for all noise measurements shown in Table 3.1. As an example, Tables 3.2 shows the size of noise reduction for the noise measurement M2a (where the manikin's head was rotated respect to the reference position) at the "noise-only" and "speech-in-noise" conditions. The negative values for the size of NR represent that the noise is boosted compared to the initial noise available before using noise suppression methods. To misalign the reference noise with the target noise, M1 (reference) is kept fixed and M2a (target) is shifted. A negative shift means that the target noise appears earlier and a positive shift means that the reference noise appears earlier.

For the target SSnoise in position M2a, the size of noise reduced by the DWS method was 18.7 dB at lag zero, in both test conditions of "noise-only" or "speech-in-noise" (Table 3.2). As expected, the amount of noise reduction for DWS method is always the same for both "noise-only" and "speech-in-noise" conditions, since the noise reduction is independent

of the speech signal with this method. At lag zero, the size of SSnoise reduction by the ANC method in position M2a was 21.3 dB in "noise-only" and was 12.4 dB in "speech-in-noise" conditions (Table 3.2).

Table 3.1: *The manikin/talker positioning and the label for each experimental condition*

Label	Manikin's head orientation (degree)	Distance from manikin's mouth to microphone (cm)
M1	0	50
M2a	15	50
M2b	−15	50
M3a	0	49
M3b	0	45
M3c	0	40
M3d	0	25

It is observed from Table 3.2 at negative and positive lags that the DWS method is more sensitive to the misalignment of the reference noise and the target noise compared to the ANC method. The later method can compensate for the time mismatch as part of the algorithm. Moreover, the size of noise reduced by the ANC method is lower for the negative lags compared to zero or the positive lags. This is because the desired output of filter (target noise) for negative lags appears earlier than the reference noise (input of filter) and the filter becomes non-causal. When the lags are increased from zero to positive numbers, the noise reduction remains slightly unchanged or increases a little. This means that when the reference noise appears earlier than the original noise, the noise reduction remains high compared to zero lag. However, after a certain value of positive lag the noise reduction gradually goes back to an amount lower than at zero lag.

Table 3.2: *The size of reduction (dB) of SSnoise in M2a (target) by SSnoise in M1 (reference) in "noise-only" condition, and at SNR=0 dB in "speech-in-noise" condition (SR=48 kHz)*

Condition	method	Lags for misalignment (samples)						
		−10	−5	−1	0	+1	+5	+10
noise-only	DWS	0.2	4.2	12.5	18.7	17.3	5.0	0.6
	ANC	6.8	9.0	18.8	21.7	22.1	22.3	22.3
speech-in-noise	DWS	0.2	4.2	12.5	18.7	17.3	5.0	0.6
	ANC	3.0	5.3	12.8	16.2	16.6	16.1	15.8

3.3.2.2.3 Playing and recording the noises with one set of equipment

It was observed that the synchronization between the reference noise and the target noise has a considerable effect on the size of noise reduction. Thus, the noises are played and recorded simultaneously using the same sound card to further analyze the selected enhancement methods.

The noises were recorded in the same noise simulation room at the Hearing Research Laboratory while there was no manikin in the room. The noise signal played in the room was the same as that used in the test explained in Section 3.3.2.2.1. In contrast, the sampling rate of the computer noise file was resampled from 24 kHz to 48 kHz and 96 kHz. Then, it was played at 75 dBC through six loudspeakers located around the room from the same sound card (TASCAM US-366, TEAC Co., Japan) which would be used to simultaneously record the noises in the room. The output noises were recorded twice as target and reference noises. In order to have comparable data, the noises were also recorded by the two different equipment (LynxOne sound card and sound recorder Olympus LS10) used in Section

3.3.2.2.1 at the two sampling rate of 48 kHz and 96 kHz. In preliminary tests, it was found that using the computer noise samples rather than noise recorded in the room as reference noise did not work well since there might be a sampling rate mismatch between the prerecorded computer noise and the noise recorded by the TASCAM sound card.

3.3.2.2.4 Results

In these data, the first recording acted as the reference noise while the second recording was taken as the target noise to be cancelled. Tables 3.3 and 3.4 show the size of cancellation of target noises using DWS and ANC methods, respectively, in the "noise-only" condition. Using the same sound card at the sampling rate of 48 kHz increased the size of noise reduction by 5.6 dB and 0.4 dB averaged across different type of noises for DWS and ANC methods, respectively, compared to that obtained with different sound cards. The increases in the size of noise reduction were 7.3 dB and 2.0 dB, respectively, at the sampling rate of 96 kHz. This verifies that only one sound card which is able to play and record the noises simultaneously should be used in the measurements. There is a bigger effect on the size of noise reduction for the DWS method since it is more sensitive to the alignment between the target and the reference noises.

Moreover, it was observed that the size of noise reduction averaged over all noises increased by 5.6 dB and 8.2 dB for DWS and ANC methods, respectively, when the sampling rate increased from 48 kHz to 96 kHz.

Table 3.3: *The size of target noise reduction by DWS method in the "noise only" condition (measurements with different sound cards and the same sound card at two sampling rates (SR))*

Measurement condition	SR (kHz)	White	Pink	Brown	SSnoise	Riveter	Helicopter
Different sound cards	48	10.4	17.7	20.4	22.4	16.7	23.1
	96	17.0	21.6	20.6	24.6	22.8	24.2
Same sound card	48	21.8	23.6	20.3	26.2	25.0	24.2
	96	27.2	28.9	26.2	31.7	30.3	29.9

Table 3.4: *The size of noise reduction by ANC method in the "noise-only" condition (measurements with different sound cards and the same sound card at two sampling rates (SR))*

Measurement condition	SR (kHz)	White	Pink	Brown	SSnoise	Riveter	Helicopter
Different sound cards	48	22.5	27.7	30.1	35.2	29.3	33.9
	96	30.8	35.8	35.2	40.4	35.9	39.4
Same sound card	48	22.9	28.3	30.2	35.0	29.4	35.0
	96	35.4	39.3	39.6	44.8	37.1	33.5

For the "speech-in-noise" condition, Table 3.5 shows an example of the size of noise removed by ANC method at the SNRs of −4 dB and +4 dB. The size of noise reduced by DWS method is not provided at the "speech-in-noise" condition due to the fact that the noise elimination for DWS method is independent of the speech signal. Thus, the noise reduced by this method in the "speech- in-noise" condition is equal in size to that obtained in the "noise-

only" condition (See Table 3.3).

Similar to the "noise-only" condition at the sampling rate of 48 kHz, using one sound card for the noise measurements increased the size of noise reduced by ANC method by 1.7 dB at the SNRs of −4 dB and +4 dB (in speech-in-noise condition). At the sampling rate of96 kHz, it increased the size of noise reduction by 2.2 dB and 1.1 dB at the SNRs of −4 dB and +4 dB, respectively. Raising the sampling rate from 48 kHz to 96 kHz increased the size of noise reduction by 2.2 dB and 0.3 dB at the SNRs of −4 dB and +4 dB, respectively.

According to the results in both "noise-only" and "speech-in-noise" conditions, a more favorable noise reduction can be obtained by using only one sound card, playing and recording the noises simultaneously in the room, at the higher sampling rate of 96 kHz than using two sound cards (Section 3.3.2.2.1).

Finally, the DWS and ANC methods are retained among different enhancement methods because of the larger size of noise reduction at different SNRs over different types of noises. Moreover, neither method affects the speech waveform, in contrast to wavelet-based methods. In Section 3.4, the performance of the two selected speech enhancement methods of DWS and ANC are more extensively compared for recovering the clean Lombard speech.

Table 3.5: *The size of noise reduction (dB) by ANC method in "speech-in-noise" condition (measurements with different sound cards and the same sound card at two sampling rates (SR)).*

Measurement condition	SR (kHz)	SNR	White	Pink	Brown	SSnoise	Riveter	Helicopter
Different sound cards	48	−4	17.3	17.4	18.0	20.0	18.1	19.1
		+4	14.9	14.5	15.9	16.9	16.2	17.5
	96	−4	19.0	20.6	18.8	21.4	20.2	19.8
		+4	17.5	17.2	17.4	18.2	16.5	16.0
Same sound card	48	−4	19.9	19.9	18.6	21.3	20.9	19.3
		+4	17.6	17.2	16.6	19.0	18.7	16.8
	96	−4	22.1	21.2	21.1	23.4	22.3	23.1
		+4	19.6	17.1	18.7	17.2	17.9	16.9

3.3.2.3 CE method

With this method, the transfer function relating the computer noise file to the noise in the room is first estimated using a broadband noise such white noise. Then, at test time, the transfer function is used to estimate the noise disturbing the recorded noisy speech in the room.

When using two different sets of recording equipments as explained in Section 3.3.2.2.1, it was observed that the noise contaminating speech signal could not be correctly estimated from the reference noise filtered by the transfer function of the room. Indeed, as an example, the transfer function obtained by computer white noise file (input) and white noise in M1 (output) was not similar to that obtained by computer SSnoise file (input) and SSnoise in M1 (output). The transfer function should be independent of the type of noise since it reflects the

acoustic characteristics of the recording room. Moreover, it was observed that the transfer function is sensitive to misalignment even by 1 sample. Therefore, no good results could be obtained by the channel estimation method. This is likely because the computer noise file played into the room through the loudspeakers and the noise file recorded in the room are obtained from separate audio playback/recoding equipment, which can create sample rate synchronization problems. The channel estimation method is adversely affected by the quality of sound equipment and the changes in the acoustic configuration of the recording booth [118]. Distortion in the estimation of the transfer function over low and high frequencies is also one of the drawbacks of this method.

In the condition of playing and recording the noises by a single sound card as explained in Section 3.3.2.2.3, the white noise samples in the computer noise file and target noise were used to estimate the transfer function of the recording room. Table 3.6 shows the size of reduction of target noise in the "noise-only" condition.

According to Table 3.6, it was observed that the channel estimation method could not suppress the noises successfully except for SSnoise. Indeed, the transfer function estimated by white noise cannot be used as channel estimator to remove the other noises (pink, riveter, and etc.) recorded in the room. As an example, the transfer function estimated by white noise is compared to that estimated by pink noise in order to assess their differences and similarities. Figure 3.6.a illustrates the spectrum of the transfer function of the room estimated by white noise and by pink noise (computer noise file as input and target noise as output) at the sampling rate of 96 kHz.

Table 3.6: *The size of target noise reduced by the CE method when the channel (transfer function of the recording room) was estimated by white noise in computer noise file and target noise as input and output of channel (measurements with the same sound card at two sampling rates (SR)).*

Measurement condition	SR (kHz)	Pink	Brown	SSnoise	Riveter	Helicopter
A single sound card	48	2.54	−6.74	18.30	0.71	2.54
	96	4.00	1.27	23.93	0.93	2.84

The two transfer functions diverge considerably above around 11 kHz (Figure 3.7a). By zooming in on the curves in the frequency range of 0−11kHz, it can be seen that curves are in a good agreement with each other (Figure 3.7b). However, the curves differ from each other in the frequency range higher than 11 kHz (Figure 3.7c). The comparison between the transfer function estimated by white noise and that obtained by other noises recorded were also similar to that for white and pink noises. Thus, this implies that the transfer function of the room obtained for different input/output noises are not the same. As explained in Section 3.2.4, it might be due to the quality of sound equipment and the distortion in the estimation of the transfer function over low and high frequencies. As a result, this speech enhancement method is not retained.

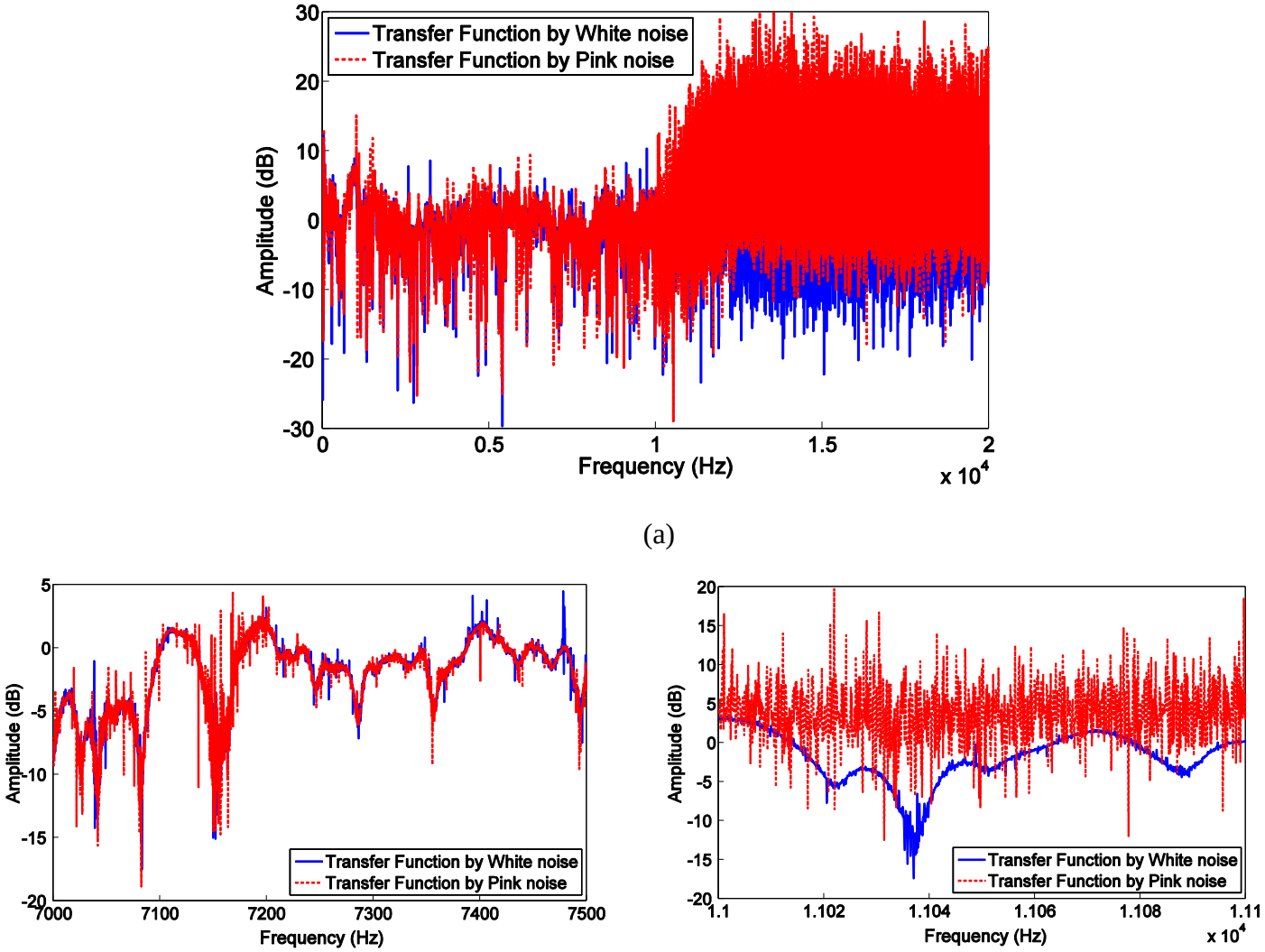


Figure 3.7: *a) The transfer function of recording room estimated by white noise and pink noise (sampling rate of 96kHz), b) Zooming in on a segment of the transfer function below 11kHz, c) Zooming in on a segment of the transfer function above 11kHz.*

3.4 A comparison of speech enhancement methods to extract Lombard speech in an external noise field

In this section, two noise suppression techniques, DWS and ANC, are used to extract the clean simulated Lombard speech from the noisy voice recorded in the external noise field of a simulation room. Apart from the noise recording in the fixed position of the manikin, several noise measurement are performed while the manikin's head and torso are rotated and relocated respect to the recording microphone to simulate the talker's body movement. This more realistic simulated experimental approach allows evaluating the effects of body movement on the recovery of clean speech signal. The size of noise reduced by two methods is compared in six different types of stationary and fluctuating noise conditions. Two basic features of speech, pitch and energy, and two objective intelligibility measures based on the speech intelligibility index (SII) and the speech transmission index (STI) are extracted from the clean and enhanced Lombard speech to assess the effectiveness of the two noise suppression methods. Results show both methods can recover the speech features at SNRs down to at least −10 dB, which make them suitable to study Lombard speech from real talkers under typical conditions in an external noise field.

3.4.1 Experimental Protocol

The experiment was conducted in the noise simulation room in the Hearing Research Laboratory at the University of Ottawa. An ANSI S3.36 acoustic manikin with artificial mouth (B&K 4128) was used to create controlled conditions simulating a talker reproducing

the same speech signal in noise and in quiet. This allows a direct comparison of the recovered speech following noise suppression against the clean speech measured in quiet.

To generate an external quasi-diffuse noise field, the background noise was played over 4 tower loudspeakers, a bookshelf speaker and a subwoofer located in the simulation room [16]. The acoustic manikin was placed near the middle of the loudspeaker arrangement, and a sound level meter (B&K Type 2235) with a free-field microphone (B&K Type 4189) was positioned 50 cm away from the mouth. The noise signals were played and recorded simultaneously by an external sound card (TASCAM US-366, TEAC Co., Japan) at a sampling rate of 96 kHz.

Two conditions were examined in this experiment: "noise-only" and "speech-in-noise". In the "noise-only" condition, repeated measurements (M1-M5) were made of the frozen noise samples recorded in the room at different positions of the manikin:

- M1 and M2: The manikin is in the reference position (head at 0 degrees with respect to the microphone),
- M3: The manikin's head is rotated to the right side by 30°,
- M4: The manikin is moved closer toward the microphone by 10 cm,
- M5: The manikin is translated to the right side by 10 cm.

The noise signal M1 was used as the reference noise and the repeated recording (M2) in the same reference position of the manikin represents an ideal situation with no talker movement. The recording of noises M3-M5, where there are small changes in the manikin’s position with respect to the reference position, allows simulating a more realistic situation in which the talker moved slightly between recordings.

In the "speech-in-noise" condition, speech was produced through the mouth of the manikin located at the reference position and recorded in quiet conditions with the same

sound card and microphone position used for the noise recordings. The speech material is a 5-sentence list selected from HINT, and the clean speech recorded was added to each noise recording (M2-M5) from the "noise-only" condition to simulate speaking in noise. In this work, the level of speech is set at different levels to evaluate the performance of the noise suppression methods over a wide range of signal-to-noise ratios (SNR) ranging from −30 dB to +20 dB.

The background noise signals chosen for this study were white, pink, International Female Fluctuating Masker (IFFM), speech-spectrum noise from the International Female Noise (IFnoise), riveter noise, and helicopter noise all presented at a level of 70 dBA. The IFFM consists of nonsense speech generated by reordering and concatenating speech segments obtained from the International Speech Test Signal (ISTS) which includes six female voices speaking in different languages [136].

3.4.2 Data Analysis

The size of the noise reduction is calculated by subtracting the power of residual noise after enhancement from the power of the initial noise. The higher the value in dB, the more noise is removed from the initial noisy signal.

The pitch, characterized by the fundamental frequency (F0), is an important attribute of voiced speech that is related to the quasi-periodic oscillation of vocal cords. Numerous algorithms have been proposed to robustly estimate the pitch due to the importance of this parameter in different fields of speech processing [137]. In this work, PRAAT (v.5.3.55), a computer software specialized in speech analysis, was used to estimate the average pitch of

clean, noisy, and noise-suppressed signals. In addition, the energy of the signal represented by the square root of the mean value of the squared signal (or RMS level), was selected as another speech feature to assess the performance of noise suppression methods in recovering the clean from the noisy speech.

The SII and STI are the most commonly used objective measures for predicting speech intelligibility and have also been used to evaluate the effectiveness of noise cancellation methods. Like the SII, the STI is an index between 0 and 1 that is highly correlated with the intelligibility of speech. However, unlike the SII, the STI accounts for the effects of temporal and nonlinear distortions on speech intelligibility [138]. The speech-based STI has been introduced which allows direct estimation of the intelligibility of speech in a noisy environment [139, 140]. The conventional SII and STI models (ANSI/ASA S3.5-1997 R2012, IEC 60268-16-2011) have been validated for stationary noises. For time-varying noisy conditions, extensions have been developed (Extended SII (ESII) and short-time STI) in which the SII and STI are averaged over short time frames of speech signals [83,140].

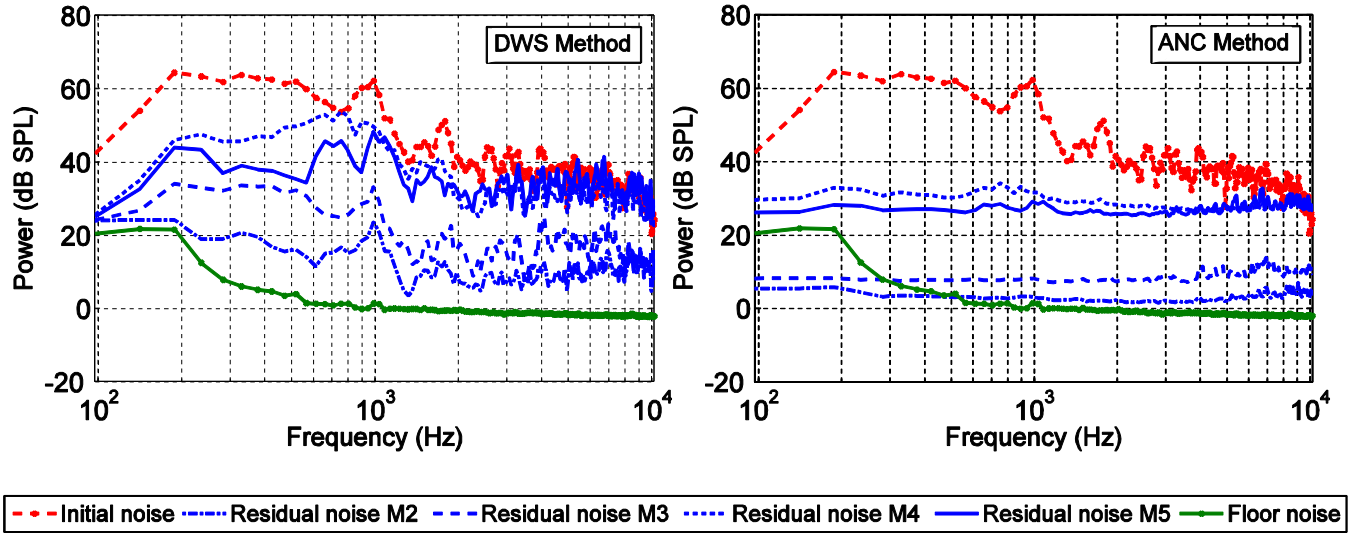
3.4.3 Results

3.4.3.1 Noise reduction in "noise-only" condition

The size of noise reduction in the "noise-only" condition, where there is no speech added, can be seen in Figure 3.8a and Figure 3.8b which show the initial noise and the residuals obtained with the two noise suppression methods, namely direct waveform subtraction and adaptive filtering, respectively. Speech-spectrum noises (SSnoise) recorded for different positions of the manikin (M2-M5) were chosen as target noises to be cancelled while

SSnoise in M1 was selected as the reference noise. Cancelling M2 by M1 represents the ideal situation in which the talker is kept still for the two recordings. The target noise to be cancelled (M2) and the reference noise (M1) are very close in this case and, as shown in Figure 3.8, the residual noise has the least spectral power (dotted-dashed curves) compared to all other choices of target noises (M3-M5). A noise reduction of up to 40 dB with direct waveform subtraction and 60 dB with adaptive filtering is achieved at low frequencies.

The influence of changes in the manikin's position is reflected as a reduction in the amount of noise suppression, as illustrated in Figure 3.8 by the elevated residual noise curves for the situations where M3, M4 and M5 are used as target noises. Inspection of Figure 3.8 and Table 3.7 shows that the adaptive method was able to suppress more noise in all conditions compared to the direct waveform subtraction method, since the former can adaptively make the reference noise to be very close to the noise to be cancelled when no speech is present.



(a) (b)

Figure 3.8: *The initial speech-spectrum noise recorded in the manikin's reference position (M1) and the residual noises obtained by (a) direct waveform subtraction, and (b) adaptive enhancement method when M1 is used as reference and different noises as target noises (M2-M5). Noise-only condition.*

Table 3.7: *Overall noise suppression (in dB) of speech-spectrum noise recorded at different positions of manikin (M2-M5) by two different suppression methods when M1 is used as reference noise. Noise-only condition.*

Target noise	Direct waveform subtraction	Adaptive method
M2	36.2	42.0
M3	27.9	37.1
M4	9.7	19.1
M5	13.9	19.7

Table 3.8: *Overall noise reduction (in dB) for speech in noise M2 and M3 at different SNRs when M1 is used as reference noise.*

	DWS method						ANC method					
SNRs	−30	−20	−10	0	+10	+20	−30	−20	−10	0	+10	+20
Target noise												
M2	36.2	36.2	36.2	36.2	36.2	36.2	34.8	30.5	25.9	20.9	15.6	9.4
M3	27.9	27.9	27.9	27.9	27.9	27.9	31.1	28.6	24.9	20.4	15.3	9.2

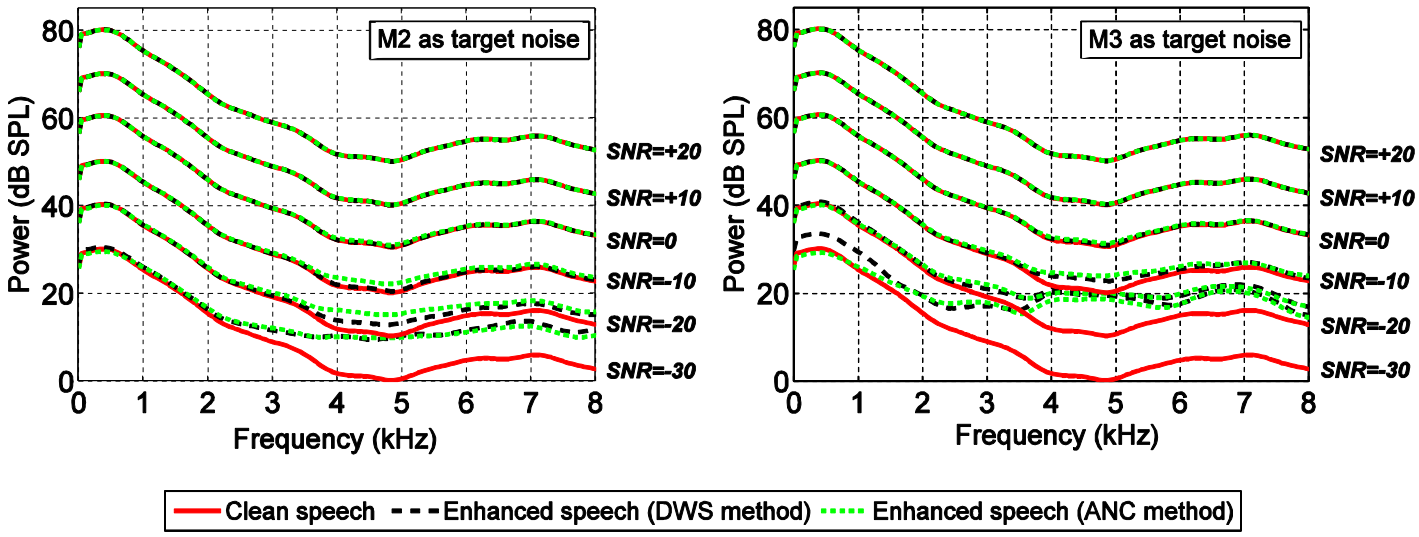
3.4.3.2 Noise reduction in "speech-in-noise" condition

In order to simulate a realistic "speech-in-noise" condition with or without movement in the talker's body during speaking, the clean speech produced by the manikin's mouth was added to the SSnoise in M2 and M3 at different SNRs. Again, the two noise suppression methods used the noise M1 as the reference. As expected, the amount of noise suppression by direct waveform subtraction is not affected by the presence of speech (Table 3.7 vs. 3.8). In contrast, the noise suppression by adaptive filtering is lower in the "speech-in-noise" condition (Table 3.8) compared to that in the "noise-only" condition (Table 3.7), in particular for higher SNRs. This is because the target noise (e.g. M2 or M3) is summed up with the clean speech signal in this case, and so the ability of the adaptive filter to match the reference noise (M1) to the target noise (M2 or M3) is adversely affected by the presence of speech. For lower SNRs, the noisy speech is dominated by the target noise so that the "speech-in-noise" condition is more similar to the "noise-only" condition. Figure 3.9 compares the clean speech and the enhanced speech by direct waveform subtraction and adaptive methods for different SNRs (-30, -20, -10, 0, +10, and +20). There is a good

agreement between the enhanced speech and the clean speech for SNRs down to about -20 dB in the best case scenario of no talker movement (M2, Figure 3.9a) and about -10 dB with movement (M3, Figure 3.9b). Although the adaptive method provides a smaller amount of noise reduction at high SNRs compared to low SNRs (Table 3.8), the noisy speech signals at high SNRs are already clean enough so that the enhanced signals can closely follow the clean speech. The noise suppressed by direct waveform subtraction and adaptive filtering at low SNRs (−30 to −10 dB) is better when there is no movement in the manikin (Figure 3.9a vs. Figure 3.8b). Specifically, Table 3.8 shows that when M2 is added to clean speech, which simulates a situation of no movement, the size of noise reduction for direct waveform subtraction increases by 8.3 dB compared to the situation where movement is simulated (i.e. M3 is added to the speech). For the adaptive method, there is 1−3 dB increase in noise suppression at low SNRs for the M2+speech situation compared to M3+speech.

3.4.4 Evaluation of enhanced speech

The two above-mentioned noise suppression methods were used to enhance the speech contaminated by six different background noises in configuration M3 at a SNR of 0 dB. Figure 3.10a shows that the average energy values extracted from the enhanced speech signals are very close to the energy of clean speech, which is 73.8 dB SPL (red line). In fact, the two methods allowed the true energy (73.8 dB SPL) to be recovered within 0.1 dB, while speech energy estimated from the noisy speech is overestimated by 2.6 to 3.6 dB across noises.



(a) (b)

Figure 3.9: *The clean and noise-suppressed speech enhanced by two methods at different SNRs, with speech-spectrum noise M1 as reference while (a) SSnoise M2, and (b) SSnoise M3 is added to the clean speech.*

The average pitch values of enhanced speech are also very close to the average pitch of clean speech at 113.4 Hz (Figure 3.10b). The unusually high value of pitch (173.5 Hz) for speech contaminated by IFFM noise is because this noise is generated from female speech which has a higher pitch than the male talker of the target HINT speech. The two methods allowed recovering the pitch within 1.3 Hz of the true pitch, while the pitch estimated for the noisy speech is overestimated by 4.1 to 60.1 Hz across noises.

Finally, Figure 3.11 shows the ESII and short-time speech-based STI intelligibility ratings for the noisy speech and speech enhanced by the two noise suppression methods. At an SNR of 0 dB, the ESII of the noisy speech ranges from 0.4 to 0.7, and the STI ranges from 0.4 to 0.5, across noises. Under perfect noise-free recovery of the speech signals, the intelligibility ratings are expected to reach 1 in each condition. As shown, there is a

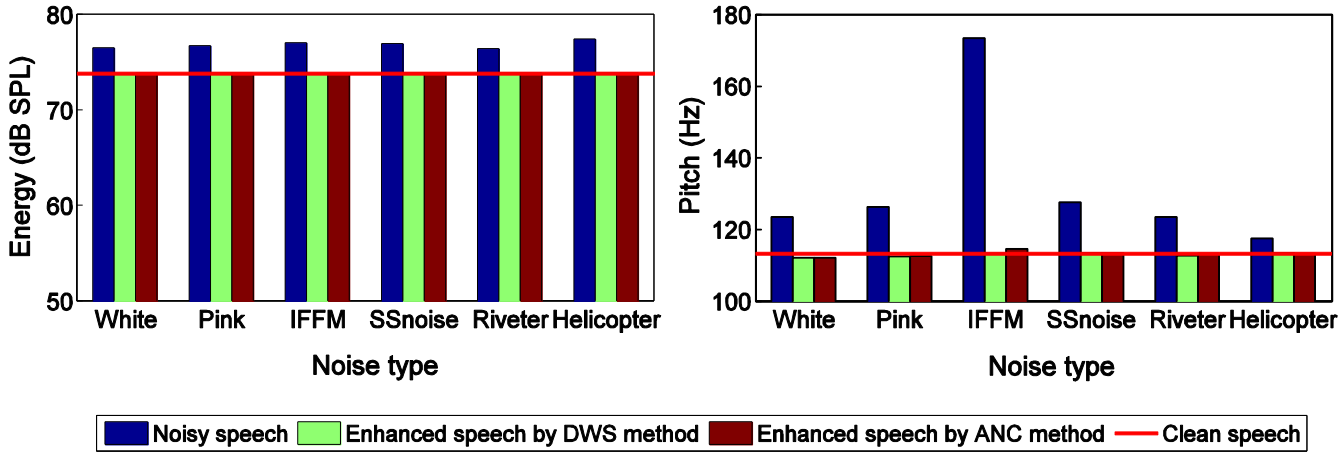
considerable increase in intelligibility ratings when using the two noise suppression methods, with ESII and STI values on the order of 0.9, indicating that both methods successfully recovered most of the audible cues masked by the noises.

3.4.5 Discussion and conclusion

There is little research investigating the Lombard effect in an external sound field, likely due to the difficulty of extracting the clean speech from the vocal output contaminated by the background noise. The available studies mostly evaluate the Lombard effect elicited by noise exposure through headphones, even though they may have a dampening effect on the vocal level. In this study, the use of two noise suppression methods was investigated to estimate the clean Lombard speech in an external diffuse noise field with different noise conditions.

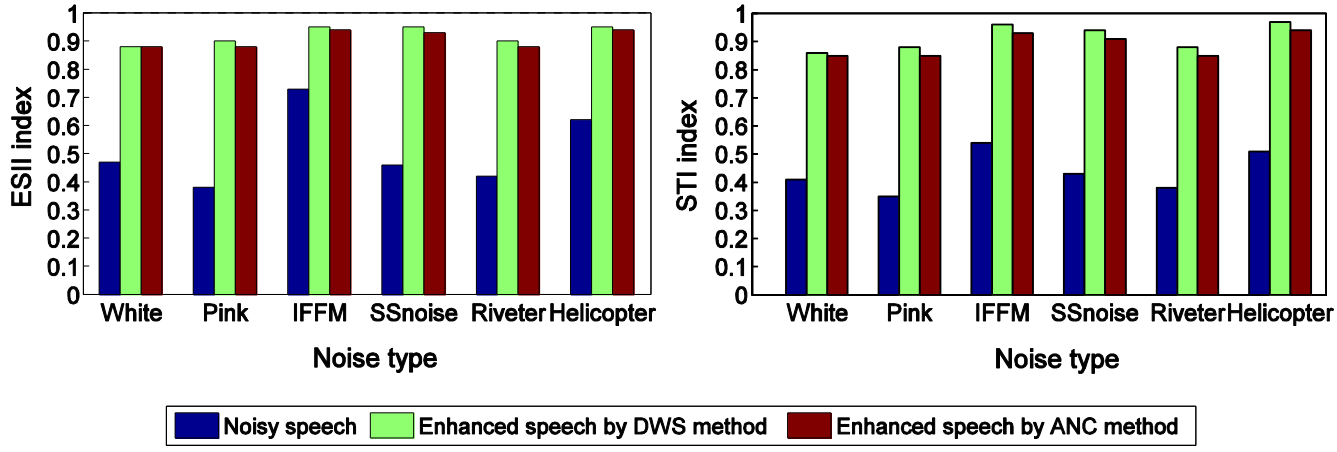
According to the results obtained in the "noise-only" condition, the largest amount of noise reduction using the direct waveform subtraction and adaptive filtering methods is achieved in the situation simulating no movement in the talker's body (M2), which ensures that the reference noise is very similar to the target noise to be suppressed. In the direct waveform subtraction method, differences between the target noise (M3-M5) and the reference noise result in a decrease in noise suppression. The adaptation procedure used in adaptive filtering to match the reference and target noise signals is also affected by the talker's movement. In the "noise-only" condition, adaptive filtering performed better than direct waveform subtraction because it is able to match the reference noise to the target noise. It can be seen that adaptive filtering can even suppress the low-frequency noise from the ventilation system in the room in situations where there is no or little movement in the

manikin relative to the reference position (residual noise signals in the M2 and M3 configurations in Figure 3.8b).



(a) (b)

Figure 3.10: *(a) Energy and (b) pitch for noisy and enhanced speech signals obtained with the two noise-suppression methods, at SNR=0 (M3 is added to clean speech and M1 is used as reference noise).*



(a) (b)

Figure 3.11: *(a) ESII and (b) Short-time speech-based STI for noisy and enhanced speech signals obtained with the two noise-suppression methods, at SNR=0 (M3 is added to clean speech and M1 is used as reference noise).*

The results of the "speech-in-noise" condition show that the performance of adaptive filtering decreases markedly when the SNR increases because the speech dominates the noisy speech signal, making it difficult for the reference noise to be adapted to the target noise mixed with the speech. Both suppression methods can successfully enhance the speech signals contaminated by background noise at SNRs above −10 dB. At low SNRs of −20 dB and −30 dB, the recovery of speech is affected at mid to high frequencies, especially when there is movement in talker's body.

The similarity between pitch and energy of enhanced speech and those of clean speech shows that the two noise suppression methods were able to extract these parameters in different types of noises, fluctuating and stationary, at an SNR of 0 dB. Moreover, the improvement in speech intelligibility, expressed by ESII and short-time speech-based STI metrics, clearly supports the effectiveness of these two methods in noise suppression. The results of this study therefore show that substantial noise reduction can be achieved using the two evaluated enhancement methods for the SNRs above −10 dB which is the most likely SNR range of Lombard speech produced in an external noise field, as reported by Brungart [117].

Chapter 4

Evaluating Noise Cancellation Methods for Recovering the Lombard Speech from Vocal Output in an External Noise Field

4.1 Introduction

Talkers are able to adjust their vocal effort with respect to their primary communicative needs, for example to be heard by listeners in noisy environments. This reflex phenomenon, referred to as the Lombard effect, was first described by Etienne Lombard in 1911 [56, 80, 91]. Under the Lombard effect, the articulatory movements and the acoustic-phonetic characteristics (e.g. level, pitch, formant frequencies, and spectral tilt) of speech are also

modified compared to normally-produced speech [76, 77, 78, 141]. In noise, the Lombard effect improves the effective signal-to-noise ratio (SNR) and leads to an improvement in the audibility and intelligibility of spoken words.

It has been found that a noise level below 45 dB SPL does not significantly affect the level of speech. Indeed, the Lombard effect as a mechanism to overcome the masking effect of noise typically appears for noise levels over 45-55 dB SPL [66, 68]. The level of speech rises in response to an increase in the level of noise, but up to the maximum human vocal effort which corresponds to about 100 dB SPL [69]. The size of the Lombard effect also varies depending on the temporal and spectral characteristics of the background noise [101, 109, 101]. In fluctuating noise, the weaker masking in the gaps, where there is little or no energy, has a damping effect on the Lombard effect. The wearing of hearing aids, hearing protectors or pre-existing hearing loss interfering with the perception of auditory feedback of one's own voice also affect the vocal effort in noise [40, 41, 117, 142].

There has been extensive research on the Lombard effect where noise has been presented through headphones worn by the talkers [83, 91, 101, 105, 106, 107]. Although such an experimental approach allows recording speech in an ideal noise-free environment, the vocal level of talkers wearing headphones can be reduced as a result of disruption in their self-monitoring auditory feedback [143]. The occlusion effect arising from the blockage of the ear canal by the headphones causes the talkers to perceive their own voice through bone conduction as boomy and louder at low frequencies [117, 118], which can affect speech production. As a consequence, results from such studies may not accurately reflect the effects of different factors, such as type and level of noise, in real life situations when talkers are immersed in noise from the surrounding environment. Furthermore, the use of headphones to elicit the Lombard effect is impractical when the talker wears hearing

protectors or hearing aids. Therefore, it is necessary to elicit the Lombard effect with an external noise field (e.g. loudspeakers) to ensure talkers have a realistic perception of their own voice and the surrounding noise in interaction with any hearing device worn, similar to that experienced in real-world environments [77, 79, 117, 118, 144]. However, the use of an external noise field introduces a new challenge because of the contamination of the recorded Lombard speech by the eliciting noise. The background noise contained in the recorded speech signal diminishes the accuracy of features extracted from the speech waveform. Therefore, noise suppression techniques need to be utilized to extract the clean Lombard speech from the recorded noisy vocal output.

The conventional denoising approaches include spectral subtraction, Wiener filtering, and wavelet-based methods [54, 127, 145]. The serious drawbacks of most noise suppression techniques include the necessity of a priori knowledge of the noise, the distortion of useful components of speech, and the appearance of artifacts called "musical noise" [145]. For our purpose, it is important to use a method that maximizes the noise rejection while introducing no or little distortion over the processed speech signal to maintain the speech characteristics as intact as possible.

A noise cancellation method called channel estimation has been proposed to extract Lombard speech parameters from talkers immersed in noises generated by two loudspeakers [118]. This method of eliciting the Lombard effect was used to make the vocal loading situation more realistic and to avoid a disruption of the talkers' auditory feedback when using headphones. Channel estimation was performed by simultaneously recording the noise signal twice as it was presented to the participant: (a) digitally at the input to the loudspeaker system (input noise), and (b) acoustically using a microphone inside the recording booth (output noise). The recordings included an initial segment of estimation noise followed by

worksite noises. A transfer function for the channel of the audio system and the recording booth including the participant's body was estimated from the initial segment of the two recordings. The input noise signal was filtered through this transfer function and the noise-suppressed speech was obtained by subtracting the filtered noise from the noisy voice output. Using the approach, the noise could be reduced by up to 40 dB under an ideal situation where there is no motion in the talker's body. The channel estimation method has some inevitable sources of error limiting the size of noise reduction, which include the quality of sound equipment, the distortion in the estimation of the transfer function over low and high frequencies, and changes in the acoustic configuration of the recording booth arising from the talker's body movement. It was shown that small movements of 1 cm and 5 cm in the talker's head to one side decreased the size of noise reduction by 3 dB to 20 dB over different frequencies. Sensitivity to the displacement of the talker's body is critical since the estimated transfer function is not adaptable to static or dynamic perturbation in room acoustics such that a movement, even as small as one centimetre, can remarkably deteriorate the size of noise reduction.

Brungart et al. (2012) used the so-called direct waveform subtraction method to recover the speech waveform from talkers wearing earplugs in different levels (50, 60, 70, and 80 dB SPL) of pink noise generated via 16 loudspeakers [117]. They stated that this technique makes it possible to evaluate the real impact of hearing protectors on speech production and to measure speech output in an actual noise field. For this purpose, a noise-only recording (frozen noise) was subtracted from the speech-plus-noise recording to recover the speech-only signal. At the highest noise level of 80 dBA, the noise suppression method could recover speech signals well below the level of noise. Although this method is easy and simple to use, inadaptability to the perturbation of the noise over time makes it sensitive to

talker body motion. Moreover, this method is only useable in a laboratory or in situations where the noise can be reproduced exactly.

In an experiment carried out to collect the Lombard speech in a noisy testing room simulating a moving vehicle, an adaptive noise cancellation method and then a Wiener filter were used to remove the background noise captured by the talker's microphone [144]. In order to avoid any disruption in the talker's voice feedback, they used loudspeakers instead of headphones to generate the noisy acoustic ambiance of a driving vehicle in a simulation room. The loudspeaker noise signals recorded by an in-ear microphone were used as the reference noise to design the adaptive filter. The adaptive filter could decrease the power of the noise by about 15 dB. The advantage of this method is the ability to adapt to the static and dynamic changes in the acoustics of the room (e.g. due to talker's body motion).

Giguere et al. (2006) used an updated version of the conventional wavelet shrinking method with time and scale adaptation of the threshold proposed in [132] to enhance speech produced in noise [87]. They extracted the Lombard effect in different types of continuous and fluctuating noise including babble, restaurant, white and speech-spectrum noise [87]. However, a possible problem with wavelet-based methods is the musical noise artifact.

While the above-mentioned noise suppression methods are promising for Lombard speech extraction, there is a general lack of knowledge about their sensitivity to the recording method (e.g. distance and type of microphone) and the talker's static and dynamic motion as well as the effective range of noise levels or SNRs. In Chapter 3, we evaluated the effect of the talker's body positioning, simulated by static displacements of a manikin's head and torso [126]. One recording setup with the omnidirectional microphone at 50 cm from the manikin's mouth was tested to assess the change in the size of noise reduction due to static shifts of the torso (laterally and towards the microphone) and head rotations. It was observed

that the largest amount of noise reduction using the direct waveform subtraction and adaptive filtering methods was achieved in the situation simulating no change in the position of the manikin's torso and head. Moreover, both suppression methods could successfully enhance the speech signals contaminated by background noise at SNRs above −10 dB. At lower SNRs, the recovery of speech was affected at mid to high frequencies, especially when there was a change in the manikin's head and torso position.

The aim of the present research is to study the performance of enhancement methods for Lombard speech extraction in more realistic experimental conditions where a real participant is used to simulate dynamic sound field disturbances arising from body movements rather than static displacements by an acoustic manikin. The dynamic disturbance of the noise field may affect differently the performance of enhancement methods that are fixed in nature like the direct waveform subtraction method versus those which are adaptive. Moreover, the effects of recording condition such as microphone distance (e.g. 50 cm and 25 cm) and the type of microphone (e.g. omnidirectional and cardioid directional) on the size of noise rejection will be evaluated. A shorter distance of microphone to the talker's mouth will increase the speech level by 6 dB if distance is halved but it may be more susceptible to noise field disturbances if the participant is moving while talking. Likewise, directional microphones may reject 4−6 dB of noise in a diffuse sound field but they typically have higher internal noise and are more prone to nonlinear distortion [146, 147]. Thus, the performance of the enhancement methods that have been proposed to extract clean speech in the noisy sound field need to also take these factors into consideration. In addition to calculating the size of noise reduction, some basic characteristics of speech (e.g. pitch and energy) and objective quality and intelligibility measures were also used to verify the effectiveness of the two enhancement methods in recovering the clean speech.

4.2 Methods and materials

4.2.1 Noise suppression methods

4.2.1.1 Direct Waveform Subtraction (DWS)

Noise in a degraded speech signal can be suppressed by subtracting the same waveform of noise recorded separately (frozen noise) [117]. A reference noise is first played and recorded in the room while the participant is asked to be silent and still. After that, the same noise waveform is replayed while the participant is talking. In order to extract clean speech, the waveform of the reference noise recorded in the first step is subtracted from the waveform of the noisy speech signal (speech plus target noise) recorded in the second step. Figure 4.1 illustrates an example of the reference noise, noisy speech signal, and the recovered clean speech waveform.

There are two main advantages to this method: (1) the size of noise reduction is independent of the speech level and (2) the speech signal is unaffected by this processing. However, the size of noise reduction is highly affected by the time alignment of the noise waveform in the first and second steps. As such, the reference and noisy speech signals must be carefully synchronized and matched before subtraction. Therefore, recording the reference noise in exactly the same conditions as the target noise in the noisy speech signal is crucial to achieve a significant amount of noise suppression. Thus, a limitation of this method is the natural movements of the talker's body or head while speaking which can affect the acoustical conditions and cause differences in the noise waveforms recorded in the two steps of processing. Another limitation in the size of the noise reduction is the possible presence of

nonlinear distortion of the recording equipment which would result in a mismatch of the reference and target noise waveforms.

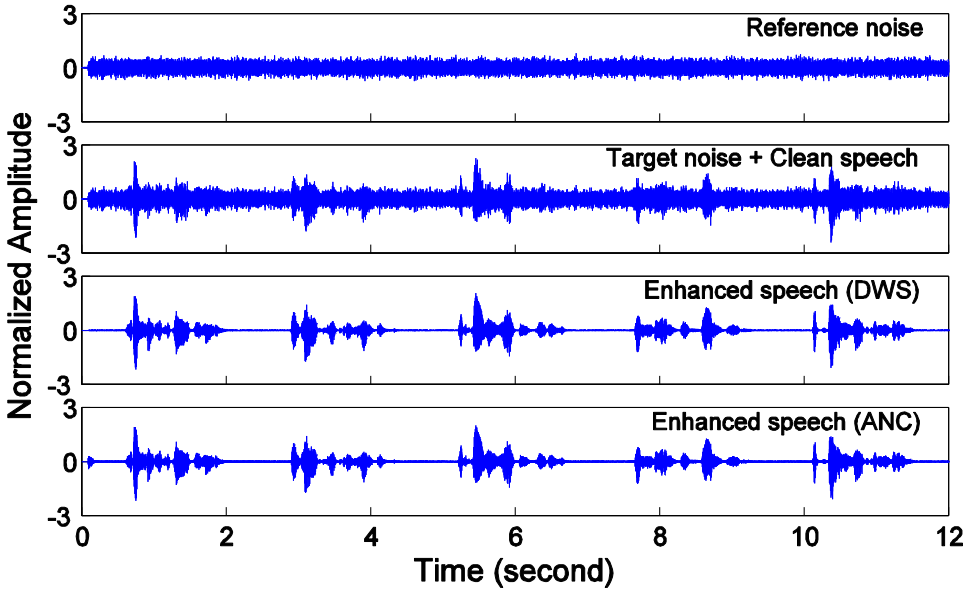


Figure 4.1: *Example of speech enhancement by the direct waveform subtraction and adaptive noise cancellation methods. First panel: reference noise, Second panel: noisy speech (SNR=0 dB), Third and Fourth panels are the speech enhanced by the DWS and ANC methods respectively.*

4.2.1.2 Adaptive noise canceller (ANC)

The adaptive noise cancellation method seems to be more effective than other filtering techniques for speech enhancement since it is able to adapt to the time-varying alterations of noise contained in the noisy speech signal [133]. The conventional adaptive noise canceller requires a second microphone, in addition to the primary one picking up the talker's voice, to separately record the background noise as reference [54]. In a typical real-time application,

the two microphones should be placed far enough such that the second microphone contains no or little speech, but also close enough so that the reference noise resembles as much as possible the target noise contained in the signal recorded by the primary microphone. Due to the different locations of the two microphones, the reference noise signal must be filtered to compensate for the alterations in noise samples relative to the target noise. The filter weights are adjusted by an adaptive algorithm. The adaptive process is controlled through a criterion (e.g. the minimization of mean-square error) in order to estimate a noise signal which is a best fit to the noise contained in the primary signal [133]. Then, the noise-suppressed speech is obtained by subtracting the filtered reference noise from the primary noisy speech signal.

In a controlled laboratory situation, or when the exact same noise waveform can be reproduced, a two-step approach to recording the primary and the second (reference) signals using the same microphone can be envisaged to achieve better performance. Similar to the direct waveform subtraction method, the first step is to record the reference noise alone and the second step is to record the speech in the same target noise. This two-step approach helps to reduce the mismatch between the reference and target noise signals and ensures that the reference signal is free of speech. Figure 4.1 illustrates an example of enhanced speech obtained by subtracting the filtered reference noise from the noisy waveform.

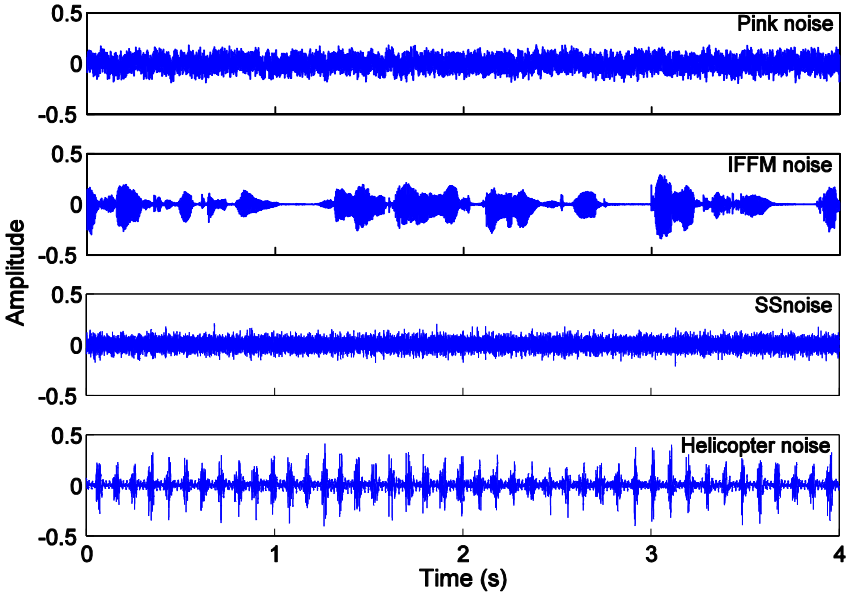
An advantage of this method is that it can adapt to static and dynamic perturbations in the recording of noise signals due to changes in the acoustical conditions (e.g. motion in the talker's body or head). However, the performance of this method can decrease at high SNRs because the noise estimation becomes more difficult when the target noisy signal is dominated by speech [133]. The LMS algorithm, with different step size values depending on the SNR, was used to adaptively update the filter coefficients.

4.2.2 Speech and noise material

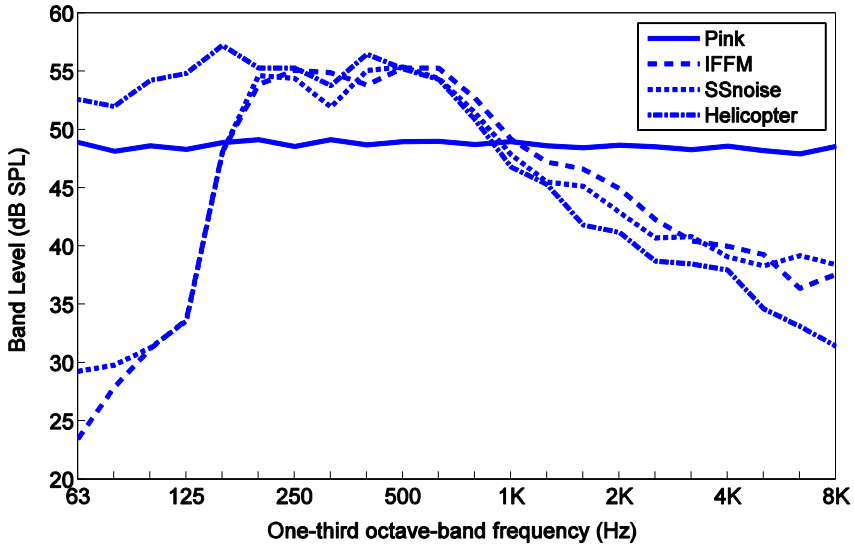
In this study, the two speech enhancement methods were tested in four different types of noise covering different temporal and spectral characteristics: pink noise, International Female Fluctuating Masker (IFFM), speech-spectrum noise (SSnoise) from the International Female Noise (IFnoise), and helicopter noise. The IFFM consists of nonsense speech generated by reordering and concatenating speech segments obtained from the International Speech Test Signal (ISTS) which includes six female voices speaking in different languages [135]. The IFnoise is generated by randomly overlapping speech material from the ISTS multiple times keeping the long-term average spectrum unchanged. Figure 4.2 shows the time-domain waveforms and the spectra of these noises.

The noise stimulus played in the room was a concatenated sequence consisting of pink, IFFM, IFnoise (SSnoise), and helicopter noise. Each 12-second noise signal separated by a 2-second pause from the subsequent noise in the stimulus signal was presented at a level of 70 dBA at the talker's ears.

The speech material is a sequence of 5 sentences from the American English Hearing-In-Noise Test (HINT) [148]. The length of the concatenated sequence of 5 sentences with 1 second pause between each pair is 11 seconds.



(a)



(b)

Figure 4.2: *(a) The time-domain waveforms, and (b) one-third octave-band spectra of the four noises used in the study*

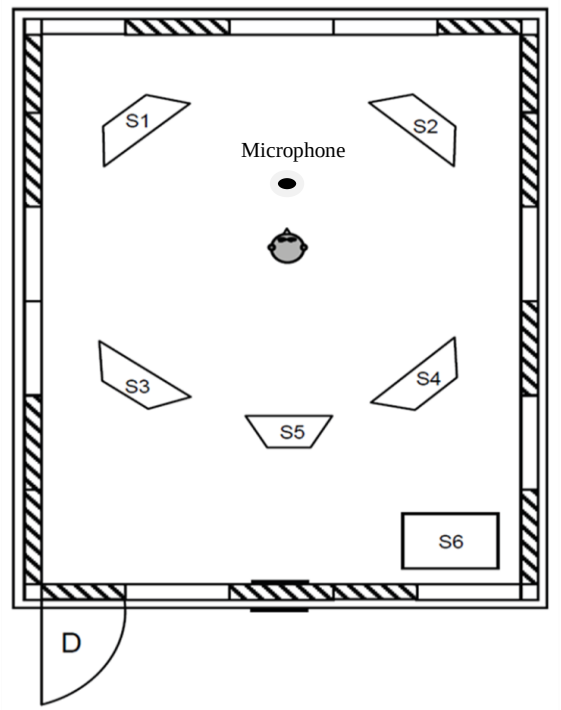


Figure 4.3: *Recording set up in the measurement room. The head icon represents a manikin when recording speech and a real person when recording noises. Four acquisition configurations were investigated using two microphone types (omnidirectional and directional) and two microphone distances (25 and 50 cm). (adapted from [16] with permission). See also Table 4.1.*

4.2.3 Experimental setup

The experiment was conducted in a sound-treated room with the dimension of 4.29 m × 3.65 m × 2.42 m (length × width × height) in the Hearing Research Laboratory at the University of Ottawa (Figure 4.3). A computer (PC) in the control room outside the recording booth sends the noise to an external sound card (TASCAM US-366, TEAC Co., Japan) (Figure 4.4). The noise is then spectrally equalized to compensate for the room response via a programmable multiprocessor (RANE RPM26z) which communicates with the sound card. The multiprocessor outputs the equalized noise signal, split into four channels. As shown on

the right side of Figure 4.4, these signals were then passed through a power amplifier (Yamaha RX-V1000) and sent to a set of loudspeakers located inside the simulation room which includes four tower loudspeakers (Sony MF600H), a bookshelf speaker on a stand (Yamaha NS-C155) and a subwoofer (Mirage Sub12) (Figure 4.3). The sound card (TASCAM US-366) was also used to simultaneously record the noise signals in the room at a sampling rate of 96 kHz. The noise was recorded while there was a real participant near the center of the recording room (Figure 4.3). The noise samples were collected by two different microphones at two different distances of 50 cm and 25 cm from the mouth of the participant in separate acquisition runs. The two microphones were a free-field omnidirectional microphone (B&K Type 4189) connected to a sound level meter (B&K Type 2235) and a cardioid microphone (Sennheiser ME 64).

The speech signals were collected separately by positioning an acoustic manikin (Head and Torso Simulator, B&K Type 4128) at the same location in the room as the real participant (Figure 4.3). The HINT speech sentences were sent through the output channel of the RANE multiprocessor, amplified (Yamaha RX-A820) and then played through the mouth simulator of the acoustic manikin, as shown on the left side of Figure 4.4. The acquisition configurations (microphone types, distances from the mouth) used to record the speech signal from the artificial mouth of the manikin were the same as those used for recoding the noise in presence of the real participant. This strategy made it possible to create speech-in-noise signals at any desired SNR by off-line mixing of speech and noise recordings.

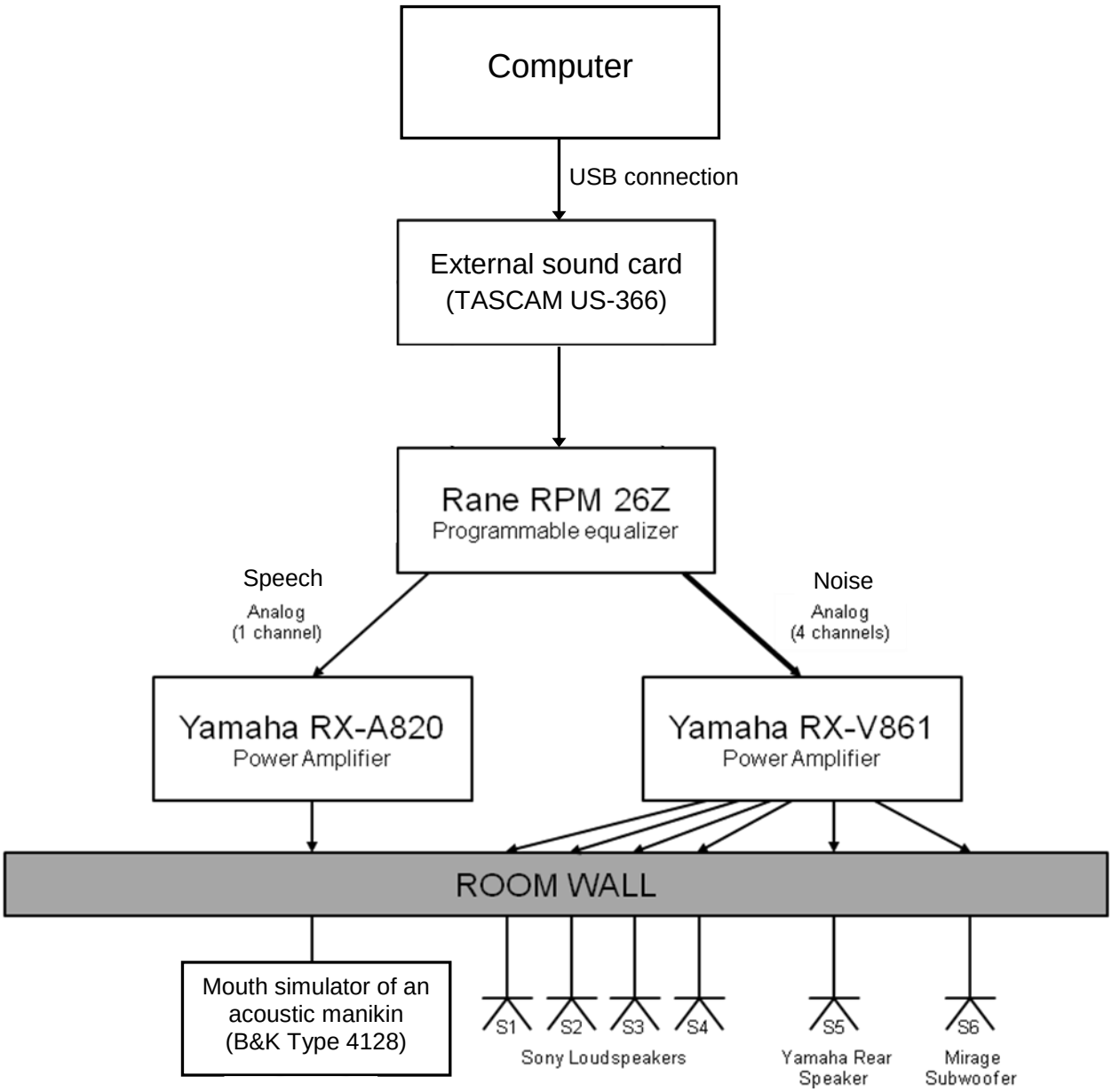


Figure 4.4: *Test set-up for playing and recording noise and speech signals in the room, (adapted from [149] with permission).*

4.2.4 Acquisition procedure

All noises were calibrated to a level of 70 dBA at the participant's head position. For each noise type (pink, IFFM, SSnoise, and helicopter) and acquisition configuration (Omni.50: omnidirectional mic. at 50cm, Omni.25: omnidirectional mic. at 25cm, Dir.50: directional cardioid mic. 50cm, Dir.25: directional cardioid mic. 25cm), three noise recordings were then made (see "Noise" column in Table 4.1). During the first and second trial (R and T1), the real participant was asked to stay still to allow recording of a reference noise and an ideal target noise to create a best case scenario for noise suppression. In the third trial (T2), the participant was asked to dynamically move (torso, head, lips, etc.) in a way that would mimic speaking to a person in face-to-face situation, but while remaining silent. This trial represented a more realistic situation where the talker's body moves, thereby disturbing the noise field and creating a less favorable scenario for noise suppression.

The comparison of enhanced speech signals is reliable and practical only when the same speech signal is uttered across all acquisition configurations and noise conditions, which is not possible with real participants. In order to have exactly the same speech waveforms in each condition, the HINT sentence speech files were played through the manikin's mouth and recorded in the same four acquisition configurations, at 50 cm and 25 cm and by two different microphones in quiet (see "Speech" column in Table 4.1). To generate noisy speech test signals at different SNRs, the clean recorded HINT speech with variable amplitude was then added to each noise recording in T1 and T2 conditions fixed at 70 dBA (Table 4.1). This process ensured that the effects of speech enhancement method, acquisition configuration, noise type and sound field disturbance by the participant motion could be compared with the same speech signal waveform.

TABLE 4.1: *Noise and speech recording trials used to prepare the test signals. Speech and noise were recorded separately. Speech signals in trial T1 and T2 were from the same recordings.*

Trial	**Noise**		**Speech**	
	Participant	Recordings	Manikin	Recordings
Reference (R)	Still in noise field	Omni50, Dir50 Omni25, Dir25	No speech	-
Ideal Target (T1)	Still in noise field	Omni50, Dir50 Omni25, Dir25	Speech from artificial mouth	Omni50, Dir50 Omni25, Dir25
Realistic Target (T2)	Natural movements of body and head	Omni50, Dir50 Omni25, Dir25	Speech from artificial mouth	Omni50, Dir50 Omni25, Dir25

4.2.5 Performance measures

The size of the noise reduction (NR) for the T1 and T2 recordings was calculated by subtracting the power level of the residual noise after enhancement from the power level of the target noise before enhancement. The residual noise was obtained by subtracting the waveform of clean speech from that of enhanced speech. The higher the value in dB, the more noise was removed from the recorded noisy signal.

Several speech characteristics and objective intelligibility and quality measures were also calculated for clean and enhanced speech to evaluate the performance of the DWC and ANC methods in recovering the clean speech from the noisy speech. The energy of signal was

reported in sound pressure level (dB SPL). The closer the energy of enhanced signal to that of clean speech indicates the better is the noise suppression.

Different algorithms in time, frequency and time-frequency domain have been proposed to precisely estimate the pitch period quantified by the fundamental frequency (F_0). In low SNR conditions, mostly below 0 dB, the estimation of pitch is challenging because of the degradation of pitch estimator performance [150, 151]. In this work, PRAAT (v.5.3.55), a specialized environment for analysis of speech, was used to estimate the average pitch of clean, noisy, and noise-cancelled signals. PRAAT calculates the pitch based on the highest value of the autocorrelation function. The non-zero lag related to the highest value of autocorrelation is taken as a candidate for the pitch period [152].

The speech intelligibility index (SII) and speech transmission index (STI), standardized by ANSI S3.5-1997 and IEC 60268-16-2011 respectively, are the most commonly used objective measures for predicting speech intelligibility [153, 154]. The SII is estimated by averaging the weighted signal-to-noise ratios (SNRs) in different frequency bands of the speech and noise spectrum [140, 155, 156]. The STI uses a modulation transfer function (MTF) rather than SNR to predict speech intelligibility. Based on the assumption that the envelope of a speech signal must be perceived correctly in order the speech to be understood correctly, the STI calculates the preservation of the envelope modulation in a degraded environment (e.g. background noise) [157]. The conventional STI cannot be used when the purpose is comparison between two speech signals in order to detect any degradation arising from different speaking styles or speech enhancement [157]. To account for these effects on intelligibility, the speech-based STI has been proposed in which the degraded speech signal is compared with the original undistorted speech. The SII and STI are indices between 0 and 1, a value highly correlated with the subjective speech intelligibility scores. The two

measures have been validated for stationary noises but they cannot predict intelligibility accurately and reliably in fluctuating noises. Extensions have been developed based on the idea that instantaneous SII or STI computed over short time frames of signal can detect short-time changes in intelligibility arising from the fluctuation in the background noise [111, 138,139, 140].

The perceptual evaluation of speech quality (PESQ), standardized by ITU-T as recommendation P.862 in 2002, is extracted from the fine structure information of the speech waveform. It is estimated based on the comparison between the clean and the degraded speech signals in a perceptual space [158, 159, 160]. The perceptual model transforms the time-aligned clean and degraded speech signals from time-amplitude domain to frequency-loudness domain. Then, the PESQ score is estimated by comparing the differences between the loudness densities of both signals. It is an index between −0.5 to 4.5, or in most cases between 1 to 4.5, correlating with subjective quality scores (e.g. subjective "Mean Opinion Score" (MOS)) [160, 161, 162].

4.3 Results

4.3.1 Noise reduction performance of the two speech enhancement methods in the omnidirectional acquisition configuration at 50 cm

Figures 4.5.a and 4.5.b illustrate the noise reduction achieved as a function of frequency with the DWS and ANC methods in a reference acquisition configuration (omnidirectional microphone at 50 cm) for the ideal motionless condition (T1) and for the more realistic condition (T2) of the participant's body in one noise (SSnoise), when no speech is present (SNR = $-\infty$). It can be seen that the size of noise cancelled by both methods is reduced when there was movement in the participant's body during noise measurements (Figure 4.5b) compared to the ideal condition (Figure 4.5a). The residual noise obtained by DWS method follows somewhat the initial noise spectrum, but in the case of ANC method it is clear that the algorithm distributed the residual noise nearly equally over frequency (Figure 4.5). Moreover, it is observed that the DWS method cannot remove the noise to a level less than the room noise, but the ANC method could even reduce the room noise at lower frequencies. Figure 4.5 thus illustrates the potential benefits of ANC method in terms of a higher noise reduction relative to that for the DWS method at very low SNR, in both ideal and realistic conditions of target noise recording.

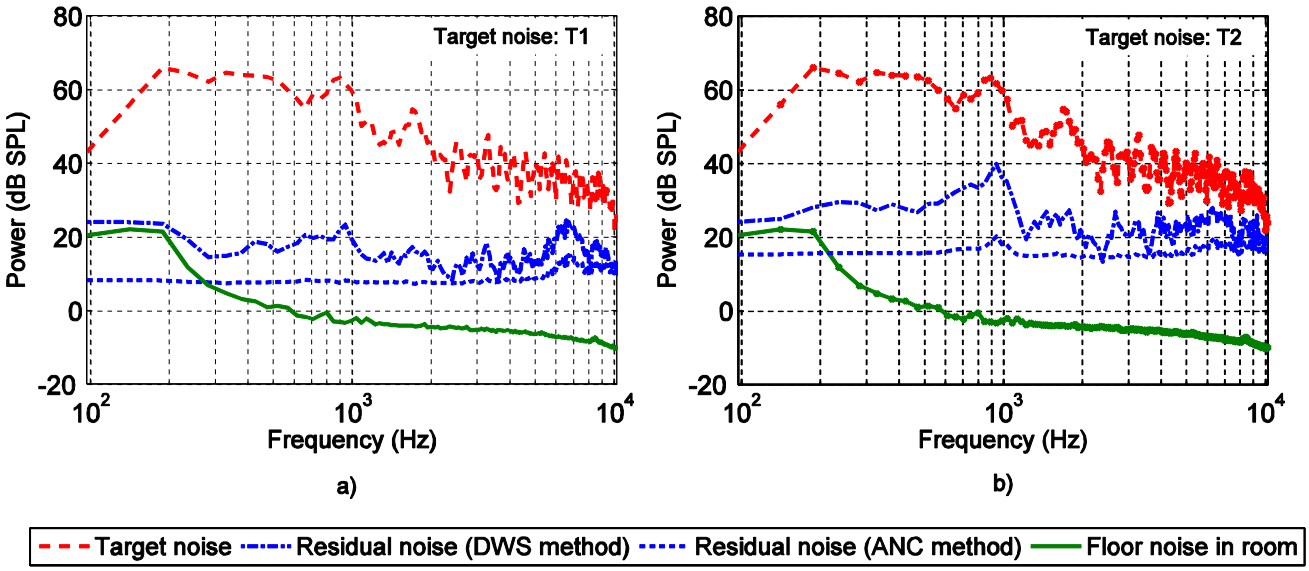


Figure 4.5: *The target noise, residual noise obtained by DWS and ANC methods at SNR of minus infinity (i.e. no speech), and the room floor noise for real participant in the simulated a) ideal situation (T1) of motionless talker and b) realistic situation (T2) of talker motion. The signals were recorded by the omnidirectional microphone at 50cm and SSnoise is used.*

Figure 4.6 shows the size of noise reduction (NR) by the DWS and ANC methods in the baseline acquisition configuration (omnidirectional microphone at 50 cm) for the ideal motionless condition (T1) and for the more realistic condition (T2) of the participant's body at different SNRs ranging from $-\infty$ (NS: no speech) to +20 dB. It illustrates that the size of noise reduced by the DWS method does not change with the SNR value since this method is independent of the presence of the speech. For the ANC method, the size of noise reduction decreases when the SNR increases from $-\infty$ to +20 dB. Estimating the target noise with a high amplitude speech signal using the reference noise is adversely affected at high SNRs.

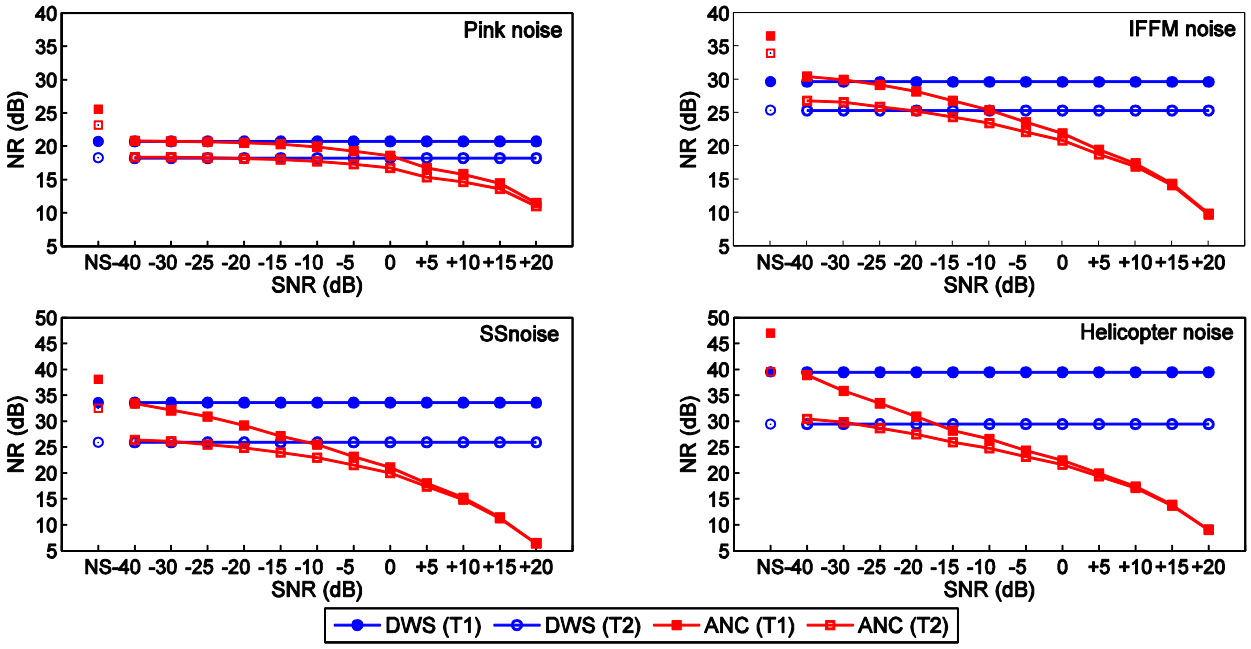


Figure 4.6: *The size of noise reduction by DWS and ANC methods for the real participant in the simulated a) ideal situation (T1) of motionless talker and b) realistic situation (T2) of talker motion. The signals were recorded by the omnidirectional microphone at 50 cm. (NS: no speech, SNR = −∞, NR: noise reduction).*

Moreover, it is observed that the amount of noise removed decreases when comparing realistic (T2) situation to the ideal motionless (T1) situation of talker's body in different types of noise. It is observed that the largest size of reduction, no matter whether ideal (T1) or real (T2) situation, is achieved for the low frequency (Figure 2.b) and fluctuating helicopter noise, while the lowest size corresponds to the high frequency pink noise. Overall, the size of noise reduction by the DWS method is superior to that by the ANC method in almost all situations above a SNR of −25 dB, in this case where there is a perfect time alignment of reference and target noises. The ANC method is better only in extremely low SNR or when there is no speech. The degradation effect of the talker's movement on noise reduction is larger for the DWS compared to the ANC method. The adaptive algorithm in the ANC

method is able to compensate for the changes in the target noise arising from body movement relative to the reference noise.

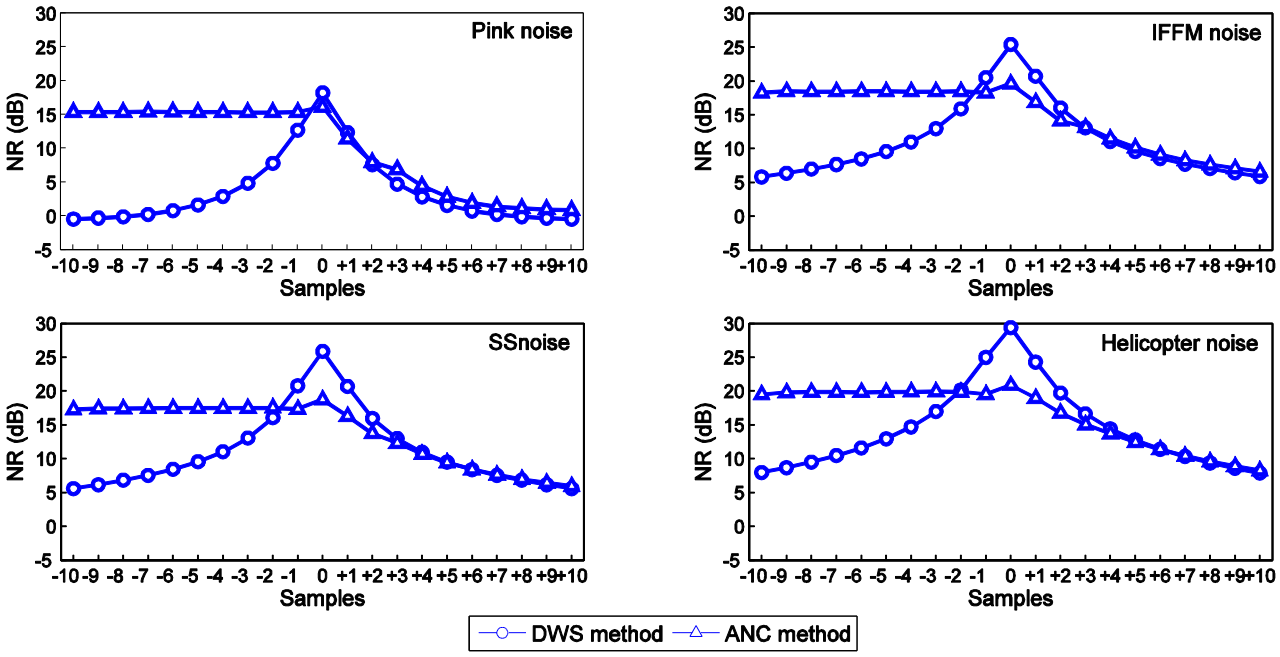


Figure 4.7: *The size of noise reduction by DWS and ANC methods at 0 dB SNR when the reference and noisy speech (T2 condition) are misaligned (lag of −10 to +10 samples). (Positive samples mean the reference appears later than noisy speech while negative samples mean it appears earlier than noisy speech).*

Figure 4.7 illustrates the effect of misalignment between the reference noise and the noisy speech (T2 condition) on the size of noise reduction by the DWS and ANC methods at an SNR of 0 dB. The performance of the enhancement methods decreases even with 1 sample misalignment. The larger drop in the size of the noise reduction by the DWS method for different delays shows that it is more sensitive to misalignment compared to the ANC method. For the negative values of delay, the ANC method is able to adaptively estimate the target noise contained in noisy speech from the reference noise samples which are present

earlier than the target samples. Thus, the size of noise reduction remains close to the value at zero lag. At positive lags, however, the noise reduction gradually degrades with increasing delay. This effect is due to the non-causality of the adaptive filter when the reference noise appears later than the noisy speech (desired output of filter).

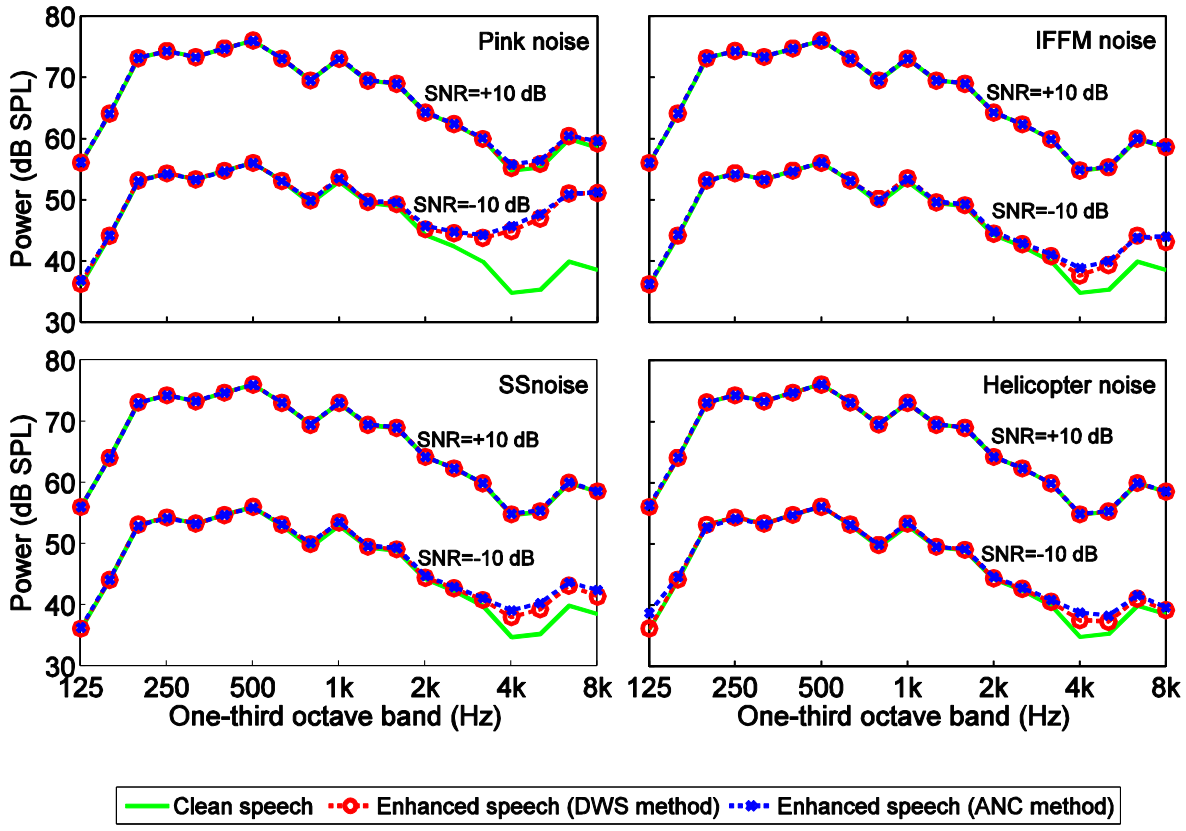


Figure 4.8: *The one-third octave band level of clean and enhanced speech signals obtained by DWS and ANC methods with four different noise types in the realistic T2 condition (SNRs= −10 dB and +10 dB). All signals were recorded with the omnidirectional microphone at 50cm.*

The spectra of speech signals recovered by the DWS and ANC methods at one third octave band frequencies for different noise types in the realistic T2 condition are shown in

Figure 4.8 using perfect time alignment. It is apparent that there is a good agreement between the clean speech and the enhanced signals at SNR +10 dB for all noise conditions. For the low SNR of −10 dB, the recovery of speech varies at frequencies higher than 2.5 kHz depending on the type of noise. The best recovery of speech was achieved in the fluctuating helicopter noise, largely because it has less high frequency energy compared to the other types of noise (Figure 4.2), followed by that in the fluctuating IFFM and the stationary SSnoises, while speech recovery in the pink noise was the least effective because it has more energy at high frequencies.

4.3.2 SNR improvement in the four acquisition configurations

A shorter distance from 50 cm to 25 cm provides a theoretical improvement of 6 dB for the speech level. A directional cardioid microphone has a theoretical directivity index of 4.8 dB which reduces the recorded noise and improves the SNR. In practice, the exact benefits of a shorter distance and directional microphone depend on the sound reflection properties of the room surfaces, noise type and noise spectrum. Table 4.2 shows the size of SNR improvement at recording arising from the type of microphone (omnidirectional or directional) and relocating the microphone to the half of the distance with respect to a reference recording at 50 cm with omnidirectional microphone for different noise types. The omnidirectional microphone at 50 cm serves as the baseline acquisition configuration in each noise. Compared to this baseline configuration, the biggest size of benefit is observed for the directional microphone at 25 cm. The SNR benefit is related to both the directivity index of

the microphone (decreased noise) and the shorter distance of microphone to the mouth (increased speech level).

TABLE 4.2: *The improvement in SNR (dB) resulting from the acquisition configurations. The omnidirectional microphone at 50 cm serves as the baseline acquisition configuration and so the SNR improvement in the first column is zero.*

Mic. type (distance to the mouth)	Noise Type			
	Omni.(50cm)	Omni.(25cm)	Dir.(50cm)	Dir.(25cm)
Pink	0	6.9	0.0	6.8
IFFM	0	6.8	3.7	9.3
SSnoise	0	6.4	3.5	9.0
Helicopter	0	6.9	3.7	9.0

Omni.: omnidirectional mic., Dir.: directional cardioid mic

It is observed that the SNR improvement arising from the shorter distance is a little larger than the theoretical value of 6 dB (Table 4.2). Further inspection of the data showed that the noise has been slightly reduced when bringing the microphone from 50 cm or 25 cm, in addition to the expected increase in speech level. Because the noise is not perfectly uniform in the room, it is hypothesized that a screening effect by the head for noise components coming from the back may occur when the microphone is closer to the head. On the other hand, the size of the SNR improvement when changing the omnidirectional to the directional microphone is found to be less that the theoretical value of 4.8 dB. As shown in Table 4.2, the SNR improvement when using the directional microphone in IFFM, SSnoise and Helicopter noises is in the range of 3.5−3.7 dB at 50 cm (Dir. 50 – Omni. 50) and 2.1−2.6 dB at 25 cm (Dir. 25 – Omni. 25). Further inspection of the data showed that the speech level was slightly affected by switching from the omnidirectional to the directional microphone

(0.7 dB at 50 cm and 0.1 dB at 25 cm). Factoring out the reduction in speech level, the size of noise reduction arising from switching to the directional microphone was 4.4 dB at 50 cm (closer to the theoretical value of 4.8 dB) and 2.2−2.7 dB at 25 cm (about half the theoretical value of 4.8 dB) in these three noises. Figure 4.9 shows the spectrum of four noises recorded by the omnidirectional and the directional cardioid microphones at the distances of 50 cm and 25 cm. The size of noise reduction arising from directivity benefit at50 cm is very close to the theoretical value for frequencies up to about 4 kHz (Figure 4.9). The smaller size of the directional noise reduction at 25 cm compared to the theoretical value is hypothesized to be caused by the shorter distance of microphone to the mouth and the change in the acoustical environment arising from the head, pinna, and torso reflections (Figure 4.9) [163].

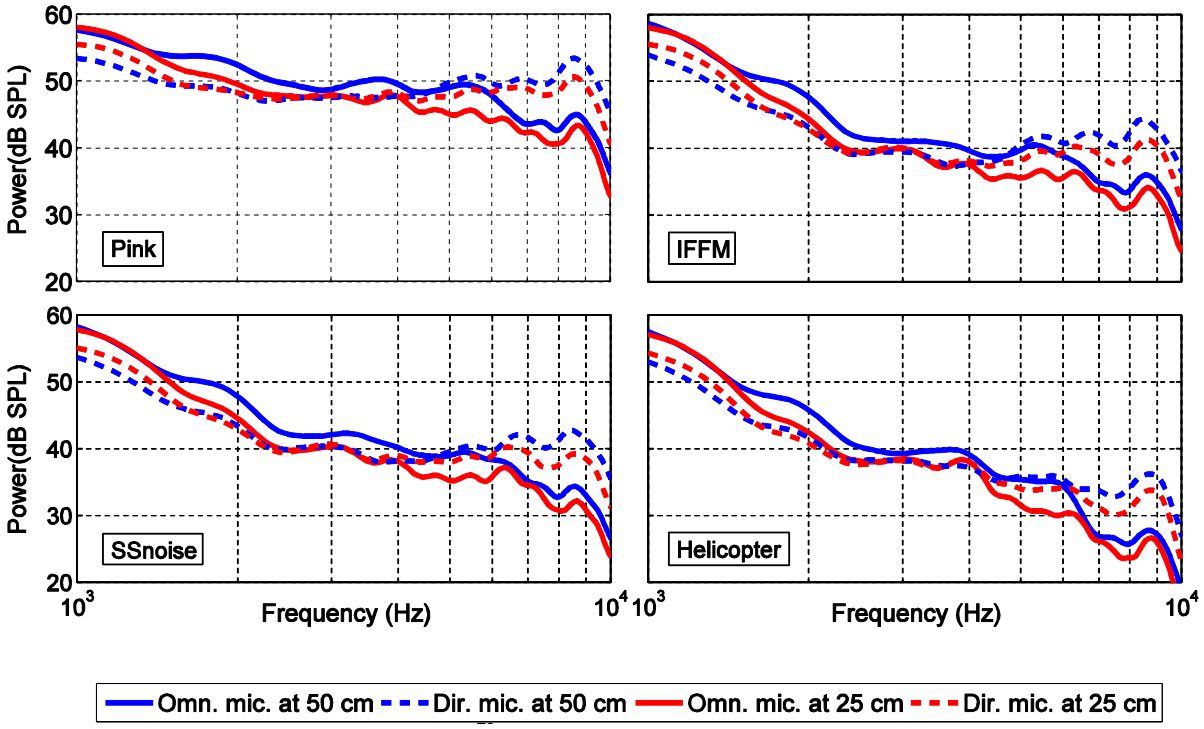


Figure 4.9: *The spectrum of four noises recorded by the omnidirectional (omni.) and the directional cardioid (Dir.) microphones at two distances of 50 cm and 25 cm.*

Table 4.2 shows that there is little directivity benefit when using the directional microphone in the pink noise condition. It was observed that the size of noise reduction arising from switching to the directional microphone in pink noise was only 0.7 dB at 50 cm and 0.1 dB at 25 cm, but the reduction in the recorded speech negated an improvement in SNR. The reason for the lower noise reduction arising from the directivity benefit is that the directional microphone was less good or even negative at high frequencies where the pink energy is dominant (Figure 4.9), so it was not able to reduce the background pink noise as much as for the other three noises which contained more energy in the lower frequencies where the directional benefit is superior.

In summary, the SNR improvement for the directional microphone at 25 cm (Dir. 25) which combines both distance effect and the directivity benefit is not simply the sum of the distance effect and the directional effect at 50 cm because the directional effect interacts with distance. As per Table 4.2, it is observed that the total measured effect is 9.0−9.3 dB in three of the four noises, a bit short of the theoretical max of 10.8 dB, while it is only 6.8 dB in pink noise due the loss of directional effect at the high frequencies.

Table 4.3 shows the total SNR improvement when the four acquisition configurations are combined with the two speech enhancement methods in the four noises in the realistic T2 condition. The reference condition is an acoustic input SNR of 0 dB at the omnidirectional microphone at 50 cm and no further processing for speech enhancement. Note that in the case of the DWS method the size of the reported improvements in Table 4.3 is independent of the acoustic input SNR (Figure 4.6). In the case of the ANC method, the numbers in Table 4.3 strictly apply to the selected acoustic input SNR of 0 dB since the performance of this adaptive speech enhancement method depends on the input SNR (Figure 4.6); larger

improvements are expected at less favorable input SNRs and the converse is expected at more favorable input SNRs.

Despite the larger size of acquisition benefit for the directional microphone compared to the omnidirectional microphone, as shown in Table 4.2 at both distances, it is observed in Table 4.3 that the total SNRs improvements with the directional microphone (Dir 50 and Dir25) are smaller than those with the omnidirectional microphone (Omni.50 and Omni.25) for both speech enhancement methods in all noises except the helicopter noise. As such, this indicates that the omnidirectional microphone provided a superior recording configuration for the purpose of speech enhancement, despite the higher acoustic input noise. It may be that the directional microphone is more prone to distortion [146, 147], which could affect the performance of DWC and ANC methods to subsequently recover the speech signal.

TABLE 4.3: *The total improvement in SNR (dB) resulting from the acquisition configurations combined with the two noise suppression methods of DWS and ANC in the realistic T2 condition. The size of total improvement is calculated with respect to the acoustic input at the omnidirectional microphone at 50 cm at a nominal SNR of 0 dB, without speech enhancement.*

Noise suppression method	Noise Type	Mic. type and distance to the mouth			
		Omni. (50cm)	Omni. (25cm)	Dir. (50cm)	Dir. (25cm)
DWS	Pink	18.2	23.4	14.3	17.6
	IFFM	25.4	29.6	22.7	27.2
	SSnoise	25.9	30.0	24.2	27.4
	Helicopter	29.4	32.0	31.5	33.0
ANC	Pink	16.8	21.1	14.1	17.2
	IFFM	20.2	24.4	18.0	23.5
	SSnoise	19.4	23.0	19.2	22.9
	Helicopter	20.7	22.9	22.3	24.3

Omni.: omnidirectional mic., Dir.: directional cardioid mic.

Overall, it is observed that the largest size of improvement for the selected acoustic input SNR was obtained with the omnidirectional microphone at 25 cm (Table 4.3). The only exception is found with the helicopter noise (both with ANC and DWS) where the total SNR improvement is maximal with the directional microphone at 25cm (Dir 25). Still, the total SNR improvements in the helicopter noise is quite large in both two microphone conditions (Omni.25 and Dir.25) and the advantage of the directional microphone is only 1−2 dB in that noise. Most importantly, the omnidirectional microphone at 25 cm is clearly superior in the most difficult noise type (pink) and in general provides the best results across different noises with less variation with noise type than is obtained with the directional microphone. The results in Tables 4.2 and 4.3 are similar if the SNR improvement is measured in A-weighted dB.

4.3.3 Evaluation of the performance of speech enhancement methods in recovering clean speech

The aim of this section is to evaluate the extent by which the enhanced speech signals recorded by the omnidirectional microphone at 25 cm estimate the clean speech in the realistic T2 condition. For this purpose, speech parameters (energy and pitch) and quality/intelligibility measures (ESII, short-time speech-based STI, and PESQ) were calculated and compared for the clean, noisy and enhanced speech signals in the four different noises at SNRs ranging from −20 dB to +20 dB.

As shown in Figure 4.10, both speech enhancement methods successfully estimate the true energy level of the speech signal down to an SNR of −10 dB in all noises. In contrast,

when there is no speech enhancement, the energy of the noisy speech overestimates that of the original clean speech at all SNRs, except when in very favorable conditions (SNR of +20 dB). This implies that there is an improvement of 20 to 30 dB in the range of SNRs where speech energy can be successfully recovered in noise when using the two speech enhancement methods. The lesser success with pink noise at SNR of −20 dB is related to the smaller size of noise reduction with the speech enhancement methods compared to the other noises, as previously observed (Table 4.3, Figure 4.8).

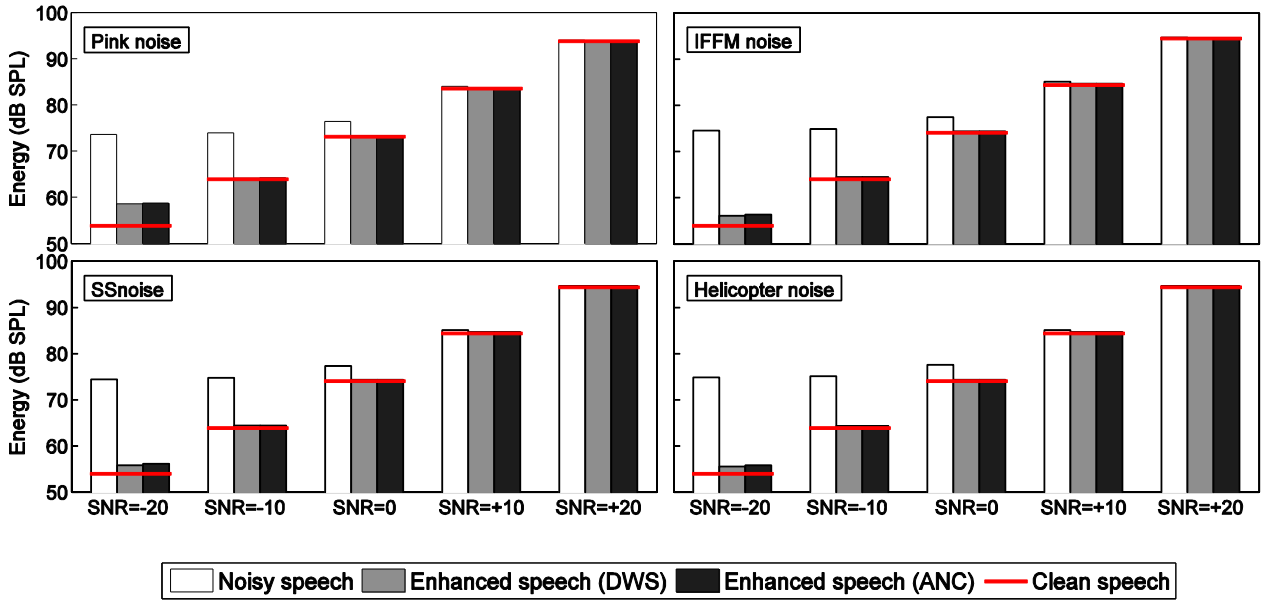


Figure 4.10: *The energy of clean, noisy, and the enhanced speech signals obtained by DWS and ANC methods in four different noise types (SNRs= −20 dB to +20 dB). All signals were recorded by the omnidirectional microphone at 25cm.*

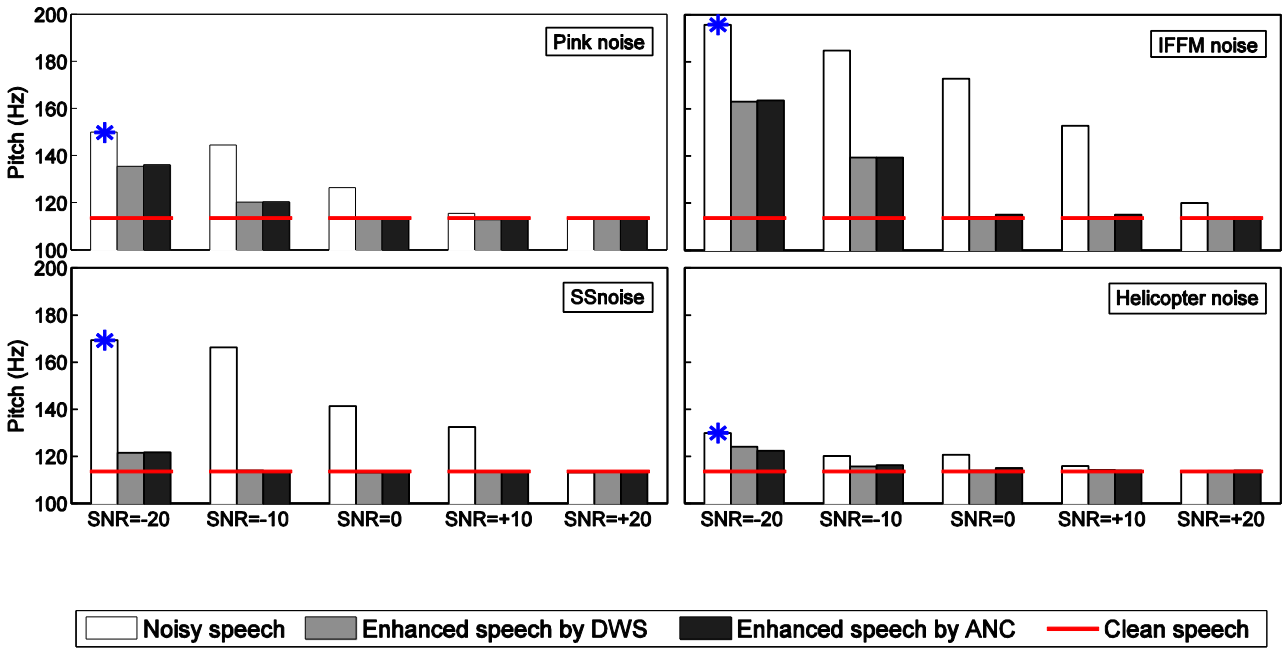


Figure 4.11: *The pitch of clean and the enhanced speech signals obtained by DWS and ANC methods in four different noise types (SNRs= −20 dB to +20 dB). All signals were recorded by the omnidirectional microphones at 25cm. The star symbol means the pitch could not be estimated reliably at SNR=−20 dB.*

Figure 4.11 shows the values of pitch extracted from the noisy, clean, and the enhanced speech signals. The stars on top of the bars indicate that the pitch could not be estimated reliably by PRAAT in the noisy speech condition at the lowest SNR. Similarly to the estimation of energy level values in Figure 4.10, the pitch estimation from the enhanced speech signal is close to that for the original clean speech at SNRs from 0 dB to +20 dB for pink, IFFM, SSnoise and helicopter noise and even at the SNR of −10 dB for the two latter noises. The overestimated pitch for the enhanced speech at SNRs of −10 dB and −20 dB in the IFFM noise condition is likely because of the competing female voiced speech embedded in this noise which interferes with the estimation of pitch from the male HINT speech signal. The situation is however much worse when there is no speech enhancement. Overall, the two

noise cancellation methods allowed recovering the true pitch within 0.2 Hz to 6.7 Hz at SNRs from −10 dB to +20 dB in three of the noise types: pink, SSnoise, and helicopter. It is within 0.2 Hz to 1.4 Hz of the true pitch at SNRs from 0 dB to +20 dB for the IFFM noise.

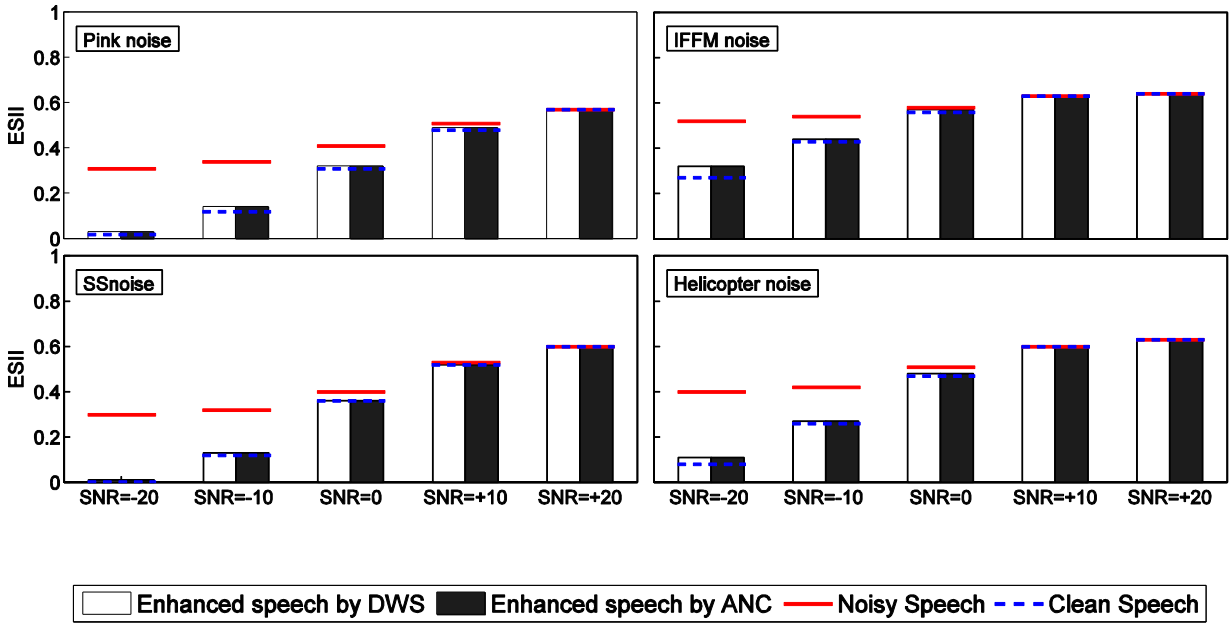


Figure 4.12: *The ESII estimated for clean, noisy, and the enhanced speech signals obtained by DWS and ANC methods in four different noise types (SNRs= −20 dB to +20 dB). All signals were recorded by the omnidirectional microphones at 25cm.*

Figure 4.12 shows the ESII index calculated for the noisy, clean, and the enhanced speech signals in the same four noises. The purpose is to determine the benefit of speech enhancement in estimating the ESII in a noisy environment. The ESII calculated for the noisy speech (red line in Figure 4.12) represents the situation when the noise reduction methods are not used and the only speech signal available is the noisy speech recording. The ESII for clean speech (blue dashed lines in Figure 4.12) is the true value of the ESII and represents the ideal situation when speech is perfectly estimated. The red lines show that

without noise reduction the calculated ESII overestimates the true value corresponding to the clean speech by a wide margin for SNR ≤ 0 dB. As shown, the ESII calculated from the enhanced speech signals is nearly identical to the true value in all noises and at all SNRs down to at least −10 dB. Thus, the noise suppression methods could remove the background noise from the recorded noisy speech signals in a way that their ESII scores approached the true ESII calculated from the clean speech. Similar to the trends observed in the figures related to the pitch and energy, the difference between the ESII index of clean speech (blue dashed-dotted lines) and those for the enhanced speech (bars) decreases for the higher SNRs. Finally, also note that the higher ESII values at the low SNRs in IFFM and helicopter noises compared to the other two noises is consistent with the fact that speech is more intelligible in the fluctuating noises than that in the steady-state noises.

The other two objective measures of short-time speech-based STI and of PESQ also demonstrate the extent by which the speech enhancement methods recover the clean speech in terms of intelligibility and quality scores.

Figure 4.13 shows that the DWC and ANC enhancement methods yielded a significant increase in the STI in all noises and SNRs by up to about 0.4. According to Figure 4.14, the PESQ scores of speech enhanced by DWS and ANC methods again increased in all noises and SNRs, by up to about 1.5, Slightly better improvements in the STI and PESQ are obtained with the DWS method. At the higher SNRs, the STI and PESQ scores become close to their maximum values of 1.0 and 4.5 respectively. The higher short-time speech-based STI index for the speech signals enhanced by DWS method compared to those for ANC method indicates that the envelope of speech signals can be better preserved by using the former method.

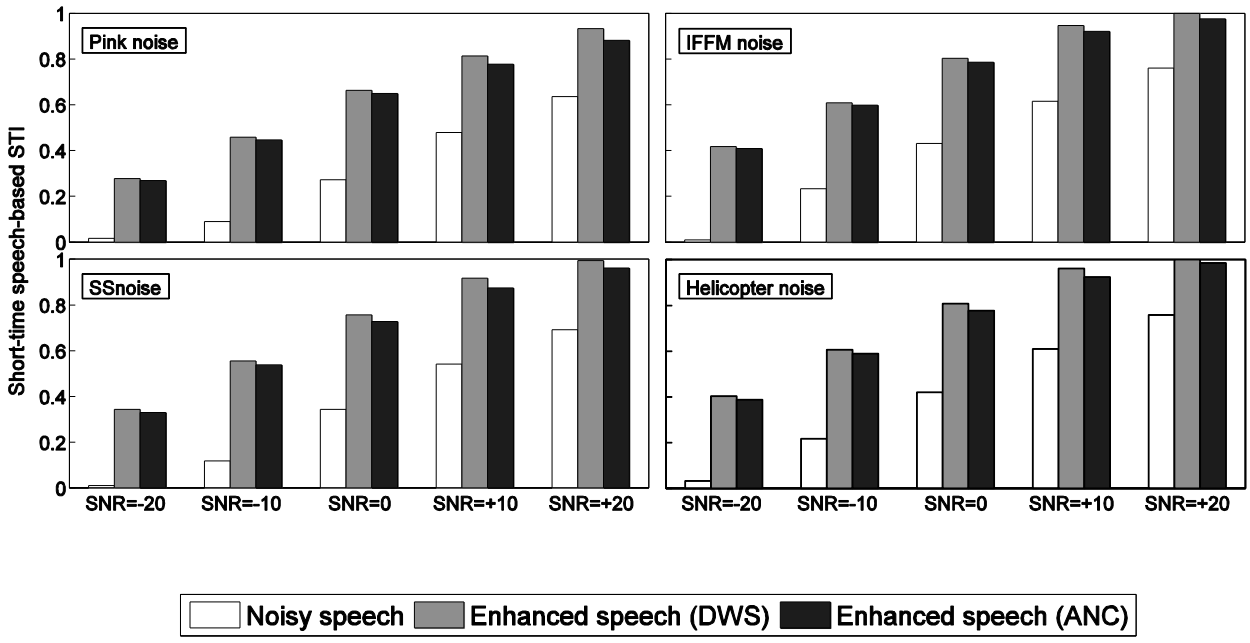


Figure 4.13: *Short-time speech-based STI estimated for noisy, and the enhanced speech signals obtained by DWS and ANC methods in four different noise types (SNRs= −20 dB to +20 dB). All signals were recorded by the omnidirectional microphones at 25cm.*

Overall, the results show that the noisy speech signals recorded by the omnidirectional microphone at 25cm and enhanced by the two noise suppression methods of DWS and ANC yield a far better estimate of the energy and pitch of speech while providing sizeable improvements in the recovery of the envelope and fine structure of speech, as measured by the STI and PESQ respectively, than when no enhancement is used.

4.4 Discussion

The available studies investigating the Lombard effect and its influences on different parameters of speech have been limited by the methodology used to generate noisy environments. When the noise is delivered through headphones to the talkers to elicit the

Lombard effect, they suffer from the occlusion effects of the headphones. Under these circumstances, the noisy condition simulated in the talkers' ears differs from a real-world noise field because the occlusion effect changes the talkers' auditory feedback of their own voice which is used to adjust the level of speech. Moreover, it is not practical to investigate the interaction of hearing protectors or hearing aids with the speech produced in noise when the noise is generated by the headphones. Thus, it is essential to elicit the Lombard effect by an external noise field, which can be simulated by loudspeakers, in order to be able to realistically account for the effects of different factors on speech production in noise. However, investigating the Lombard effect in an external noise field is challenging due to the contamination of the speech by the background noise, which makes it difficult to recover and analyze a clean version of the produced speech signal. Due to the limited research related to the recovery of clean Lombard speech from the noisy recording, this study aimed at evaluating the performance of two noise suppression methods used to extract the clean speech in the simulated external noise field. In order to investigate the effects of different recording methods, noise conditions, and the talker's movement on the performance of the clean speech extraction, two microphones (omnidirectional and cardioid directional) at two distances of 50 cm and 25 cm were employed to record the simulated Lombard speech in four different fluctuating and continuous noise conditions.

It was observed that both noise suppression methods could significantly decrease the target background noise when there is no movement in the talker's body. When the talker moved during the recording of noises, as would be the typical situation in practice, the size of noise reduction was less compared to the ideal condition of no movement, because the of recorded target noise becomes less matched to the reference noise (Figure 4.5).

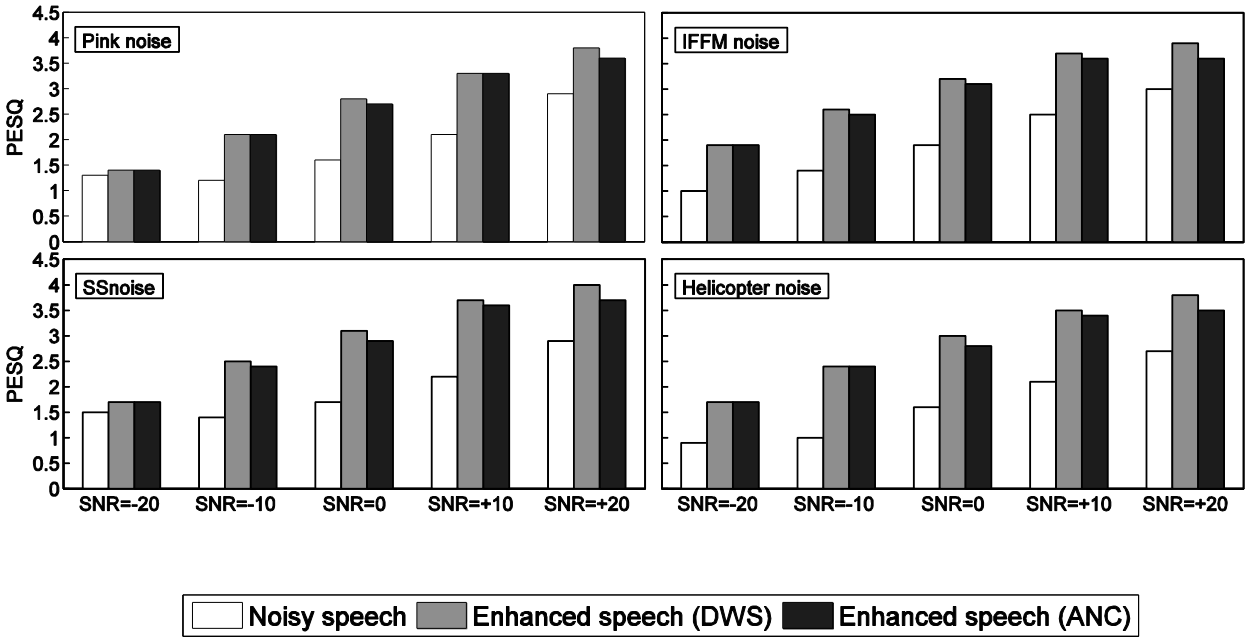


Figure 4.14: *The PESQ estimated for noisy, and the enhanced speech signals obtained by DWS and ANC methods in four different noise types (SNRs= −20 dB to +20 dB). All signals were recorded by the omnidirectional microphones at 25cm.*

When no speech is present or at very low SNRs, the ANC method cancelled more noise than that cancelled by DWS method (Figure 4.6). ANC could even suppress the noise to an amount lower than the floor noise in the room at low frequencies (Figure 4.5). This is because the ANC method is able to filter the reference noise with respect to the target noise to be cancelled until their subtraction reaches the optimal minimum value. The ANC method is also less sensitive to the alignment issues arising from the delay in the recording of the target and the reference noises (Figure 4.7). It was observed, however, that the performance of the ANC method gradually deteriorates at higher SNRs due to the greater prominence of speech hindering the ability of the ANC filter to adaptively match the reference noise to the target noise. Nonetheless, the speech spectra recovered with the ANC method are as good as those with the DWS method at representative low and high SNRs (Figure 4.6).

The amount of noise reduction with both methods was affected by the type of noise. Greater noise reduction was achieved in the fluctuating and low-frequency helicopter noise, while the least reduction was obtained in the continuous pink noise which has more energy at higher frequencies. We speculate that noise reduction for the pink noise is lower because the effects of movement on the sound field may be more important at high frequencies, resulting in greater perturbation of the target pink noise with respect to the reference noise. As a consequence, the recovery of the speech spectrum was better with the helicopter noise than the pink noise at lower SNRs (Figure 4.8).

Different acquisition configurations were investigated for recording the speech and noise signals. Halving the microphone distance from 50 to 25cm increased the speech level and associated SNR by about 6 dB, as expected. Use of a cardioid directional microphone picked up less noise by about 2−4 dB below 2 kHz, but did not consistently lead to better noise reduction, especially in pink noise. Overall, the omnidirectional microphone at 25 cm was the preferred acquisition configuration as it was found to provide good noise reduction for all noise types.

The pitch and energy values extracted from the enhanced speech at different SNRs showed results very close or identical to clean speech, except at the lowest SNR of −20 dB or pitch extraction in a voice masker like the IFFM noise. Moreover, the objective speech measures calculated for the enhanced speech demonstrated that the noise suppression methods could substantially increase the predicted intelligibility and quality scores starting at SNRs as low as −10 dB.

In practice, both DWS and ANC noise suppression methods provide adequate noise reduction for SNRs above −10 dB when using the omnidirectional microphone at 25 cm. Brungart et. al. (2012) showed that the SNR range of speech produced by open-ear talkers in

pink noise levels of 50 dBA to 80 dBA was 1 dB to 13 dB at the distance of 50 cm from the mouth [117]. When wearing a set of combat earplugs in linear mode, the SNR ranged from −8 dB at 80 dBA to 9 dB at 50 dBA while in the nonlinear mode, the SNR ranged from −4 dB to 12 dB. Hormann et. al. (1984) indicated that when talker and listener were open ears in the pink noise levels of 76 dBA to 92 dBA, the SNR ranged from −11 dB to −4 dB at a distance of 1 m [15]. At the condition where both talker and listener wore earplugs, the SNR ranged from −16 dB to −8 dB at the same noise levels and distance. At the preferred acquisition configuration selected in our study, the SNR values reported above would increase by 6 to 12 dB due to the shorter recording distance of 25 cm compared to the 50 cm or 1 m recording distances used in those investigations. Consequently, the methods evaluated in this study are suitable to successfully recover Lombard speech produced in an external noise field with or without hearing protector.

The methods evaluated in this study are applicable in laboratory settings and for investigations of the Lombard effect where the same noise waveform can be presented twice, once to generate a reference noise and once to collect the noisy Lombard speech. However, a limitation of the present study is that the noise and speech signals were recorded separately, then mixed to create noisy speech signals at different SNRs. Speech was produced through the artificial mouth of the manikin located at the reference position in the room while the noise signals were presented by loudspeakers in the presence of a real participant with/without dynamic body movements. This was done in order to have an identical "clean" speech signal in all conditions with different acquisition set-ups and so allow direct comparison of the recovered speech following noise suppression. Therefore, the speech material could not be recorded while it was read by a real participant in the noisy room which would have resulted in no single comparator. We expect that this recording

methodology would not have any significant effect on the results, since the purpose is only focused on the size of noise reduced by two methods, which is not affected by the way speech is recorded.

In the next chapters, the Lombard speech produced by real talkers wearing hearing protectors will be recovered using these two noise cancellation methods in order to evaluate the effects of different noise conditions and hearing status on different speech parameters.

4.5 Conclusions

Use of two noise suppression methods, direct waveform subtraction and active noise cancellation, was found to adequately remove noise from speech recorded in the context of Lombard effect studies carried out in an external noise field, for example when hearing protectors or other hearing devices are worn. The method requires presenting the noise waveform twice. It is important to minimize movements in the talker's body during recordings in order to increase the amount of noise reduction. The type of microphone and its distance to the talker's mouth also affect the amount of noise removal and speech pick up. The noise reduction achieved by the DWS method is independent of the initial SNR, but noise reduction by the ANC method decreases when increasing the SNR. On the other hand, the DWS method is sensitive to time alignment of the two noise recordings, whereas the ANC method is more robust. Speech feature extraction and objective intelligibility and quality measures calculated for the enhanced Lombard speech verified the effectiveness of the two noise suppression techniques in recovering speech at SNRs as low as −10 dB.

Chapter 5

Changes in Speech Production Induced by a Conventional Hearing Protector Worn by the Talker and/or the Target Listener in Different Noise Conditions

5.1 Introduction

Exposure to high level of noise is one of the main health-related concerns in many industrial and military workplaces. Excessive occupational noise can cause temporary or permanent hearing loss [6, 8, 48]. Moreover, loud noise can jeopardize workers' safety through physical and psychological stress, annoyance, fatigue, and interference with verbal communication and warning sound perception [4, 5, 6]. Military services and industries have been advised to

use protective devices (e.g. earplugs or earmuffs) in order to prevent noise-induced hearing damages [27, 40, 113] when the noise cannot be further reduced through engineering and administrative control methods. Although hearing protectors attenuate the level of noise to a lower more comfortable level, they can impede speech communication and users' awareness of the surrounding acoustic environment [31, 32, 37].

Hearing protectors affect the users' perception of speech and other desirable surrounding sounds by reducing equally the level of unwanted sounds (noise) and wanted sounds such as speech and warning signals [22, 24, 31, 32, 37, 49, 119]. Many investigations have been carried out to evaluate the effect of wearing conventional hearing protectors on speech and auditory perception [15, 17, 18, 19, 25, 27, 28, 29, 32, 33, 34, 36, 37, 123, 164]. The effects of the fixed attenuation provided by conventional protectors can cause communication problems, particularly in the case of workers with pre-existing hearing loss [22, 31]. For these individuals, blocking the ear canal by the hearing protector increases the challenges since they already suffer the problems of decreased audibility.

The type and characteristics of the background noise affect the size of masking and consequently the size of changes in auditory perception and speech production [96, 97, 98, 66]. The masking effect of noises varies depending on the temporal and spectral characteristics, for instance, amplitude modulations or fluctuations, the energy distribution over frequency spectrum, and the existence of any linguistic information [95]. Normal hearing individuals have a better understanding of speech spoken in fluctuating maskers compared to that spoken in continuous noises [97-101]. The phenomenon of "dip listening" or "release from masking" is referred to the listener's ability to extract speech cues audible in relatively silent and short periods of noise where the local signal-to-noise ratio is high because of momentary decrease of fluctuating noise level [97, 99, 102, 103, 110, 165, 166,

167]. Blocking the ear canal by the hearing protector can disrupt the ability of listeners to take advantages of brief "acoustic glimpses" of speech in dips of fluctuating noise, especially in listeners with hearing loss. The effect of protectors on speech communication and auditory perception also depends on the level of noise in which they are worn. It was observed that wearing protectors causes little degradation on the ability of normal-hearing listeners to understand speech in noise levels above 85-90 dBA [25, 27, 40]. Furthermore, it was found that wearing hearing protectors could even provide a benefit in speech intelligibility at high noise levels [16, 25, 28, 168, 169] for individuals with normal hearing. At low level of noise (<70 dBA), the protectors cause a reduction in speech intelligibility [25] probably because of the attenuation of the lower level of speech to a level insufficient for recognition.

Hearing protectors can not only affect speech perception but also speech production [22, 25, 113, 114]. When the ear canal is occluded by a hearing protector, the perception of one's own voice increases relative to the open-ear condition. Indeed, the occlusion effect blocks the air conduction of sounds through the ear canal while resonates the bone-conducted vibrations of talker's own voice, resulting in a "boomy" and "hollow" auditory feedback amplified in low-frequency [117, 170]. Thus, hearing protectors reduce the talkers' natural tendency to speak loud enough against the background noise, known as the Lombard effect, by increasing the users' perception of their own voice and decreasing their perception of the surrounding noise [39].

There have been relatively few studies investigating the effect of wearing conventional hearing protectors on speech production [15, 25, 27, 40, 41, 117, 171]. In quiet, wearing the conventional hearing protectors has been found to promote higher speech level by 4−6 dB compared to open-ear condition [25, 41, 171] although some authors reported a small

decrease of 1 dB [39, 40]. In noise, studies showed that using conventional hearing protectors decrease the talkers' speech level by typically 3−6 dB compared to when unprotected [15, 25, 27, 40, 117, 171]. Some authors reported a slightly greater size of decrease in the talkers' speech level when using the hearing protectors in some situations [40, 117]. Brungart et al. (2012) showed that talkers wearing a combat earplug spoke less intensely by 10.1 dB when they wore the earplug in the linearly attenuating mode and by 5.8 dB when they used the nonlinear mode designed to protect against high-impulse noises only, compared to no protection [117]. Tufts and Frank (2003) observed a drop in overall speech levels and corresponding SNRs by 4 to 11 dB in different noise levels for the talkers wearing the earplugs compared with the open-ear condition [40]. They also observed that although the spectral center of gravity (SCoG) of speech increased in frequency in the presence of background noise compared to quiet condition, the SCoG was consistently lower when the talkers wore the earplugs. In a noise level of 100 dB SPL, the mean SCoG were 972 Hz for the talkers in the open ear condition while it was 624 and 756 Hz for the talkers with foam and flange earplugs, respectively. Furthermore, the changes in the speech intelligibility index (SII) when talkers wore earplugs compared to the unprotected condition showed that the speech information available for listeners decreased in the same noise in the protected condition [15, 27, 40, 117].

The effect of wearing the hearing protectors on the speech of talkers can depend on the level and type of noise affecting the masking size of noise. Most available studies evaluating the acoustical characteristics of speech produced in noise, with or without hearing protection, have been mostly used continuous noises [15, 28, 29, 32, 33, 34, 37, 64, 77, 78, 83, 91, 105, 106, 107, 108, 172], while many natural noises interfering with daily-life speech communication are often temporally fluctuating [97, 109, 110, 111, 112]. As a result, it is

necessary to evaluate the speech production for the talkers wearing the hearing protector under a more realistic noisy condition in order to account for the interaction effect between real workplace noises and the hearing protectors on speech communication. There have been relatively few studies on speech production when using hearing protectors, particularly in fluctuating noises and across a range of noise levels. Also, with exceptions, most studies focused exclusively on speech levels, and not in the other features of speech.

The effect of wearing hearing protectors on speech perception and production also depends on the role of wearer being a listener, a talker or both. It was observed that speech intelligibility was reduced when both the talkers and listeners wear protectors [27]. Hormann et al. (1984) investigated the effect of wearing the earplugs in pink noise by a talker and a listener on verbal communication and the extent of speech intelligibility in all four possible combinations of hearing protector usage: (1) both unprotected, (2) only talker protected, (3) only listener protected, and (4) both protected [15]. This allowed evaluating the effect of the listener's ear condition as well as the talker's ear condition and their interactions on speech communication with HPDs. The presence of a real participant as the listener prompted the talker to adjust the level of Lombard speech more accurately. Overall, the authors observed a 40 percent reduction in the understanding of speech and a considerable delay in the exchange of information when both talker and listener wore hearing protectors. The highest speech intelligibility was related to the situation where both talkers and listeners were with open ears, while the lowest score were obtained when the protector was worn by both of them [15].

The aim of this chapter is to evaluate the effect of an earmuff-type hearing protector on the level, pitch and the spectral distribution of speech produced in quiet and in different types of continuous and fluctuating noises presented at different levels in an external sound field.

Following Hormann et al. [15], different ear conditions are investigated for the talker and listener in terms of hearing protection use.

5.2 Methods and materials

5.2.1 Participants

Twenty-four participants were recruited to participate in the study. They were equal numbers of males (12) and females (12), ranging in age from 19 to 32 years, native speaker in the English language, without any history of smoking, speech/language impediments, and respiratory, postural or other chronic conditions that could affect speech output. All participants had normal hearing thresholds ≤ 15 dB HL for frequencies up to 4000 Hz in each ear and ≤ 25 dB HL at 6000 Hz and 8000 Hz, and normal tympanograms. They filled out a questionnaire about their auditory history and the type and duration of the noise exposure they experienced, if any, in their daily activities. Note that the data for two other participants were collected but discarded and replaced because it was found that English was not the first language for one participant. For another participant, the hearing thresholds were higher than 15 dB HL at three low frequencies of 250, 500, and 1000 Hz.

5.2.2 Noises, speech material, and hearing protector

Four different noise types were chosen in this study to explore a range of temporal and spectral variations in the characteristics of the background noise. They are pink noise,

International Female Fluctuating Masker (IFFM) [135], speech-spectrum noise (SSnoise) from the International Female Noise (IFnoise) [135], and helicopter noise (Figure 4.2b). Pink noise and SSnoise are continuous noises while the other two represent the fluctuating noises (Figure 4.2a).

The noise stimulus was a concatenated sequence consisting of pink, IFFM, IFnoise (SSnoise) and helicopter noise segments, in this order, each 27 second long separated by a 5 second pause. The noise stimulus sequence was presented at two different levels of 70 dBA and 85dBA at the talker's head position in the noise simulation room.

A list of 12 sentences selected from the American English Hearing-In-Noise Test (HINT) [148] was used as speech material to be read by the talkers. The HINT sentences chosen were representative of conversation in noisy workplaces (e.g. *The engine is running*) (Table 5.1).

The hearing protector used in the experiment was the earmuff PELTOR Optime^{TM} 98 (3M, Minneapolis MN) with a Noise Reduction Rating (NRR) value of 25 dB calculated from the laboratory-based attenuation data as reported by the manufacturer.

5.2.3 Procedure

The experiment was conducted in a noise simulation room with the dimension of 4.29 m × 3.65 m × 2.42 m (length × width × height) in the Hearing Research Laboratory at the University of Ottawa (Figure 4.3). The equalized and amplified noise signals were sent to the loudspeakers located around the simulation room by an external sound card (TASCAM US-366, TEAC Co., Japan) at the sampling rate of 96 kHz. There were six loudspeakers

including four tower type (Sony MF600H), a bookshelf speaker on a stand (Yamaha NS-C155) and a subwoofer (Mirage Sub12) to generate a quasi-diffuse sound field in the room [16]. The spectral equalization of noise signals was used to compensate for the variation arising from the room response. The same soundcard was also used to record the signals in the room at the same sampling rate.

The microphone used to pick up the speech was a free-field omnidirectional microphone (B&K Type 4189) located 25 cm away from the talker's mouth and connected to a sound level meter (B&K Type 1625). This acquisition configuration was found to be the best overall to extract speech in noise among four configurations previously studied (Chapter 4).

TABLE 5.1: *The HINT sentences chosen as speech material for the talkers*

No.	HINT sentences
1	The engine is running.
2	He climbed up the ladder.
3	A janitor swept the floor.
4	The buckets fill up quickly.
5	It's getting cold here.
6	There are branches everywhere.
7	Someone is crossing the road.
8	A fire truck is coming.
9	The paint dripped on the ground.
10	A bag fell off the shelf.
11	The exit is well lit
12	The scissors are very sharp

The participant was asked to sit on a chair at the centre of the simulation room 1-m away from an ANSI S3.36 acoustic manikin (B&K 4128). Vocal output is known to increase less when the task is just reading text in isolation compared to when the emphasis is on transmitting intelligible message to a target listener [56]. Thus, the presence of the acoustic manikin prompted the talker to adjust the voice level accordingly. The talker was asked to read the list of 12 sentences attached to the manikin's torso and was instructed to speak in such a manner to be understood at the target distance of the manikin in the different noises at different ear conditions. The participants were asked to start speaking two seconds after the beginning of each noise. The first two seconds of each noise was assigned for the talker's preparation to adapt to each noise condition. The talker was monitored during the test by the experimenter in the control room outside the recording booth.

The participant's output speech signal was collected in quiet and in noise at two different levels and under four different ear conditions: (1) when both talker and listener were in open ears ("Open-Open" condition), (2) when the talker was wearing the hearing protector while the listener was in open ears ("HPD-Open" condition), (3) when the talker was in open ears while the listener was wearing the hearing protector ("Open-HPD" condition), and (4) when both talker and listener wore hearing protection ("HPD-HPD" condition). The order of ear conditions (4), types (4) and levels (2) of noise was counterbalanced across the participants.

As discussed in Chapter 4, two enhancement methods consisting of active noise cancellation (ANC) and direct waveform subtraction (DWS) were developed to recover the clean vocal output from the noisy recording in the room. Both methods could adequately remove the noise from the vocal output at SNRs as low as −10 dB. This was verified by extracting some basic characteristics of speech (e.g. pitch and energy) and objective quality and intelligibility measures from the enhanced speech, which found to be very close to those

features extracted from the clean speech at SNRs down to −10 dB (Section 4.3.3). In order to be able to use the enhancements methods the noise signal needs to be send twice in the simulation room, one for recording the reference noise while the talker is asked to be silent and one when the talker is speaking. In this work, the speech signal was recovered by using the ANC method in order to assure the better recovery in the case of low SNRs, since this method was found to be a little more superior at very low SNRs and was less prone to alignment errors. The LMS algorithm, with different step size values depending on the SNR, was used to adaptively update the filter coefficients.

5.3 Results

5.3.1 Overall speech energy level

5.3.1.1 Quiet condition

Figure 5.1 and Table 5.2 show the mean overall speech level (dB SPL) for different ear conditions of the talker and listener (with and without the earmuff) in quiet. The effect of the talker status on speech levels is relatively small. In the situation where the target listener (manikin) was open ears, speech levels increased by 1 dB when talkers wore the HPDs relative to open ears (i.e. 74.5 dB SPL in "HPD-Open" condition vs. 73.5 dB SPL in "Open-Open" condition). In the situation where the target listener was wearing the HPD, a small increase of 0.2 dB SPL was also observed in the speech level when the talkers were wearing

protection relative to open ears (i.e. 80.3 dB SPL in "Open-HPD" condition vs. 80.5 dB SPL in "HPD-HPD" condition).

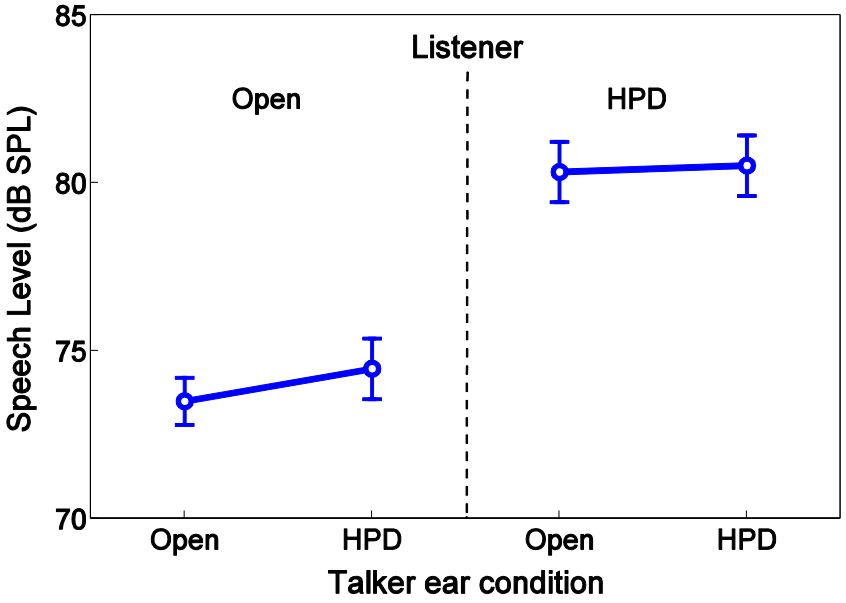


Figure 5.1: *The mean overall speech level (dB SPL) for four different ear conditions of the talker and listener in quiet (Open ears or with a hearing protector). Error bars are the standard error of the mean: ± standard deviation/√sample size).*

TABLE 5.2: *The mean and standard deviation (std.) of the overall speech level (dB SPL) separately for males and females and when averaged over both genders for different ear conditions of the talker and listener in quiet*

Gender	Speech level	Talker-Listener ear condition			
		Open-Open	HPD-Open	Open-HPD	HPD-HPD
Females (n=12)	mean (std.)	72.8 (2.7)	73.6 (3.4)	81.6 (4.4)	80.3 (4.6)
Males (n=12)	mean (std.)	74.2 (3.6)	75.3 (5.4)	79.1 (3.7)	80.7 (4.3)
Average (n=24)	mean (std)	73.5 (3.2)	74.5 (4.5)	80.3 (4.2)	80.5 (4.3)

The effect of the listener status on talkers' speech levels was more important. When the talkers were open ears, speech levels increased by 6.8 dB when the listener wore the HPD relative to open ears ("Open-Open" vs. "Open-HPD" conditions). Likewise, when the talkers wore the HPD, the speech levels increased by 6.0 dB when the listener was also wearing the HPD relative to open ears ("HPD-Open" vs. "HPD-HPD" conditions). It shows that when the listener wore the HPD, the talkers raise their voice no matter the ear conditions they were in.

A two-way mixed ANOVA was performed on the speech level data. The within-subjects and between-subjects factors are ear condition and gender, respectively. Ear condition was a significant main effect ($F(3,66)=59.0,\ p<0.001$) while the effect of gender was not significant ($F(1,22)=0.03,\ p=0.86$). There was a significant interaction between ear condition and talkers' gender ($F(3,66)=3.81,\ p=0.014$). Multiple pairwise t-test comparisons using Bonferroni correction (p=0.05/6=0.008) showed that the "Open-Open" vs "HPD-Open" and the "Open-HPD" vs "HPD-HPD" ear condition comparisons were not significant. All other four pairwise comparisons reached statistical significance. Figure 5.2 shows the interaction between ear condition and gender. When the target listener is with open ears, the males and females both increased their speech level when wearing the HPD compared to open ears. In contrast, when the target listener wears the HPD, the males increased their speech level while the females decreased theirs.

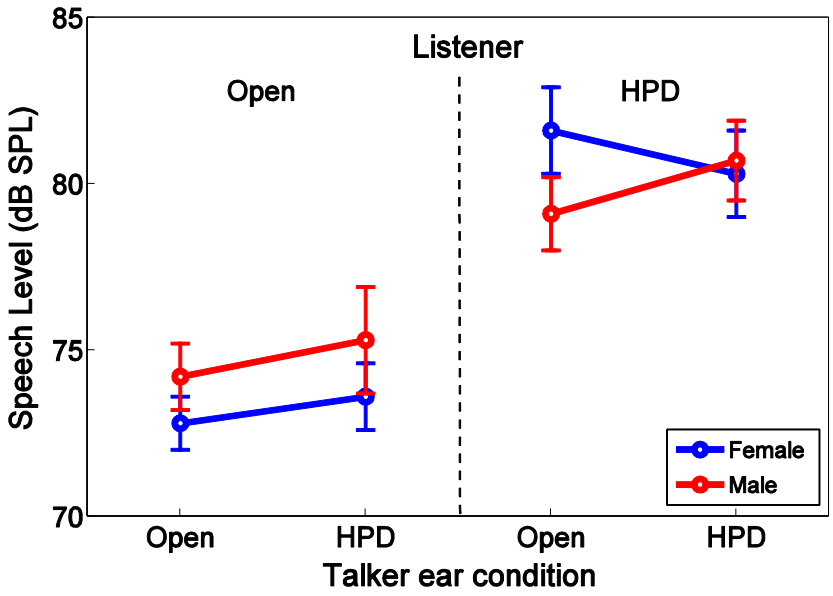


Figure 5.2: *The mean overall speech level (dB SPL) separately for males and females for four different ear conditions of the talker and listener in quiet (Open ears or with a hearing protector)*

5.3.1.2 Noise condition

Figure 5.3 illustrates the mean speech level for four different ear conditions in quiet and in four types of noise presented at 70 dBA and 85 dBA. It is obvious that the talkers raise their voice level in noise relative to that in quiet. At the noise level of 70 dBA, there was an increase in speech levels by 8.4 dB to 11.2 dB ("Open-Open" condition), 4.0 dB to 4.9 dB ("HPD-Open" condition), 5.2 dB to 7.8 dB ("Open-HPD" condition), and 2.2 dB to 3.0 dB ("HPD-HPD" condition) across different types of noises (Table 5.3). At the noise level of 85 dBA, the increase in speech level relative to that in quiet was 14.0 dB to 18.3 dB ("Open-Open" condition), 5.6 dB to 8.7 dB ("HPD-Open" condition), 8.8 dB to 13.1 dB ("Open-HPD" condition), and 3.0 dB to 6.0 dB ("HPD-HPD" condition) across the four noises.

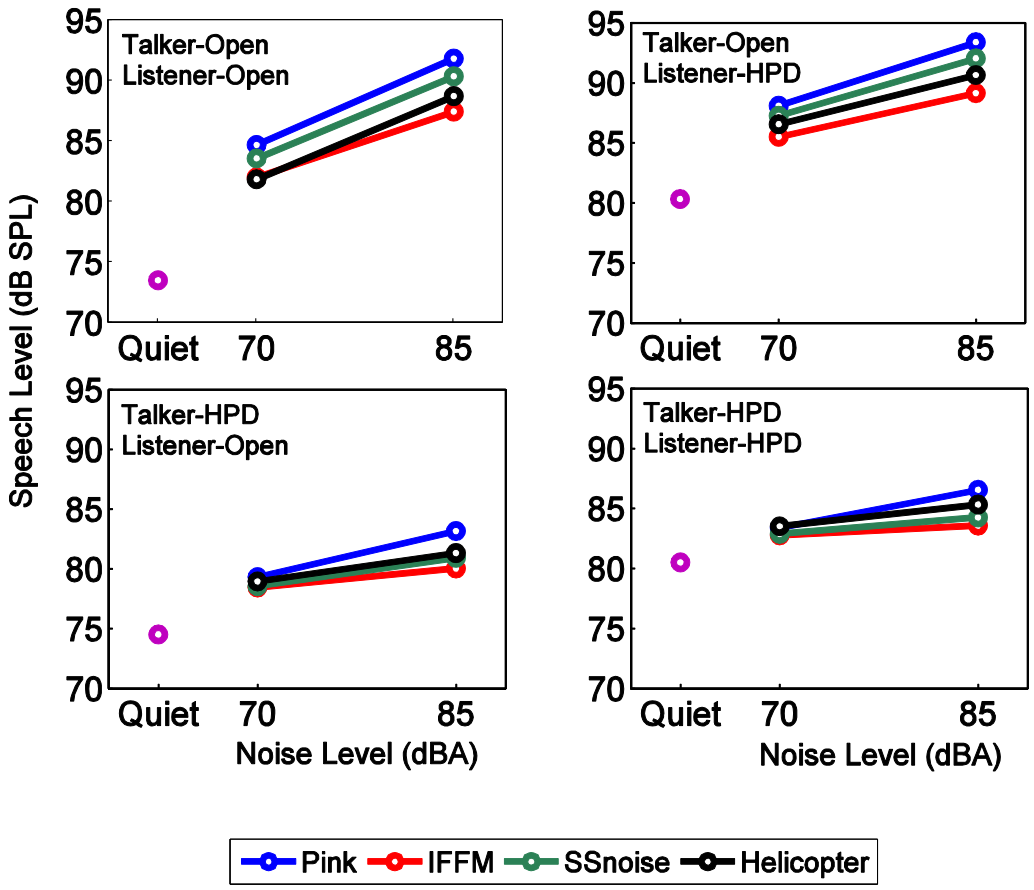


Figure 5.3: *Mean overall speech level (dB SPL) for four different ear conditions of the talker and listener (open ears and with a hearing protector) in quiet and at four different noises at two levels.*

In the higher noise level of 85 dBA compared to the noise at 70 dBA, the talkers raised the level of speech by 3.6 dB to 7.1 dB (talkers in open ears) and by 0.8 dB to 3.8 dB (talkers with protection) across different types of noises and listener's ear conditions.

In 70 dBA of noise, the mean speech level decreased by 2.9 dB to 5.4 dB (listener in open ears) and by 2.8 dB to 4.7 dB (listener with protection) when the talkers wore the HPD relative to open ears (Figure 5.4). The corresponding size of the decrease was 7.4 dB to 9.4 dB (listener in open ears) and 5.3 dB to 7.8 dB (listener with protection) at the noise level of

85 dBA. It is observed that when the talkers wore the HPDs at the lower noise level of 70dBA, the speech levels in the different types of noises were very close to each other (all symbols overlap) for a given listener ear condition. The maximum difference in speech levels across the different types of noises was 0.9 dB when the listener was open ears and 0.7 dB when the listener wore hearing protection. Thus, wearing the HPD at the noise level of 70 dBA did not cause the talkers to react appreciably to the differences in the noise type.

In contrast, the effect of noise type on the speech level is much larger when the talker is wearing the HPD at the higher noise level of 85 dBA (by up to 3.2 dB) or when in open ears at both noise levels (by up to 4.4 dB).

TABLE 5.3: *The overall speech level (dB SPL), averaged over both genders (n=24), for four different ear conditions of the talker-listener in four different noise types at two levels (mean and standard deviation)*

Noise Level:		70 dBA				85 dBA			
Noise Type:		Pink	IFFM	SSnoise	Helicopter	Pink	IFFM	SSnoise	Helicopter
Open-Open	mean	84.7	82.0	83.6	81.8	91.8	87.4	90.4	88.7
	(std.)	(3.3)	(3.2)	(2.9)	(2.7)	(3.5)	(3.6)	(3.3)	(3.2)
HPD-Open	mean	79.3	78.4	78.5	79.0	83.2	80.0	80.9	81.3
	(std.)	(4.9)	(4.7)	(5.0)	(5.1)	(5.0)	(5.1)	(5.0)	(5.2)
Open-HPD	mean	88.1	85.5	87.3	86.5	93.4	89.1	92.0	90.6
	(std.)	(3.8)	(3.7)	(3.7)	(3.9)	(4.0)	(4.0)	(3.5)	(3.5)
HPD-HPD	mean	83.4	82.8	82.9	83.5	86.5	83.6	84.2	85.3
	(std.)	(4.9)	(4.5)	(4.7)	(4.5)	(5.1)	(5.0)	(5.0)	(5.1)

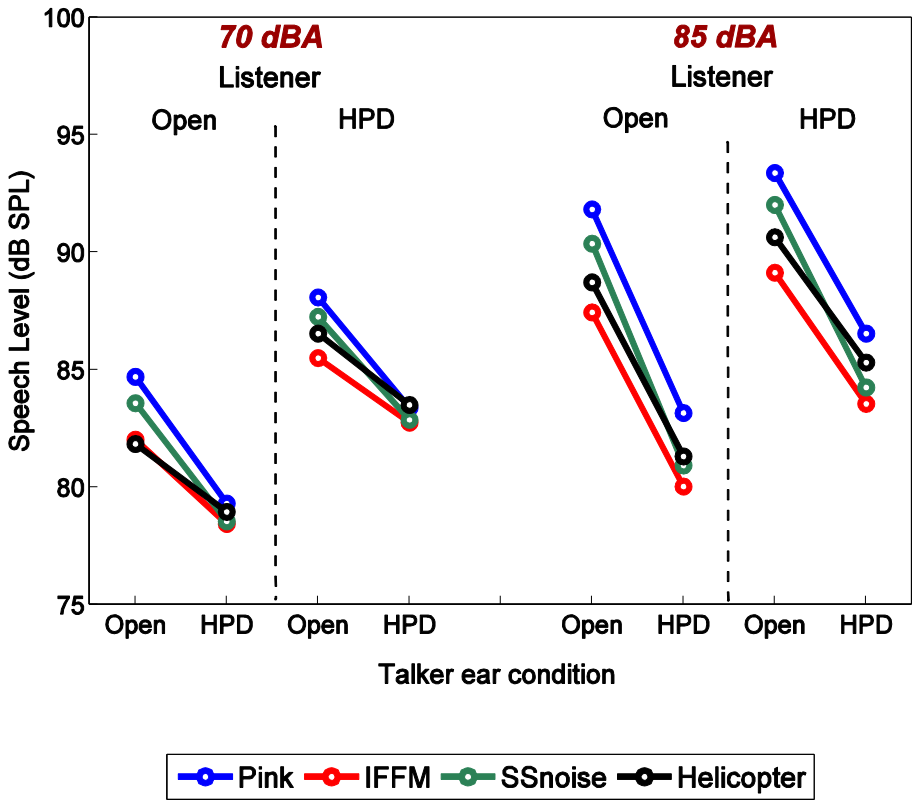


Figure 5.4: *Mean overall speech level (dB SPL) in four noise types at two levels for four different ear conditions of the talker and listener (open ears and with a hearing protector)*

Comparing the curves on the right side of the dashed line to the curves on the left in Figure 5.4, it is observed that the talkers increased their voice level when the target listener (manikin) wore the hearing protector. At the noise level of 70 dBA, the average rise in speech level across the different noises is 3.8 dB (talkers in open ears) and 4.4 dB (talkers with protection). At the noise level of 85 dBA, the increase is 1.7 dB and 3.6 dB, respectively.

The speech level increased by 15.0 dB, 11.4 dB, 13.7 dB, and 12 dB in pink, IFFM, SSnoise, and helicopter noise, respectively, when comparing "Open-Open" condition in quiet to that in noise (average of both noise levels) (Figure 5.3). The larger increase for the

pink and SSnoise could indicate that the talkers perceived these two continuous noises as stronger maskers than the two fluctuating noises. As a result, they tended to talk louder in the pink and SSnoise to be heard by the listener (manikin).

At the noise level of 70 dBA (Table 5.3), an interaction between the HPD and the type of noise was observed as the decrease of speech level for the talkers by 5.1 dB, 3.2 dB, 4.8 dB, and 2.9 dB in pink, IFFM, SSnoise and helicopter noises, respectively, when comparing the condition of talkers with HPD to the condition of open ears (no matter if the target listener is in open ears or wearing hearing protection). At the noise level of 85 dBA (Table 5.3), decreases in speech level by 7.8 dB, 6.6 dB, 8.7 dB, and 6.4 dB in pink, IFFM, SSnoise and helicopter noises, respectively, were found.

The speech level data in noise were statistically evaluated using a four-way mixed ANOVA. There are three within-subjects factors (ear condition, noise type, and noise level) and one between-subjects factor (gender). The Huynh-Feldt corrected probabilities (*F* value) are reported when the Mauchly's sphericity test was violated. Significant main effects were found for ear condition ($F(2.0,44.2)=60.8$, $p<0.001$), noise type ($F(2.3,50.7)=102.2$, $p<0.001$) and noise level ($F(1,22)=267.1$, $p<0.001$), but not for gender ($F(1,22)=0.12$, $p=0.73$). The two-way interactions of ear condition with noise type ($F(7.5,165.3)=22.8$, $p<0.001$), ear condition with noise level ($F(2.6,58.2)=32.7$, $p<0.001$) and noise type with noise level ($F(3,66)=60.1$, $p<0.001$) were also statistically significant, but not ear condition with gender ($F(2.0,44.2)=2.8$, $p=0.08$), noise type with gender ($F(2.3, 50.7)=1.2$, $p=0.3$), nor noise level with gender ($F(1.0, 22.0)=1.3$, $p=0.3$). The three-way interaction of ear condition with noise type and noise level was significant ($F(7.2,158.6)=2.8$, $p=0.008$), but not the other three-way interactions nor the four-way interaction.

Multiple pairwise t-test comparisons among the four ear conditions (pooled over noise type, noise level and gender) using Bonferroni correction (p=0.05/6=0.008) show that the speech level in each ear condition was significantly different from that in other conditions, except for "Open-Open" vs. "HPD-HPD" conditions which did not reach significance. Multiple pairwise t-test comparisons among the different types of noises (pooled over ear condition, noise level and gender) using Bonferroni correction (p=0.05/6=0.008) also confirmed that the speech level in each noise type differed significantly relative to that in all other noise types.

5.3.2 Pitch

5.3.2.1 Quiet condition

The pitch of speech signals produced in the quiet condition was calculated by PRAAT (v.5.3.55) and is illustrated in Figure 5.5 and Table 5.4 for each talker-listener ear condition and separately for males and females. The general trend is similar to that for the speech level in Figure 5.2.

In the situation where the target listener was open ears, the mean pitch increased by 4.5 Hz and 8.6 Hz for male and female respectively when the talker wore the protector relative to open ears. When the target listener wore hearing protection, the wearing of HPDs by the talkers caused a rise of 6.6 Hz in the mean pitch of male participants while a decrease of 6.7 Hz was observed in the mean pitch for the females relative to open ears. Comparing "Open-Open" and "Open-HPD" conditions revealed that the effect of wearing hearing protection by the listener resulted in an increase in the mean pitch of open-ear male and female talkers by

11.6 Hz and 48.9 Hz, respectively. An increase of 13.7 Hz and 33.6 Hz in pitch was found when comparing "HPD-Open" to "HPD-HPD" condition.

A two-way mixed ANOVA on the speech pitch data revealed a significant main effect for ear condition ($F(2.4,53)=27.7$, $p<0.001$) and talkers' gender ($F(2.4,53)=8.4$, $p<0.001$). The two-way interaction of ear condition with talkers' gender was also significant ($F(2.4,53)=8.4$, $p<0.001$). Similar to the speech level data, multiple pairwise t-test comparisons (pooled over gender) using Bonferroni correction (p=0.05/6=0.008) indicated that pitch in "Open-Open" vs "HPD-Open" and in "Open-HPD" vs. "HPD-HPD" ear conditions did not differ significantly while all other pairwise comparisons were significant. Figure 5.5 shows the interaction between ear condition and gender. When the target listener is with open ears, the males and females both increased pitch frequency when wearing the HPD compared to open ears. In contrast, when the target listener wears the HPD, the males increased while the females decreased pitch frequency.

5.3.2.2 Noise condition

Figure 5.6 and Table 5.5 illustrate the mean pitch of speech produced by male and female talkers separately in the noise levels of 70 dBA and 85 dBA. In the "Open-Open" condition, the pitch of female talkers increased by 59.4 Hz, 46.6 Hz, 50.5 Hz and 39 Hz in pink, IFFM, SSnoise and helicopter noise at 70 dBA, respectively, relative to the quiet (Tables 5.5 and Table 5.4). Corresponding pitch increases by 29.6 Hz, 19.7 Hz, 22.9 Hz and 17 Hz were found for male talkers. At the noise level of 85 dBA relative to the quiet condition, female talkers raised the pitch by 121.2 Hz, 74.5 Hz, 104.1 Hz and 90.6 Hz in pink, IFFM, SSnoise and helicopter noise while it increased by 81.5 Hz, 56 Hz, 70.9 Hz and 62.4 Hz for male

talkers. The results follow the trends for speech level (Figure 5.3, Tables 5.2 and 5.3); both speech level and pitch are higher in noise than in quiet and increase with noise level when the talker is open ears.

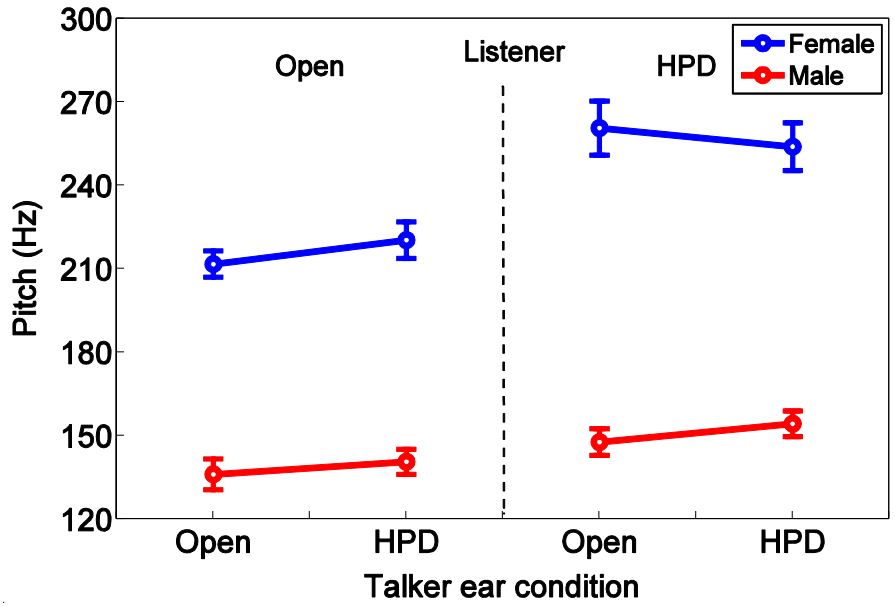


Figure 5.5: *The average pitch of speech (Hz) for four different ear conditions of the talker and listener in quiet (open ears and with a hearing protector) in quiet (two upper lines: female talkers, two lower lines: male talkers). Error bars are the standard error of the mean: ± standard deviation/√sample size.*

TABLE 5.4: *The mean and standard deviation (std.) of pitch (Hz) for four different ear conditions of the talker and listener in quiet*

Gender	Pitch	Talker-Listener ear condition			
		Open-Open	HPD-Open	Open-HPD	HPD-HPD
Females (n=12)	mean	211.7	220.3	260.6	253.9
	std.	(16.4)	(22.8)	(33.8)	(29.9)
Males (n=12)	mean	136.1	140.6	147.7	154.3
	std.	(19.1)	(15.5)	(16.8)	(15.8)

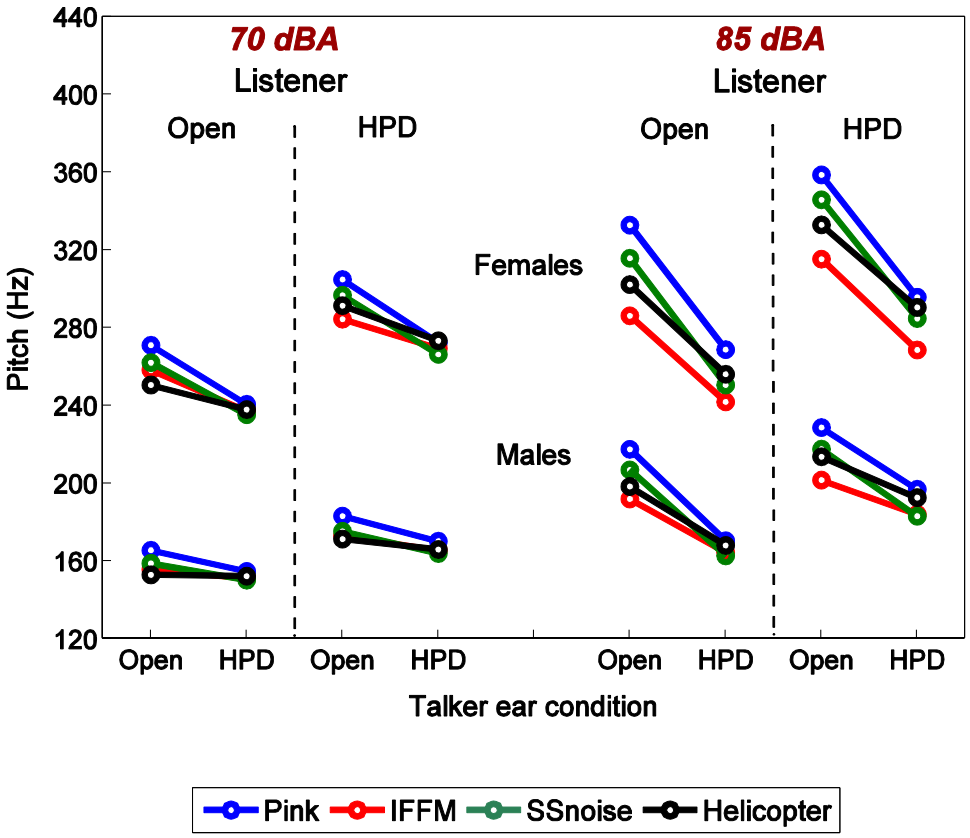


Figure 5.6: *The average pitch of speech (Hz) in four different noises at two levels of 70 dBA and 85 dBA for four different ear conditions of the talker and listener (with and/or without a hearing protector) (upper lines: female talkers, lower lines: male talkers).*

In the situation where the target listener was open ears, the pitch decreased by an average of 22.7 Hz for the females and by 6.2 Hz for the males across the different types of noises at 70 dBA when the talkers wore the protector relative to open ears. When the target listener wore hearing protection, the decline in the pitch arising from wearing the protector by the talkers was 24.0 Hz for the females and 9.4 Hz for the males. At the noise level of 85 dBA, the size of decrease in the pitch when the talkers wore the protector were 55 Hz (listener with open ears) and 57.8 Hz (listener with protector) for the females, and 37.3 Hz (listener with open ears) and 26.2 Hz (listener with protector) for the males.

TABLE 5.5: *The pitch (Hz) for a) female and b) male talkers and listener for four different ear conditions in four different noise conditions*

a) female talkers (n=12)

Noise Level:		70 dBA				85 dBA			
Noise Type:		Pink	IFFM	SSnoise	Helicopter	Pink	IFFM	SSnoise	Helicopter
Open-Open	mean	271.1	258.3	262.2	250.7	332.9	286.2	315.8	302.3
	std.	(33.6)	(33.5)	(36.5)	(31.1)	(35.0)	(32.2)	(40.2)	(37.0)
HPD-Open	mean	240.7	237.3	235.4	238.1	268.8	241.9	250.6	256.1
	std.	(34.7)	(33.2)	(32.1)	(35.0)	(44.1)	(31.0)	(35.4)	(39.9)
Open-HPD	mean	304.9	284.5	296.8	291.4	358.7	315.4	345.9	332.9
	std.	(39.3)	(32.3)	(43.2)	(41.9)	(47.0)	(36.8)	(47.5)	(45.2)
HPD-HPD	mean	272.5	269.6	291.4	273.4	295.8	268.6	284.9	290.5
	std.	(44.2)	(44.1)	(47.6)	(50.7)	(56.1)	(42.0)	(49.3)	(52.1)

b) male talkers (n=12)

Noise Level:		70 dBA				85 dBA			
Noise Type:		Pink	IFFM	SSnoise	Helicopter	Pink	IFFM	SSnoise	Helicopter
Open-Open	mean	165.7	155.8	159.0	153.1	217.6	192.1	207.0	198.5
	std.	(17.0)	(17.5)	(18.3)	(17.5)	(34.4)	(26.6)	(32.3)	(29.5)
HPD-Open	mean	154.8	151.2	150.3	152.4	170.5	164.6	162.8	168.2
	std.	(20.4)	(20.5)	(21.8)	(23.6)	(34.5)	(30.5)	(31.1)	(32.7)
Open-HPD	mean	183.3	173.2	175.4	171.5	228.7	201.7	217.7	213.8
	std.	(32.2)	(29.7)	(28.4)	(25.3)	(40.8)	(34.7)	(37.2)	(38.1)
HPD-HPD	mean	170.3	165.5	164.1	166.1	197.0	184.2	183.2	192.7
	std.	(23.8)	(19.4)	(20.5)	(22.6)	(41.0)	(36.0)	(35.3)	(40.1)

The effect of wearing the hearing protector by the listener was an increase in the mean pitch of the talkers' speech in noise. When the talkers were open ears, they raised their pitch by 33.9 Hz (females) and 17.5 Hz (males) at the noise level of 70 dBA, and by 28.9 Hz(females) and 11.7 Hz (males) at the noise level of 85 dBA due to this listener HPD effect. When the talkers were wearing the protector the size of increase in the pitch arising from wearing the protector by the listener was 32.6 Hz (females) and 14.3 Hz (males) at the noise level of 70 dBA, and 30.6 Hz (females) and 22.8 Hz (males) at the noise level of 85 dBA.

A four-way mixed ANOVA was performed on the pitch data in noise. The Huynh-Feldt corrected probabilities (*F* value) are reported when the Mauchly's sphericity test was violated. The analysis showed that there are significant main effects for ear condition (*F(2.5,55.4)=39.8, p<0.001*), noise type (*F(2.8,62)=58.6, p<0.001*), noise level (*F(1,22)=130.5, p<0.001*) and talkers' gender (*F(1,22)=68.9, p<0.001*). The two-way interactions of ear condition with noise type (*F(7.3,160.1)=18.3, p<0.001*), ear condition with noise level (*F(3,66)=31.9, p<0.001*), ear condition with gender (*F(2.5, 55.4)=4.1, p=0.02*), noise type with noise level (*F(2.6, 57.8)=43.1, p<0.001*) and noise type with gender (*F(2.8, 62)=5.5, p=0.003*) were statistically significant, but not noise level with gender (*F(1.0, 22.0)=0.004, p=0.95*). The three-way interactions of ear condition with noise type and noise level (*F(6.8,149.7)=2.9,p=0.008*) and of noise type with noise level and gender (*F(2.6, 57.8)=8.0, p<0.001*) were statistically significant, but not the other three-way interactions nor the four-way interaction.

Multiple pairwise t-test comparisons (pooled over noise type, noise level and gender) showed that the pitch produced in the four ear conditions was significantly different from each other ear condition at the p=0.05/6=0.008 level, except for "Open-Open" condition

compared to "HPD-HPD" condition. Multiple pairwise t-test comparisons (pooled over ear condition, noise level and gender) among the different types of noises revealed that the pitch in each noise type was significantly different from that in other noise types, except in SSnoise compared to helicopter noise.

5.3.3 One-third octave-band spectrum analysis

Figure 5.7 shows the distribution of speech produced by male and female talkers in the quiet condition over one-third octave frequency bands. The male talkers produced higher speech levels than the female talkers in the lower frequency bands up to 200 Hz. Conversely, the female talkers have slightly higher speech levels than the male talkers in the frequency bands from 200 Hz up to about 630−800 Hz when comparing the same ear condition. In the frequency bands over 800−1000 Hz, the female talkers had higher speech levels than the male talkers when the listener condition was HPD, but both genders produced nearly identical levels when the listener condition was open ears. As expected from the overall energy results (Figure 5.1), it was observed that the talkers, whether males or females, had slightly higher speech levels over most frequency bands in the "HPD-Open" compared to the "Open-Open" talker-listener ear condition. Moreover, the speech band levels were slightly higher in "HPD-HPD" relative to those in "Open-HPD" for the male talkers, but the opposite was observed for the female talkers particularly over the frequency bands from 1000 to 4000 Hz. This gender effect is consistent with the overall energy (Figure 5.2) and pitch (Figure 5.5) results.

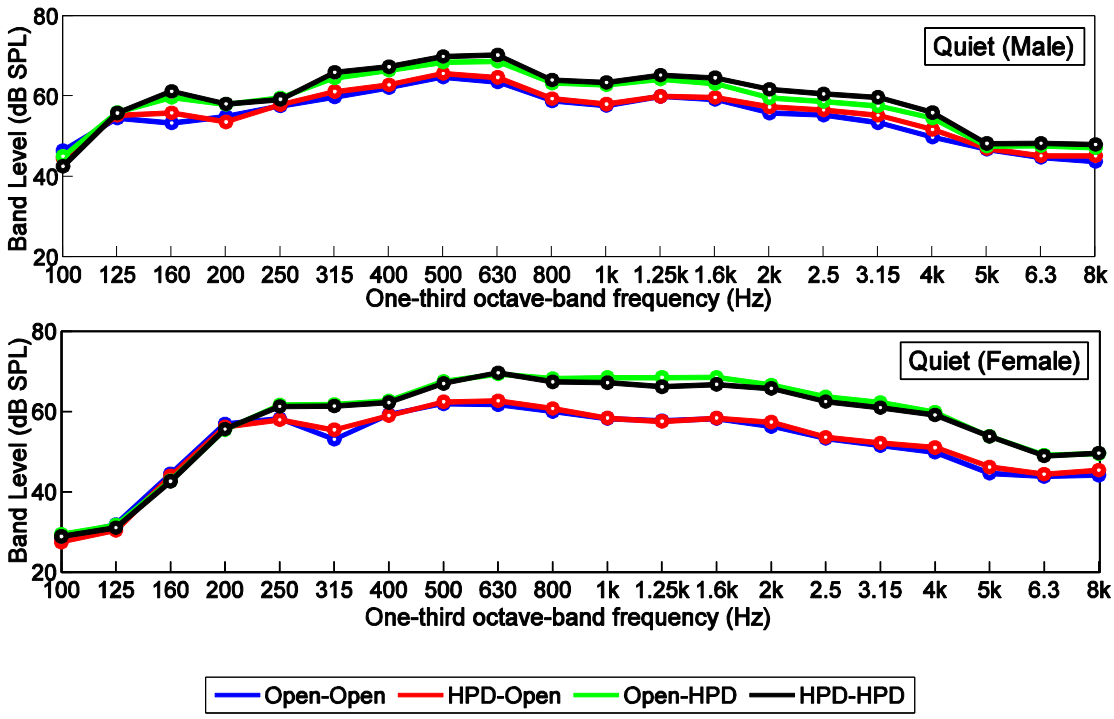


Figure 5.7: *The speech level distribution by one-third octave band (dB SPL) for males (top panel) and females (bottom panel) in the quiet condition for different ear conditions of the talker and listener (with and/or without a hearing protector).*

Figure 5.8 illustrates the spectrum of speech produced by males and females in pink noise at 70 dBA and 85 dBA, respectively. As expected, the level of speech was higher in each frequency band when the noise level was 85 dBA relative to 70 dBA. At the noise level of 70 dBA (Figure 5.8a), the male talkers had again much higher levels of speech than the female talkers in the lower frequency bands up to 200 Hz. At the higher noise level of 85 dBA (Figure 5.8b), these low-frequency spectral differences between males and females were not as pronounced.

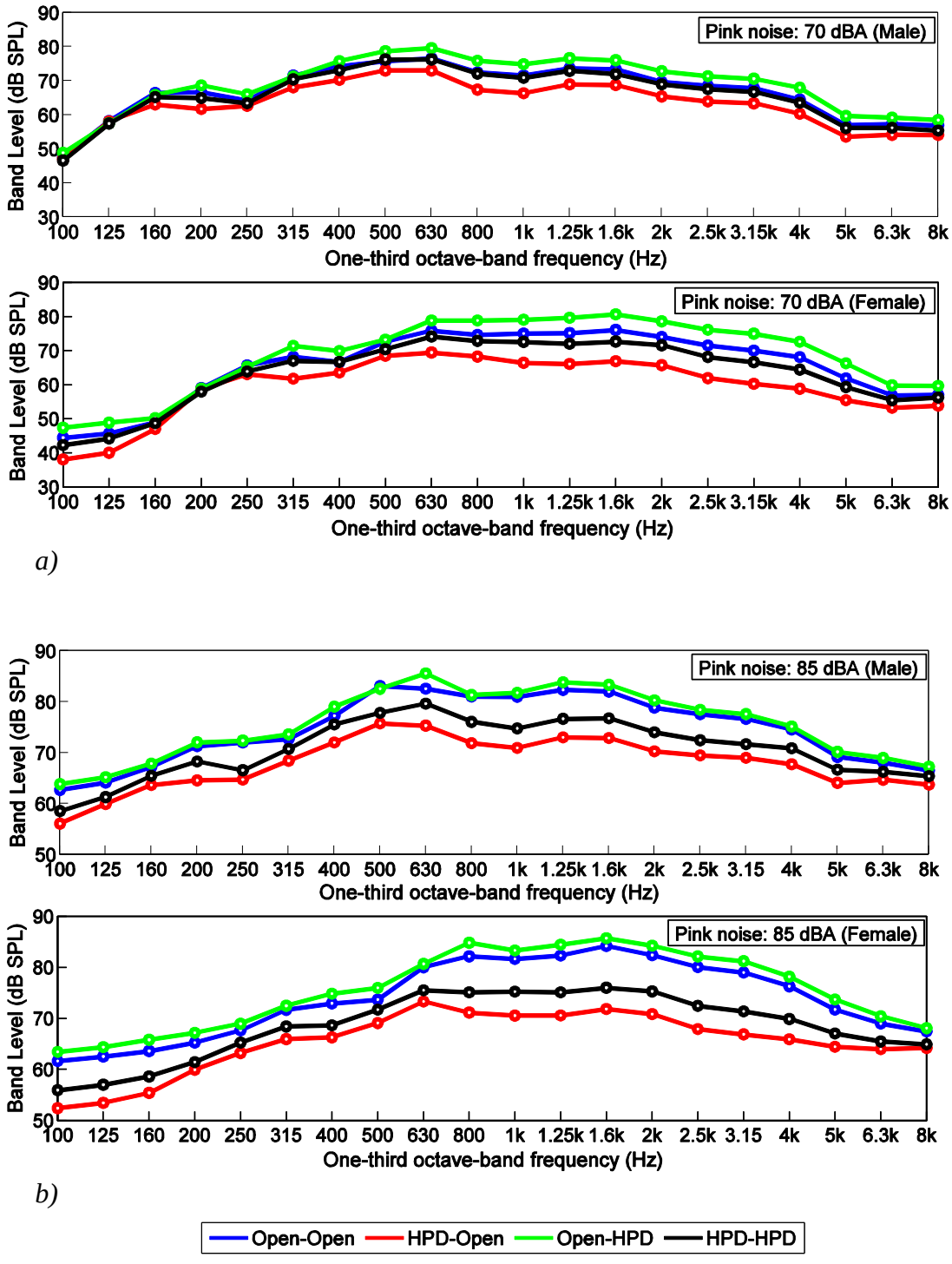


Figure 5.8: *The speech level distribution by one-third octave band (dB SPL) for males and females in a) 70-dBA and b) 85 dBA of pink noise for different ear conditions of the talker and listener (with and/or without a hearing protector).*

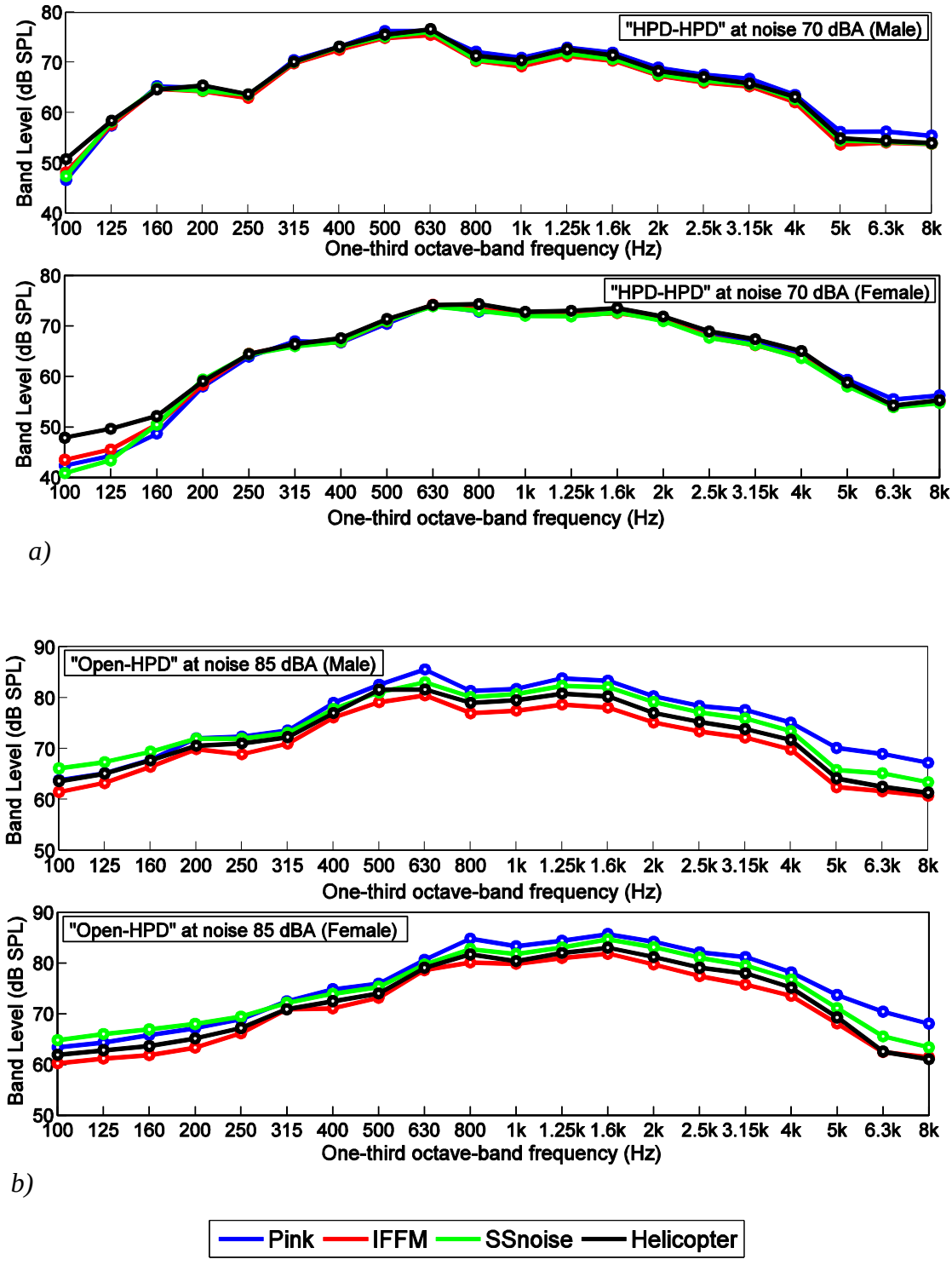


Figure 5.9: *The speech level distribution by one-third octave band (dB SPL) for males and females in four noises a) at 70 dBA for the "HPD-HPD" condition and b) at 85 dBA for the "Open-HPD" condition.*

At both noise levels, the lowest speech level in each frequency band was found in the "HPD-Open" condition, consistent with what was found for the overall energy results (Figure 5.4) and as reflected in the lower pitch (Figure 5.6). The highest speech level was observed in each one-third octave band in the "Open-HPD" condition, similar to the observations for the energy and pitch in Figure 5.4 and Figure 5.6, respectively. Figure 5.9 illustrates the effect of noise type on the spectral distribution of speech for males and female talkers for two separate combinations of ear condition/noise level. Figure 5.9a shows the results for the "HPD-HPD" ear condition at the noise level of 70 dBA where the talker is least exposed to noise. Whether males or females, the speech spectra are nearly identical to each other in the four noise types. As for the overall speech level (Figure 5.4) and pitch (Figure 5.6) results, wearing of the HPD did not induce the talkers to react appreciably to the noise type and change their voice spectrum in this situation. On the other hand, Figure 5.9b shows the results for the "Open-HPD" ear condition at the noise level of 85 dBA where the talker is most exposed to noise. In this higher noise level, a more prominent response to the type of noise by the unprotected talkers is now observed. As shown, whether males or females, the talkers speak louder in the two continuous noises (pink, SSnoise) than the two fluctuating noises (IFFM, helicopter).

5.4 Discussion

5.4.1 Effects of hearing protector on speech in quiet condition

Overall talkers' speech levels (pooled over gender) were slightly higher when wearing the HPDs in quiet (Figure 5.1) irrespective of whether the target listener was protected or unprotected; however, the effect was non-significant. This may be the consequence of two opposite effects occurring in this situation: (1) the occlusion effect of HPDs and (2) the attenuation of external sounds. The former effect causes the talkers to perceive their own voice louder due to the trapping of bone-conducted sounds within the ear cups, while the latter dampens the air-conduction pathway from their own voice. The small non-significant increase in speech levels observed in quiet may be indicative that the attenuation effect was slightly more important than the occlusion effect for the given HPD used (a mid-range earmuff). It is possible that the interplay between these two opposite effects would be different for different types of hearing protectors (earmuffs or earplugs) and the amount of volume enclosed or insertion depth of hearing protectors [173]. As an example, earmuffs with large cup sizes have typically less of an occlusion effect and larger attenuation value, which could promote higher speech levels, than earmuffs with smaller cup sizes.

The results verify that the use of an earmuff by the talker does not overly affect the speech produced in quiet (around 1 dB), in accordance with the data from earlier studies by Navarro (1996) [39] and Tufts and Frank (2003) [40]. Some authors reported a bigger size of change in speech level for the talkers in the order of 4−6 dB while wearing hearing protectors (earplug and earmuff) [25, 171]. As explained by Tufts and Frank (2003) [40], in the case of earplugs it might be due to the different amount of acoustic seal provided by the

different earplugs used which affects the size of attenuation of air-conducted component and the size of enhancement of bone-conducted component of speech.

In contrast, a large significant increase in talkers' speech level was noted when the target listener wore the HPD relative to open ears (Figure 5.1). Thus, when the talkers have knowledge that the target listener is protected, a higher speech level is produced in quiet perhaps in anticipation that the listener could have difficulty understanding speech. Overall, the highest level of speech in quiet was produced in "HPD-HPD" condition and the least one is related to "Open-Open" condition. No comparable data was uncovered in the literature on the effect of the listener's hearing protector in quiet; the results of this study show that knowledge that the listener is protected promotes higher talkers' speech levels, presumably to maintain intelligibility high enough.

It is to be noted that there is an interaction between ear condition and gender for speech level (Figure 5.2). While both genders increased speech levels when wearing the HPD if the listener was unprotected, males increased speech levels if the listener was protected but females decreased theirs. The same trends are seen in the pitch extracted from the talkers' speech (Figure 5.5). The females' pitch slightly decreases when the talkers wear the HPD for the ear-protected target listener ("HPD-HPD" vs. "Open-HPD" condition), similarly observed for the speech level. The similar effect is also observed in the spectra for males and females (Figure 5.7). The cause of these small differences in level, pitch and spectrum is not clear, but they are likely due to slight differences in vocal effort chosen by each gender.

5.4.2 Effects of hearing protector on speech in noise conditions

No matter the ear condition, the talkers' speech levels were higher in noise than in quiet and higher at the noise level of 85 dBA than 70 dBA, as shown in Figure 5.3, which was expected due to the commonly known Lombard effect. On the other hand, the effect of wearing an earmuff by the talkers in noisy conditions consistently led to a decrease in speech levels. When wearing the HPD in noise, the occlusion effect is still present as in quiet that amplifies the talker's own voice and gives the impression that the voice is loud, but at the same time the external noise is less intense due to the HPD's attenuation which dampens the Lombard effect. So, whereas in quiet the occlusion and Lombard effects could counteract each other, in noise both effects direct the talkers to speak less loudly.

The size of the decrease in speech level when the talkers wore the HPD varied depending on the level of noise; less reduction of speech level at the noise level of 70 dBA compared to that at the noise of 85 dBA (Table 5.3, Figure 5.4), as reported in earlier studies [40, 117]. The size of drop is up to 5.5 dB at the former noise condition while it is up to 10 dB at the latter one, in accordance with data shown by other studies [40, 117].

The decrease in speech level when the talker wore the HPD in noise also depended on the type of noise. The size of the effects was larger in the continuous noises than the fluctuating noises. No matter the ear condition, the talkers' tendency to increase voice levels in noise was consistently higher in pink noise and SSnoise than in the IFFM and helicopter noises (Figures 5.3 and 5.4). Fluctuating noises are known to be less masking than continuous noises [97, 98, 99] due to the phenomenon of "listening in the gaps" [97, 99, 103]. Listening to own voice may thus be facilitated by this "release from masking" effect promoting talkers to adjust their voice accordingly and speak not as loud in fluctuating than

continuous noises. This effect was found earlier in an earlier study for the unprotected ear condition [101], while in this study this observation is extended to conditions when the talkers wear HPD.

The effect of wearing the HPD on the talkers' speech level not only depended on noise type and noise level but also on the interaction between these two factors. At the lower level of noise (70 dBA), when the talkers wear the HPD the speech levels are very close to each other in all types of noise whether the listener is with the HPD or open ears (Figure 5.4). It is an indication that the talkers' speech levels are not strongly related to the noise in this condition ("HPD-Open" or "HPD-HPD" at the noise level of 70 dBA). When wearing the HPD, the residual noise under the earcups may be low enough (around 50 dBA assuming 20 dB attenuation) to appreciably dampen the Lombard effect so that talkers adjust their voice primarily due to the occlusion effect. At the higher level of noise (85 dBA), the HPD does not attenuate the noise enough (around 65 dBA assuming 20 dB attenuation) and the talker reacts to the noise as well as the occlusion effect. Thus, the speech levels are different in different noise types in "HPD-Open or "HPD-HPD" conditions at 85 dBA. More research is needed to document the effects of noise type and its interaction with noise level on the talkers' speech wearing the HPD since there is no comparable data available in the literature.

The effect of wearing an earmuff by the listeners in noisy conditions consistently led to an increase in speech levels. This is seen when the talker is open ears ("Open-Open" vs "Open-HPD") as well as when the talker is wearing the HPD ("HPD-Open" vs "HPD-HPD") as shown in Figure 5.4. On the other hand, when both the talker and listener wear HPDs compared to both open ears, then there is interplay between the effects of the talker and listener HPDs. At the lower level of noise (70 dBA), the speech level in "HPD-HPD" condition was slightly higher (IFFM, helicopter) or lower (pink, SSnoise) than that in

"Open-Open" condition depending on the type of noise, but there is always a drop in speech level between the same conditions at the noise level of 85 dBA. However, this interplay between talker-listener and use of HPD could depend on the type of HPD. Hormann et al. found lower speech levels when both talker and listener wear earplugs compared to when they both are open ears in pink noise at 76 dBA and 92 dBA, but not at 84 dBA [15].

The pitch increases in noise relative to quiet condition for both males and females (Table 5.5 vs. Table 5.4). Pitch patterns across noise types, levels and ear conditions in Figure 5.6 follows essentially the same patterns for speech level in Figure 5.4. Thus it seems pitch and level are dependent parameters and likely predicated by a common source: vocal effort. The HPD worn by the talkers causes that the pitch decreases while it increases when the target listener wears the HPD. At the lower level of noise (70 dBA), when both talker and listener are with the hearing protector, the raising effect of wearing the HPD by the listener on pitch is stronger than the damping effect of the HPD worn by the talker (Figure 5.6). Thus, there is a slight increase in pitch when comparing "HPD-HPD" to "Open-Open" for both males and females. At the higher level of noise (85 dBA), there is a slight to moderate decrease in pitch when comparing "HPD-HPD" to "Open-Open" for both males and females (Figure 5.6).

In noise, the speech spectra show the lowest level in each one-third octave frequency band in the "HPD-Open" condition while they indicate the highest level in each band in the "Open-HPD" condition (Figures 5.8a and 5.8b). Likewise, the female talkers have more high-frequency speech energy than males especially for the listener wearing the HPD (Figures 5.8a and 5.8b). At the lower level of noise (70 dBA), the negligible differences between the speech band levels across four different noises at "HPD-HPD" condition show the stronger role of occlusion effect of HPD than the noise type in adjusting the talkers' speech level (Figure 5.9a vs. Figure 5.9b). At the higher level of noise (85 dBA), the spectra

across noise types show the speech band levels being highest in the two continuous noises and lowest in the two fluctuating noises reflecting differences in the amount of noise masking inducing the Lombard effect.

5.5 Conclusion

In general, although the HPDs can protect hearing from damage due to noise, the effects of wearing the protector on speech communication need to be taken into account. This chapter evaluated the effects of the hearing protector on the talkers' speech produced in different noise types at different levels in an external sound field under conditions where both the talker and the target listener could wear or not the protector.

Ear condition was a main effect on the speech level and it interacted with noise level and type, but not gender. The effect of wearing an earmuff by the talkers in the high noise condition consistently led to a decrease in speech levels. Further, the size of adverse effects of hearing protectors on talkers' speech levels was larger for the continuous noises compared to the fluctuating ones. On the other hand, when the talker has knowledge that the listener wears hearing protection, speech levels were found to increase. These two factors are in play when communicating with hearing protectors and when both talkers and listeners wore HPDs, the speech levels decreased compared to when they did not. Information exchange can be significantly impeded when both individuals wear hearing protection compared to when they do not or when only one of the two conversation partners wear hearing protector [15].

The limitations of this study include: (1) the target listener was a manikin, so there was no interaction between the talker and listener which could have resulted in dynamic vocal adjustments by the talker; and (2) the participants were naive users of hearing protectors. Workers with experience using hearing protection may have developed strategies to speak louder to compensate for wearing an HPD in noise. In any case, this study shows it is important to inform workers about the effect of their HPDs, not just when listening but when they are talking, and to train them to speak louder when using HPDs.

The level-dependent hearing protection devices (HPD) or those equipped with active noise cancellation technology can be good alternatives for the conventional passive protectors (earmuffs and earplugs). The next chapter evaluates the effects of one such type of hearing protector on speech production in noise.

Chapter 6

The Effect of Level-dependent Hearing Protection on Speech Production in Quiet and Noise

6.1 Introduction

Conventional passive hearing protectors can impede speech perception and user awareness of the surrounding acoustic environment [19, 22, 24, 31, 32, 37, 49, 119]. They can also interfere with production of speech by altering the talker's auditory feedback of their own voice and the loudness of residual noise [15, 25, 27, 40, 41, 117, 171, Chapter 5]. By increasing the users' perception of their own voice (occlusion effect) and decreasing their

perception of the surrounding noise (attenuation effect), hearing protectors dampen the talkers' tendency to speak louder against the background noise, known as the Lombard effect [39].

Although hearing protectors are expected to be used in noisy conditions above 80-85 dBA, there are some situations where the talkers are wearing hearing protectors in quieter periods, such as when the noise sources are intermittent during a work shift (e.g. drilling, welding, and stoppages). It has been observed that talkers wearing hearing protectors in quiet produce higher speech level by 1−6 dB compared to open-ear condition [25, 41, 171, Chapter 5] while some authors reported a small decrease of 1 dB [39, 40]. Depending on the type of hearing protector (earmuff or earplugs) the size of volume enclosed or the depth of insertion into the ears would be different so that the different interplay between the two opposite effects (occlusion effect and attenuation effect) of the protector might cause a raise or a drop of speech level in quiet [173].

When wearing conventional hearing protectors in noise, unlike in quiet, the occlusion and attenuation effects described above both prompt the talkers to speak less loudly [Chapter 5]. Studies have reported a decrease of 3−11 dB [15, 25, 27, 40, 117, 171, Chapter 5] in speech level compared to open ears for different types of hearing protectors and noise conditions (e.g. type and level of noise). The size of the decrease when wearing the hearing protectors reduces as the level of noise decreases [40, 117], and was found to be as much as 5.5 dB at the noise level of 70 dBA while it was as much as 10 dB when the level of noise increased to 85 dBA [Chapter 5]. Larger decreases in speech level were also observed in continuous noises than in fluctuating noises with respect to open ear condition [Chapter 5]. The phenomenon of "listening in the gaps" [97, 99, 103], known to produce less masking in fluctuating than continuous noises at the same overall level [97, 98, 99], likely causes the

talkers to have a better perception of their own voice which promotes them to speak even less loudly in fluctuating noises [101].

Thus, conventional passive hearing protectors, which offer a fixed amount attenuation regardless of the level of incident sound, can affect the talker's speech level and impede the exchange of information between the talker and the listener [15]. The advent of electronic hearing protectors with level-dependent transmission gain into the marketplace offers some advantages over conventional passive protectors. By providing increasing amounts of attenuation with increasing noise level, they can provide protection against high level continuous noise and intense transient sounds while maintaining or even enhancing audibility and situational awareness in low level noise or quieter periods, especially when the noise is intermittent in the workplace and/or in the presence of workers with hearing loss [20, 21, 22, 23]. In a qualitative assessment, noise-exposed workers rated the level-dependent HPDs higher than the passive protectors in terms of their performance in improving communication and situational awareness [22].

Despite these advantages, the inevitable impacts of automatic gain control and the dynamics of the compression circuitry of level-dependent hearing protectors on speech production cannot be ignored. The available research on these devices is limited to the studies focussing on the listener, typically evaluating the impacts of different gain settings and level-dependent attenuation on the perception, recognition, detection and the localization of speech and other desirable sounds [17, 18, 23, 31, 33, 34, 35, 36, 37, 47, 120, 121, 122, 123, 124, 125]. Thus, there is a gap in knowledge regarding the effects of wearing level-dependent hearing protectors on speech production in quiet and in different noise conditions. Changing the gain of a level-dependent HPD is not expected to alter the occlusion effect, which for earplugs is primarily related to the depth of insertion, while for earmuffs it depends

on the size of the earcups. Thus, level-dependent HPDs are expected to produce the same amount of occlusion effect as their conventional passive counterparts. On the other hand, the air-conducted auditory feedback from own voice and the noise transmitted through the device is determined by transmission gain characteristics of the level-dependent circuitry. Thus, it is expected that speech production will be different with level-dependent HPDs compared to conventional passive protectors.

In this chapter, the changes in level, pitch and spectral distribution of speech arising from wearing an electronic level-dependent earmuff-style headset are evaluated in quiet and in two noise types presented at three different levels in an external noise field. It is expected that talkers will speak louder and closer to the level spoken in open ear conditions when the device operates in level-dependent mode than when the device acts as a conventional passive device with a larger attenuation. As mentioned above, the same occlusion effect is expected in both level-dependent and passive attenuation operation modes, but the external noise will be perceived as being louder in the level-dependent mode than in the passive operation and it should induce talkers to speak louder in noise due to the increased Lombard effect. Moreover, it is hypothesized that the effect of level-dependent HPD on the talker's speech level will depend on the type of noise, as for conventional hearing protection and open ear conditions, and be lower in fluctuating noises than the continuous noise types.

6.2 Methods

6.2.1 Participants

Twenty-four adults (12 males and 12 females) participated in the study. These participants were native English speakers, between 19 to 32 years of age, had normal hearing and tympanometry, with no history of smoking, speech/language impediments, and respiratory, postural or other chronic conditions. Normal hearing was defined as thresholds ≤ 15 dB HL for frequencies up to 4000 Hz in each ear and ≤ 25 dB HL at 6000 Hz and 8000 Hz. The history of each participant's exposure to noise was collected through a questionnaire. The project was approved by the University Ethics Board on research with human subjects.

6.2.2 Hearing protection device

The device used in this study is the Peltor Power Com Plus^{TM} level-dependent earmuff-style headset (3M, St. Paul, MN). The built-in electronics provide stereo pass-through listening to maintain audibility at low to moderate levels of noise and to protect hearing at high noise levels [21, 23]. The device can operate in one of five so-called "surround settings", S1 to S5, providing a gain control range of 18 dB. Low level sounds (<60 dBA) are slightly attenuated at the two lower surround settings S1 and S2, while they are amplified at surround settings S3 to S5. Higher sound levels (>60 dBA) are subjected to automatic gain control in all surround settings which reduces the transmission gain and increases attenuation to protect hearing against high noise levels [21, 23]. Figure 6.1 shows the input-output function of the device in the 5 surround settings. When the surround mode is turned off (not shown in Figure

6.1), the device behaves like a passive HPD with a NRR of 25 dB, with a mean attenuation of 19 dB to 39 dB from 125 Hz to 8000 Hz (as per laboratory-based attenuation data reported by the manufacturer). In this study, the protector was used in passive (surround off) and two level-dependent modes (surround settings S1 and S4). The former provides the maximum attenuation of external sounds, as listed above, and the latter two a transmission gain of about −6 dB and +9 dB, respectively, at low sound levels and gradually more attenuation with increasing sound level (see Figure 6.1)

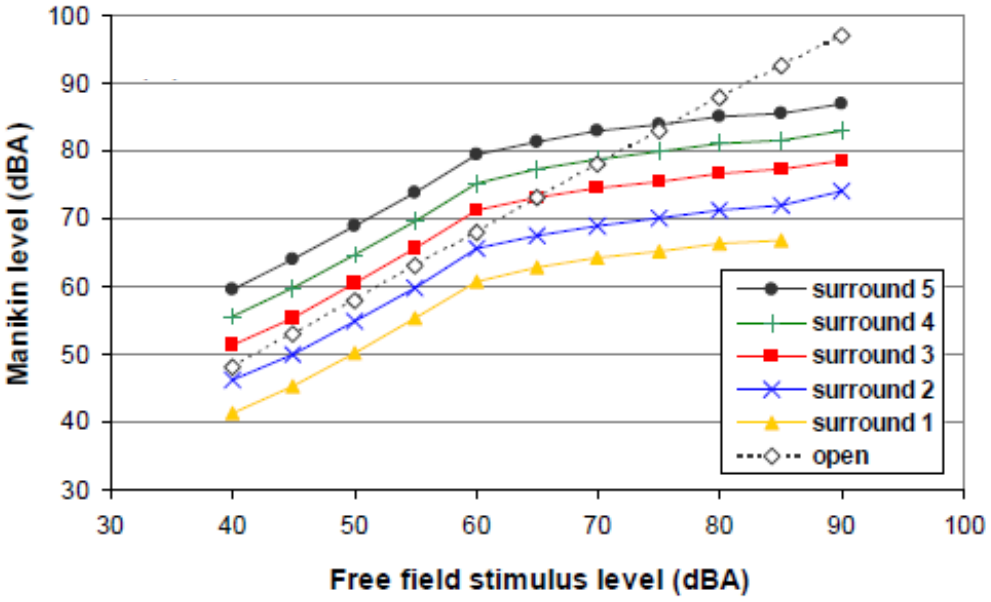


Figure 6.1: *Input-output level functions for speech-spectrum noise in the surround modes (S1-S5) of level-dependent earmuff used in the tests (Reproduced from [21, 23] with permission). The transmission gain at any surround setting and stimulus level is given is the difference between the manikin sound level with the device in position and the open ear response.*

6.2.3 Materials and Procedures

Two types of noises, pink and pulse pink, were chosen in this study (Figure 6.2). They have exactly the same spectral content (Figure 6.2.b) but one is continuous and the second one is temporally fluctuating (Figure 6.2.a). In order to generate pulse pink noise, the continuous pink noise was modulated by a pulse waveform of width 50 ms with exponential decay with a time constant of 10 ms, at a rate of 16 Hz. The noise stimulus was a concatenated sequence consisting of pink and pulse pink noises, each 27 second long separated by a 5 second pause. The noise stimulus sequence was calibrated at three different levels of 55 dBA, 70 dBA, and 85dBA at the talker's head position in a noise simulation room.

The noise simulation room with the dimension of 4.29 m × 3.65 m × 2.42 m (length × width × height) was located in the Hearing Research Laboratory at the University of Ottawa. In order to generate a quasi-diffuse sound field, the equalized and amplified noise signal was played through six loudspeakers including four tower type (Sony MF600H), a bookshelf speaker on a stand (Yamaha NS-C155) and a subwoofer (Mirage Sub12) [16] in the room by an external sound card (TASCAM US-366, TEAC Co., Japan) at the sampling rate of 96 kHz.

A free-field omnidirectional microphone (B&K Type 4189) connected to a sound level meter (B&K Type 2235) at 25cm away from the talker's mouth and the same external soundcard was used to pick up the signals in the room at the same sampling rate of 96 kHz. This acquisition configuration was found to be the best to extract speech successfully and accurately in an external noise field (Chapter 4).

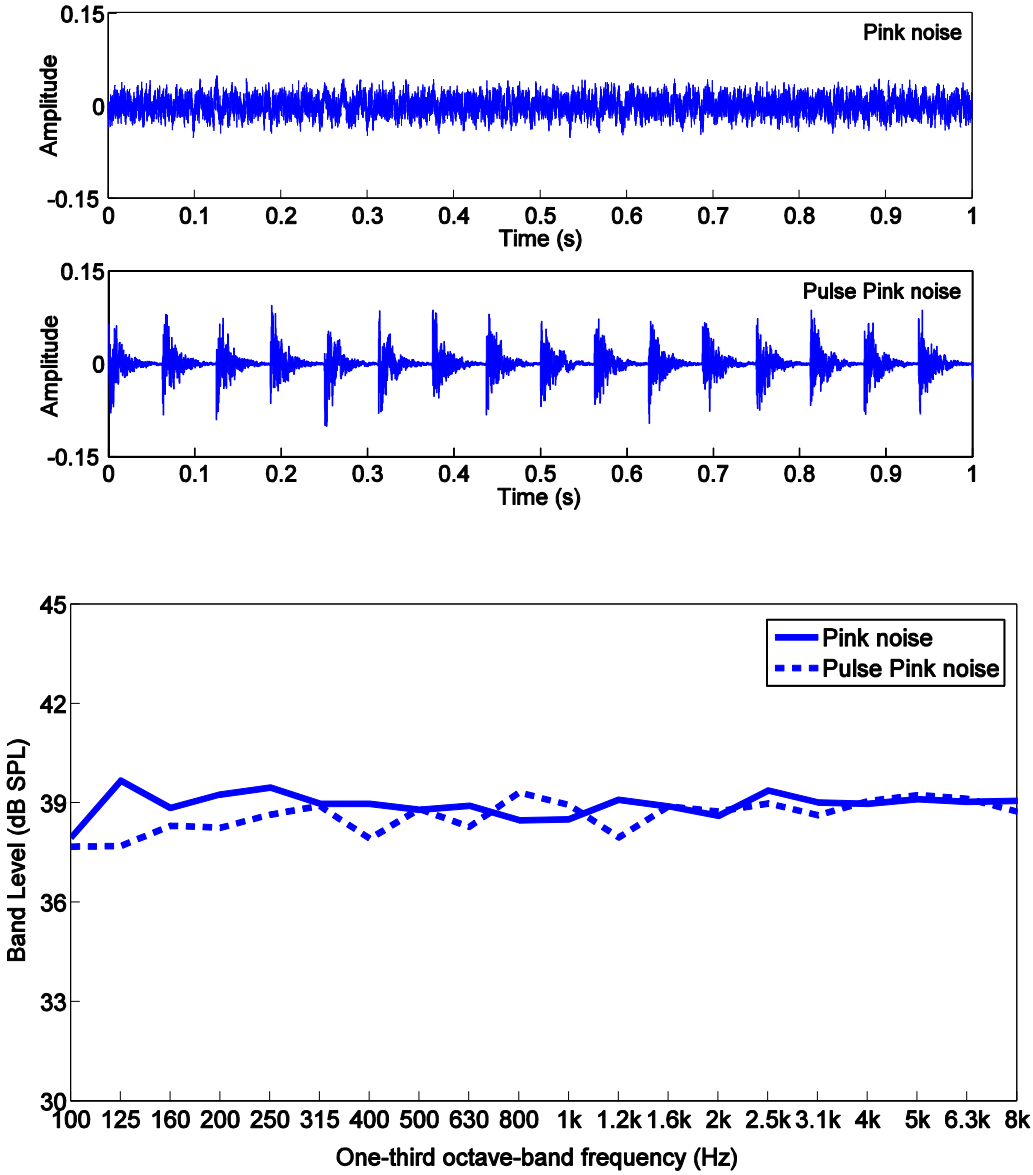


Figure 6.2: *(a) The time-domain waveforms, and (b) one-third octave-band spectra of the two noises used in the study***.**

The participant sat near the centre of recording room 1-m away from an ANSI S3.36 acoustic manikin (B&K 4128) which served as a target listener (Chapter 5). Since vocal output depends on the communication situation facing the talkers [56], the presence of the manikin at a fixed distance prompted all talker participants to adjust their voice to a common scenario. The first 2 second of each noise was assigned for the talker's adaptation to the

noise condition and then during the remaining 25 second of the noise the participant was asked to read the list of 12 sentences attached to the manikin's torso and was instructed to speak in such a manner to be understood at the target distance of the manikin in the different noises. The sentences which selected from the American English Hearing-In-Noise Test (HINT) represented a conversation in noisy workplaces (e.g. *The engine is running*) (Table6.1) [148]. The participant was monitored during the test by the experimenter in the control room outside the recording room.

The participants' output speech signal was recorded in quiet and in the two noises presented at three different levels. There were four ear conditions for the talkers: (1) open ears ("OpenEar"), (2) when the level-dependent HPD was worn with surround setting S4 ("HPD-S4"), (3) when wearing the level-dependent HPD with surround setting S1 ("HPD-S1"), (4) when wearing the level-dependent HPD in the "Off" mode ("HPD-Off"). The order of ear conditions (4), noise types (2), and noise levels (2) was counterbalanced across the participants. In order to limit the number of experimental conditions and the session duration to avoid vocal fatigue, only one manikin target "listener" situation could be tested. The open ear condition was selected.

In order to recover the clean speech signal from the noisy vocal output collected in the external noise field, the active noise cancellation (ANC) method was used because it was verified to be able to adequately remove the noise from the vocal output at SNRs as low as −10 dB (as explained in Chapter 4). The LMS algorithm, with different step size values depending on the SNR, was used to adaptively update the filter coefficients. Based on the methodology of using the ANC method provided in Chapter 4, the noise signal needs to be sent twice in the simulation room, one for recording the reference noise while the talker is asked to be silent and one when the talker is speaking.

TABLE 6.1: *The HINT sentences chosen as speech material for the talkers*

No.	HINT sentences
1	The engine is running.
2	He climbed up the ladder.
3	A janitor swept the floor.
4	The buckets fill up quickly.
5	It's getting cold here.
6	There are branches everywhere.
7	Someone is crossing the road.
8	A fire truck is coming.
9	The paint dripped on the ground.
10	A bag fell off the shelf.
11	The exit is well lit
12	The scissors are very sharp

6.3 Results

6.3.1 Overall speech energy level

6.3.1.1 Quiet condition

Figure 6.3 and Table 6.2 show the mean overall speech levels produced in quiet for different talkers' ear conditions. The lowest speech level (71.1 dB SPL) occurred when the talkers wore the HPD at the surround setting S4 ("HPD-S4"). As shown in Figure 6.1, the talkers reduced their speech level by 1.2 dB (Figure 6.3) at the surround setting S4 compared to the

unprotected condition ("OpenEar"). In the surround setting S1, the talkers' speech level slightly increased compared to that in "OpenEar" condition. The highest speech level (73.1 dB SPL) was produced when the HPD was worn in the passive mode ("HPD-Off"). Table 6.2 and Figure 6.4 show that there was a similar trend due to wearing the HPD in both male and female talkers, except that the males had less differences between ear conditions, especially at surround setting S4 ("HPD-S4"). Also, males spoke about 3-6 dB louder than females across the ear conditions.

A two-way mixed ANOVA was conducted on the speech level data, with ear condition and gender as within-subject and between-subject factors, respectively. There was a significant main effect for ear condition ($F(3,66)=14.7, p<0.001$) and gender ($F(1,22)=6.9, p=0.015$). The interaction between ear condition and talkers' gender was also significant ($F(3,66)=3.8, p=0.015$), as illustrated in Figure 6.4. Multiple pairwise t-test comparisons using Bonferroni correction (p=0.05/6=0.008) showed there was no effect between ear conditions for males, but for females speech levels in the "HPD-S4" were significantly lower than in the three other ear conditions.

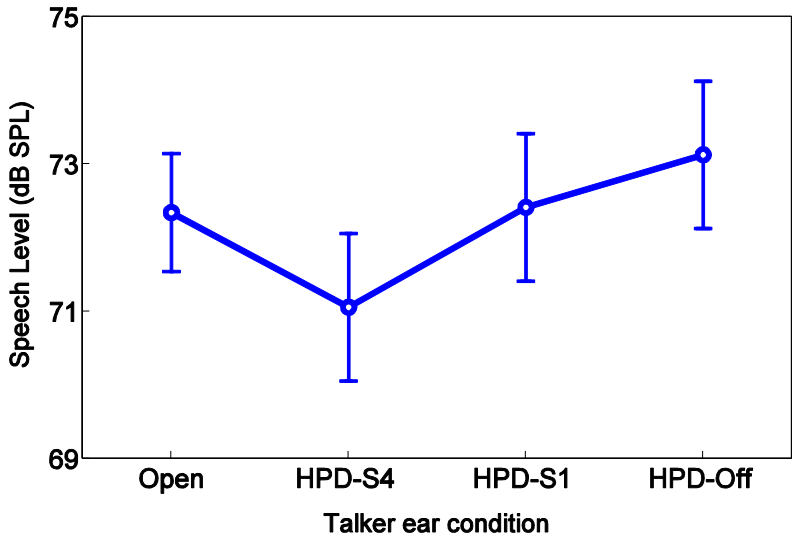


Figure 6.3: *The mean overall speech level (dB SPL) averaged over both genders for four different talker ear conditions in quiet (Open ears and with the level-dependent hearing protector at surround gain settings S1 and S4 and at surround Off). Error bars are the standard error of the mean: ± standard deviation/√sample size)*

TABLE 6.2: *The mean and standard deviation (std.) of the speech level (dB SPL) separately for males and females and when averaged over both genders for different talker ear conditions in quiet*

Gender	Speech level	Talker ear condition			
		OpenEar	HPD-S4	HPD-S1	HPD-Off
Females (n=12)	mean (std.)	70.7 (4.1)	68.4 (4.3)	70.2 (4.4)	71.0 (4.4)
Males (n=12)	mean (std.)	74.0 (3.4)	73.7 (4.0)	74.6 (4.1)	75.3 (4.1)
Average (n=24)	mean (std)	72.3 (4.0)	71.1 (4.9)	72.4 (4.7)	73.1 (4.7)

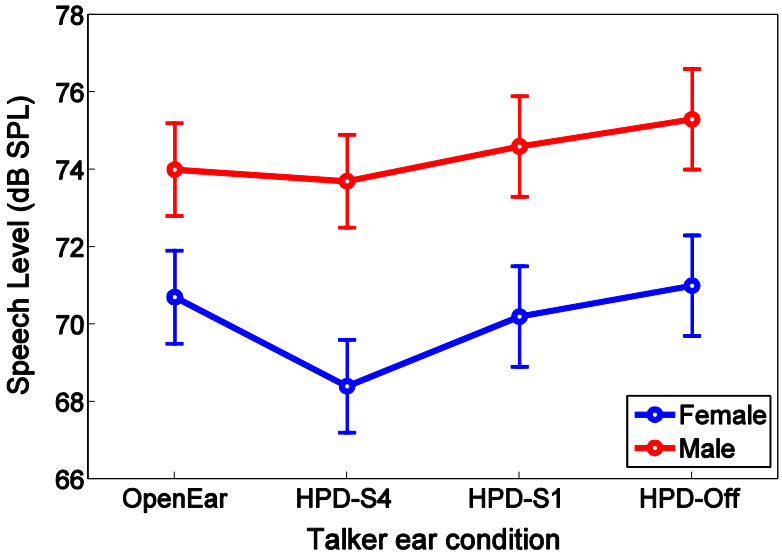


Figure 6.4: *The mean speech level (dB SPL) separately for males and females for four different talker ear conditions in quiet (Open ears or with a level-dependent hearing protector at three surround gain settings). Error bars are the standard error of the mean: ± standard deviation/√sample size)*

6.3.1.2 Noise condition

Figure 6.5 shows that the talkers, no matter the ear condition, increased their speech level when speaking in noise compared to quiet condition, conforming to the Lombard effect. The speech level increased more when the level of noise increased from 55 dBA to 85 dBA. When the talkers were with open ears in noise, the increase in speech levels compared to quiet condition was 4.3, 9.7, 17.1 dB at the three noise levels of 55, 70 and 85 dBA, respectively, averaged over the two noise types (Table 6.3). The corresponding increase in speech levels was 4.0, 8.1 and 13.0 dB at the “HPD-S4” condition, 3.1, 6.4 and 9.4 dB at the "HPD-S1" condition, and 1.9, 4.0 and 7.1 dB at the "HPD-Off" condition in the three noise levels compared to quiet. The highest speech levels in the presence of noise were produced

in the "OpenEar" condition, indicating that wearing the HPD dampens the vocal output in noise not only in passive mode but also in the two level-dependent modes.

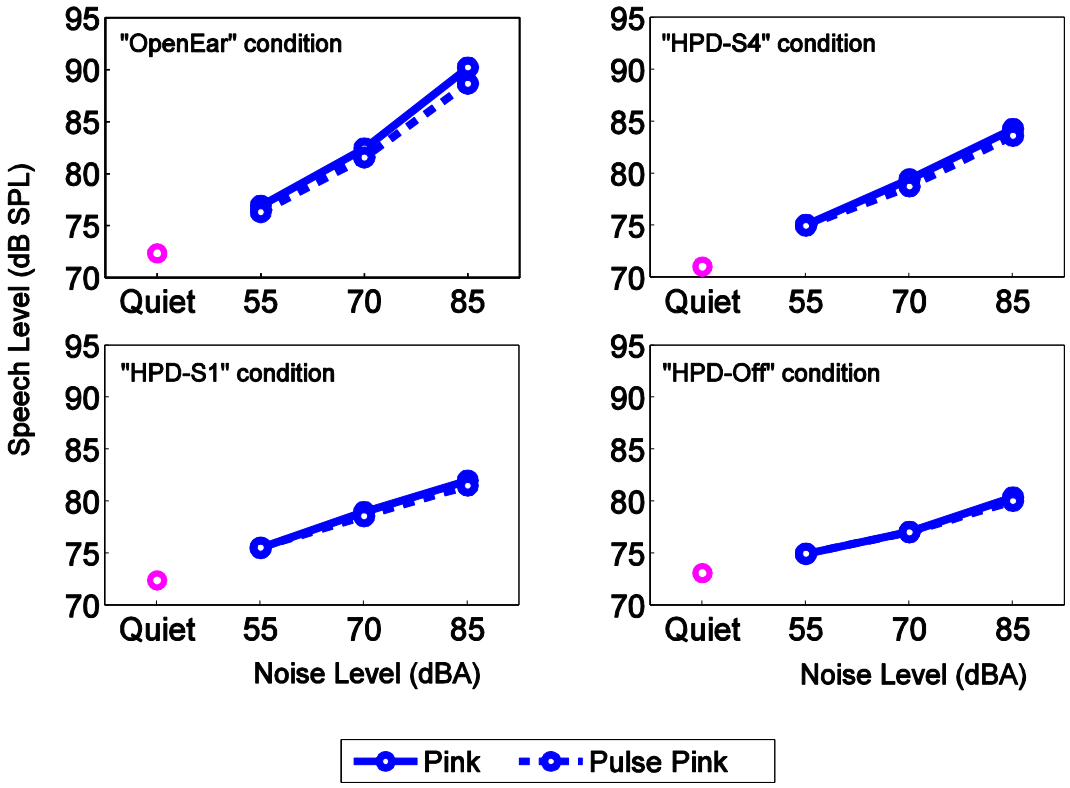


Figure 6.5: *Mean overall speech level (dB SPL) averaged over both genders for four different talker ear conditions (open ears and a level-dependent HPD at three surround gain settings) in quiet and in two different noises at three levels.*

Figure 6.6 shows the same data as in Figure 6.5 so as to illustrate that the effect of HPD on the speech level depended on the surround gain setting of HPD, at the three different noise levels. In 55 dBA of noise, the speech level decreased by 1.6 dB and 1.3 dB at "HPD-S4", 1.4 dB and 0.8 dB at "HPD-S1", and 2.0 dB and 1.3 dB at "HPD-Off" in continuous and pulse pink noises, respectively, when the talkers wore the HPD relative to open ears. At the noise level of 70 dBA, the corresponding sizes of decrease were 3.0 dB and 2.8 dB at"HPD-

S4", 3.4 dB and 3.0 dB at "HPD-S1", and 5.3 dB and 4.5 dB at "HPD-Off" relative to open ears in the same two noises respectively. At the noise level of 85 dBA, the talkers' speech levels decreased by 5.9 dB and 5.0 dB at “HPD-S4” relative to the "OpenEar" condition at the two noise types (pink and pulse pink) of 85 dBA. When comparing the "HPD-S1" condition to the "OpenEar" condition dBA, the size of drop in speech level were 8.2 dB and 7.1 dB, while they were 9.8 dB and 8.6 dB when the HPD was at the passive mode ("HPD-Off" vs. "OpenEar").

TABLE 6.3: *Overall speech levels (dB SPL), averaged over both genders (n=24), for four different talker ear conditions in two different noise types at three levels (mean and standard deviation)*

Noise Level:		55 dBA		70 dBA		85 dBA	
Noise Type:		Pink	Pulse Pink	Pink	Pulse Pink	Pink	Pulse Pink
OpenEar	mean	76.9	76.3	82.4	81.6	90.2	88.7
	(std.)	(4.2)	(4.4)	(3.4)	(3.5)	(3.7)	(3.8)
HPD-S4	mean	75.1	75.0	79.5	78.8	84.3	83.7
	(std.)	(4.8)	(5.0)	(4.8)	(5.1)	(4.3)	(4.7)
HPD-S1	mean	75.5	75.6	79.0	78.6	82.0	81.5
	(std.)	(5.0)	(5.2)	(4.6)	(4.9)	(3.8)	(4.1)
HPD-Off	mean	75.0	75.0	77.1	77.0	80.4	80.1
	(std.)	(5.1)	(5.0)	(4.9)	(5.3)	(4.9)	(5.1)

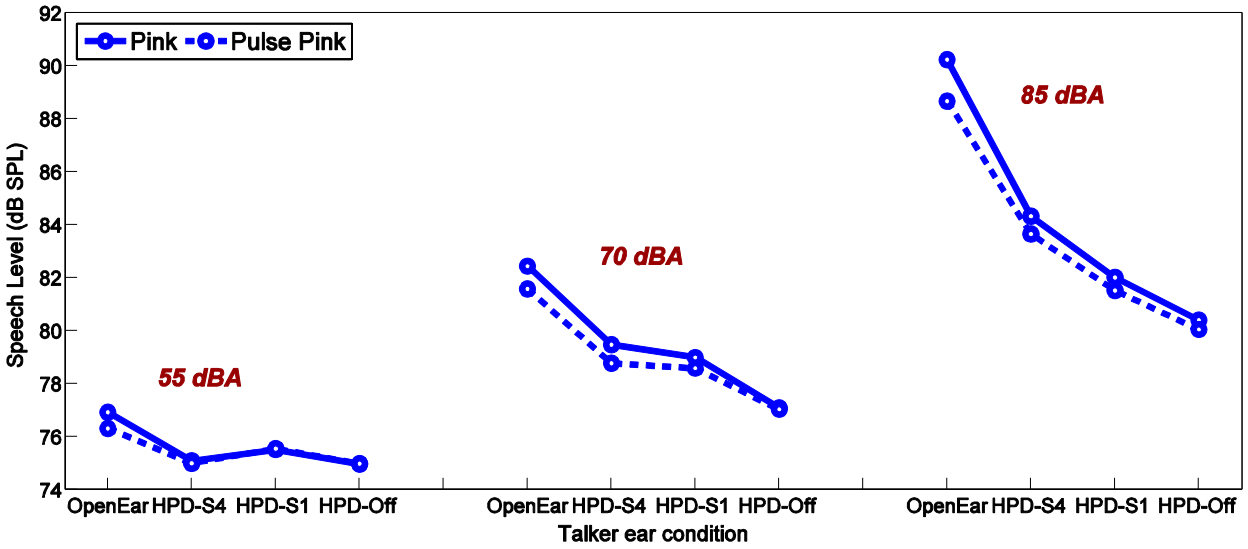


Figure 6.6: *Mean overall speech level (dB SPL) averaged over both genders in two noise types at three levels for four different ear conditions of the talker (open ears and with a level-dependent HPD at three surround gain settings).*

The effect of HPD on the speech level also depended on the level of noise. At the noise level of 85 dBA, the maximum difference in speech levels between the protected ear conditions and open ears was 9.8 dB and was observed in the "HPD-Off" condition. At the noise levels of 70 dBA and 55 dBA, the maximum difference was 5.3 dB and 2.0 dB, respectively, again in the "HPD-Off" condition.

As shown in Figure 6.6, the effect of noise type appeared as a slightly lower speech level in the fluctuating pulse pink noise compared to that in the continuous pink noise, especially, at the high noise levels of 70 dBA and 85 dBA. The speech level was lower by 0.9 dB ("OpenEar"), 0.7 dB ("HPD-S4"), 0.4 dB ("HPD-S1"), and 0.1 dB ("HPD-Off") at the pulse pink noise level of 70 dBA relative to the same level of continuous pink noise. At the noise level of 85 dBA, the speech level was lower by 1.6 dB ("OpenEar"), 0.7 dB ("HPD-S4"), 0.5 dB ("HPD-S1"), and 0.4 dB ("HPD-Off") in pulse relative to continuous pink noise. The size

of differences between speech levels across two types of noises decreased as the attenuation of the HPD increased (or transmission gain decreased) from "HPD-S4" to "HPD-S1" and "HPD-Off". At the low noise level of 55 dBA, there were only very small differences in speech levels over the two types of noises when wearing the HPD.

A four-way mixed ANOVA was conducted on the speech level data. In the analysis, ear condition, noise type, and noise level were within-subjects factors and gender was a between-subjects factor. The Huynh-Feldt corrected probabilities (*F* value) are reported for the main effects of ear condition and noise level and the interaction of ear condition and noise level, since the Mauchly's sphericity test was violated. There were significant main effects for ear condition ($F(2.5, 54.1)=75.6$, $p<0.001$), noise type ($F(1,22)=10.1$, $p=0.004$), and noise level ($F(1.2, 25.6)=233.6$, $p<0.001$), but not gender ($F(1,22)=3.6$, $p=0.07$). The two-way interactions of ear condition with noise level ($F(4.1,91.1)=57.4$, $p<0.001$), ear condition with noise type ($F(3,66)=8.8$, $p<0.001$), noise type with noise level ($F(2,44)=7.4$, $p=0.002$), and ear condition with gender ($F(2.5,54.1)=3.6$, $p=0.025$) were statistically significant, but not noise level with gender ($F(1.2,25.6)=2.2$, $p=0.15$), and noise type with gender ($F(1,22)=0.12$, $p=0.73$). The three-way interaction of ear condition with noise level and gender ($F(4.1,91.1)=57.4$, $p<0.001$) was significant, but not the other three-way interactions nor the four-way interaction.

Multiple pairwise t-test comparisons among three different levels of noises (pooled over ear condition, noise type and gender) using Bonferroni correction (p=0.05/6=0.008) show that the comparison between speech energy levels in each pair of noise levels was statistically significant. Multiple pairwise t-test comparisons among the four ear conditions (pooled over noise type, noise level and gender) using Bonferroni correction (p=0.05/6=0.008) show that the comparison between the speech levels in all pairs of ear

conditions were significant, except in "HPD-S4" vs. "HPD-S1". However, given the strong effect of noise level on results (Figure 6.6), the same analysis was repeated separately by noise level. At the noise level of 55 dBA, the differences between speech levels in all pairs of ear condition were non-significant. In 70 dBA and 85 dBA of noises, all paired comparisons were significant except for "HPD-S4" vs. "HPD-S1" at the former noise level.

6.3.2 Pitch

6.3.2.1 Quiet condition

Pitch frequency was calculated by PRAAT (v.5.3.55) and is shown in Figure 6.7 and Table 6.4 for the four different ear conditions in quiet, separately for females and males. It is observed that the general trend is similar to that for the speech level in Figure 6.4. When the talkers wore the HPD at the surround setting S4 ("HPD-S4"), the pitch decreased by 5.3 Hz and 1.4 Hz for females and males, respectively, compared to the open ear condition. When the HPD was worn at the surround setting S1 ("HPD-S1"), there was an increase in pitch by 1.1 Hz for females and a small drop of 0.4 Hz for males. Comparing the "HPD-Off" condition to the "OpenEar" condition showed an increase of 7.1 Hz in the females' pitch and a decrease of 0.5 Hz in the males' pitch.

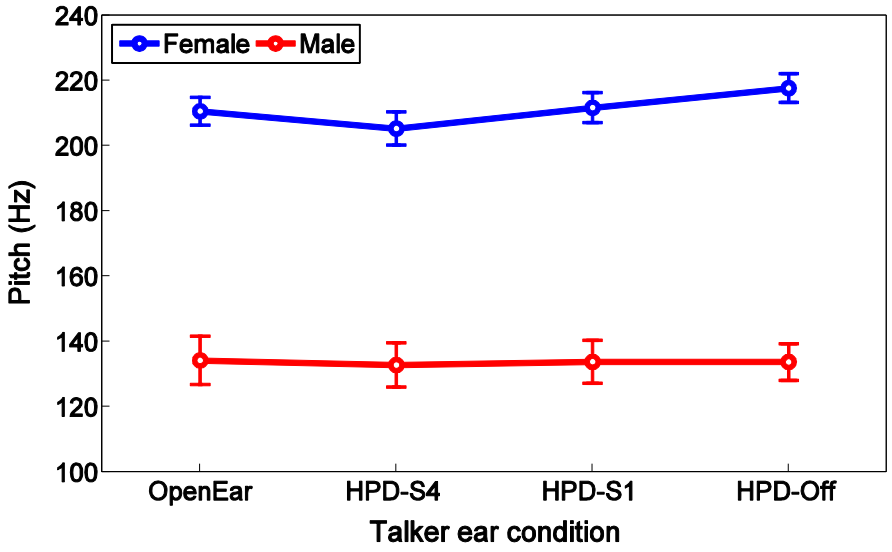


Figure 6.7: *The mean pitch (Hz) for four different ear conditions of the talkers in quiet (open ears and with a level-dependent hearing protector at three surround gain settings). Error bars are the standard error of the mean: ± standard deviation/√sample size.*

TABLE 6.4: *The mean and standard deviation (std.) of pitch (Hz) for four different talker ear conditions in quiet*

Gender	Pitch	Talker-Listener ear condition			
		OpenEar	HPD-S4	HPD-S1	HPD-Off
Females (n=12)	mean	210.6	205.3	211.7	217.7
	std.	(14.8)	(17.6)	(16.1)	(15.2)
Males (n=12)	mean	134.2	132.8	133.8	133.7
	std.	(25.6)	(23.7)	(22.7)	(19.4)

A two-way mixed ANOVA was conducted on the pitch data in quiet. The Huynh-Feldt corrected probabilities (*F* value) are reported since the Mauchly's sphericity test was violated. There was a significant main effect for ear condition ($F(2.42,53.24)=4.2$, $p=0.015$) and gender ($F(1,22)=100.74$, $p<0.001$). The two-way interaction of ear condition with

gender was also significant (*F(2.42,53.24)=3.2, p=0.041*). Multiple pairwise t-test comparisons using Bonferroni correction (p=0.05/6=0.008) showed that the pitch in each ear condition did not differ significantly relative to that in other ear conditions, whether this analysis was done separately for females or males, or done for the pooled gender data.

6.3.2.2 Noise condition

The mean pitch produced by male and female talkers in two noise types presented at three different levels is illustrated in Figure 6.8 and Table 6.5. Comparing to "OpenEar" condition, it is observed that the pitch decreased for both males and females when wearing the HPD at the higher noise level of 85 dBA, and at 70 dBA except for the males at "HPD-S4" and "HPD-S1" in pulse pink noise at that level.

The size of decrease in female pitch was 12.1 Hz and 13.2 Hz in continuous and pulse pink noises of 70 dBA, and 49.1 Hz and 39.0 Hz respectively at 85 dBA, when comparing ""HPD-S4" condition relative to the “OpenEar” condition. For the same comparison of ear conditions, the male pitch decreased by 4.6 Hz in continuous pink noise while it increased by 0.6 Hz in pulse pink noise of 70 dBA. At the noise level of 85 dBA, the male pitch in "HPD-S4" condition reduced by 33.0 Hz and 28.1 Hz relative to "OpenEar" condition in continuous and pulse pink noises, respectively.

When wearing the HPD at the surround setting S1 ("HPD-S1") relative to the "OpenEar" condition, the female pitch reduced by 14.5 Hz and 12.5 Hz in 70 dBA of continuous and pulse pink noises and by 56.7 Hz and 49.1 Hz in 85 dBA of the two same noises, respectively. There was a drop of 3.3 Hz but an increase of 2.4 Hz in the male pitch in 70

dBA of continuous and pulse pink noises when comparing "HPD-S1" to "OpenEar". At 85 dBA, the male pitch reduced by 46.7 Hz and 38.9 Hz in the two noises.

Comparing "HPD-Off" condition to "OpenEar" condition shows a decrease of 23.7 Hz and 21.7 Hz in female pitch, and 10.3 Hz and 5.5 Hz in male pitch in 70 dBA of continuous and pulse pink noises. At the noise level of 85 dBA, the size of drop was 73.2 Hz and 61.4 Hz (pink and pulse pink) in female pitch and 46.4 Hz and 39.2 Hz (pink and pulse pink) in male pitch. In contrast, at the low noise level of 55 dBA, wearing the HPD did not have much effect on pitch compared to the other two noise levels (Figure 6.8).

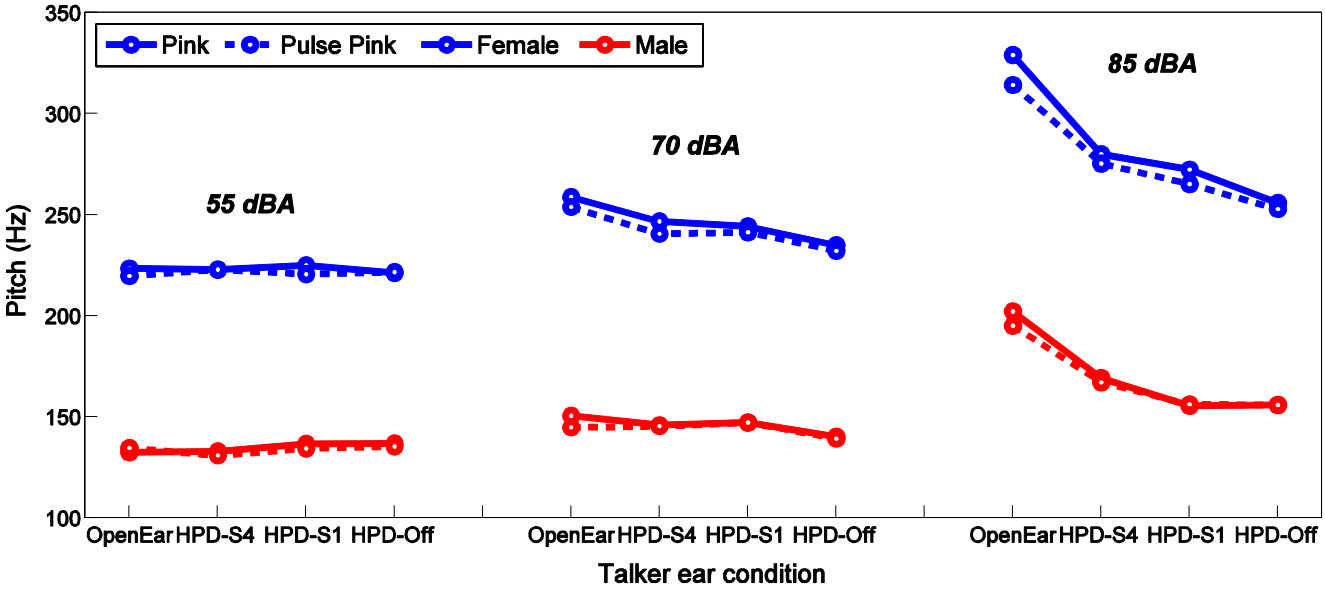


Figure 6.8: *The mean pitch (Hz) in two different noises at three levels of 55 dBA, 70 dBA and 85 dBA for four different talker ear conditions (open ears and with a level-dependent hearing protector at three surround gain settings) (upper lines: female talkers, lower lines: male talkers).*

TABLE 6.5: *The pitch (Hz) for a) female and b) male talkers for four different talker ear conditions in two different noise types*

a) female talkers (n=12)

Noise Level:		55 dBA		70 dBA		85 dBA	
Noise Type:		Pink	Pulse Pink	Pink	Pulse Pink	Pink	Pulse Pink
OpenEar	mean	223.5	219.7	258.8	253.9	329.1	314.4
	(std.)	(14.8)	(16.2)	(14.4)	(21.2)	(34.2)	(33.7)
HPD-S4	mean	222.9	222.9	246.8	240.7	280.0	275.4
	(std.)	(15.9)	(20.4)	(24.9)	(25.8)	(29.5)	(31.0)
HPD-S1	mean	225.0	220.7	244.3	241.5	272.4	265.2
	(std.)	(16.7)	(19.2)	(21.1)	(25.2)	(31.8)	(30.7)
HPD-Off	mean	221.3	221.7	235.1	232.2	255.9	252.9
	(std.)	(21.4)	(22.9)	(19.8)	(22.4)	(26.1)	(28.5)

b) male talkers (n=12)

Noise Level:		55 dBA		70 dBA		85 dBA	
Noise Type:		Pink	Pulse Pink	Pink	Pulse Pink	Pink	Pulse Pink
OpenEar	mean	132.6	134.7	150.6	144.9	202.2	195.2
	(std.)	(19.2)	(20.5)	(16.6)	(15.1)	(39.6)	(34.6)
HPD-S4	mean	133.0	131.0	146.0	145.5	169.2	167.1
	(std.)	(18.7)	(21.6)	(19.1)	(23.6)	(24.2)	(28.8)
HPD-S1	mean	136.7	134.5	147.2	147.3	155.6	156.3
	(std.)	(22.5)	(21.0)	(24.7)	(27.0)	(22.8)	(27.8)
HPD-Off	mean	136.9	135.5	140.2	139.3	155.9	156.0
	(std.)	(20.7)	(23.5)	(20.8)	(22.1)	(35.1)	(37.6)

A four-way mixed ANOVA was performed on the pitch data in noise. The Huynh-Feldt corrected probabilities (*F* value) are reported since the Mauchly's sphericity test was violated, except for the two-way interaction of noise type with noise level and the three-way interaction of ear condition with noise type and noise level. There were significant main effects for ear condition ($F(2.6,56.8)=35.6$, $p<0.001$), noise type ($F(1,22)=13.1$, $p=0.001$), noise level ($F(1.1,25.1)=126.4$, $p<0.001$) and gender ($F(1,22)=136.3$, $p<0.001$). The two-way interactions of ear condition with noise type ($F(2.1,45.7)=5.6$, $p=0.006$), ear condition with noise level ($F(3.5,76.3)=44.2$, $p<0.001$), ear condition with gender ($F(2.6, 56.8)=3.0$, $p=0.045$), noise type with noise level ($F(2,44)=5.9$, $p=0.005$), and noise level with gender ($F(1.1, 25.1)=7.5$, $p=0.009$) were statistically significant, but not noise type with gender ($F(1,22)=3.0$, $p=0.095$). The three-way interactions of ear condition with noise type and noise level ($F(6,132.0)=2.8$, $p=0.02$) was statistically significant, but not noise type with noise level and gender ($F(2,44)=2.6$, $p=0.09$), ear condition with noise level and gender ($F(3.5,76.3)=1.7$, $p=0.17$), and ear condition with noise type and gender ($F(2.1,45.7)=1.06$, $p=0.40$) nor the 4-way interaction ($F(5.1,113.3)=1.3$, $p=0.30$).

Multiple pairwise t-test comparisons (pooled over ear condition, noise type and gender) revealed that the pitch at each noise level was significantly different from that at other noise levels. Multiple pairwise t-test comparisons (pooled over noise type, noise level and gender) showed that the difference between the pitch in each ear condition and that in others was statistically significant at the p=0.05/6=0.008 level (using Bonferroni correction), except for the pitch in "HPD-S4" vs. "HPD-S1" and "HPD-S4" vs. "HPD-Off". However, given the strong effect of noise level on results (Figure 6.8), the same analysis was repeated separately by noise level. At the noise level of 55 dBA, all paired comparisons were non-significant. In 70 dBA of noise, the speech level differences in "OpenEar" vs. "HPD-Off" and "HPD-S1"

vs. "HPD-Off" were significant but not for other pairs of ear condition. At the noise level of 85 dBA, the comparison between the speech levels in all pairs of ear condition were significant except for "HPD-S1" vs. "HPD-Off".

6.3.3 One-third octave-band spectrum analysis

Figure 6.9 illustrates the distribution of speech produced by male and female talkers over one-third octave frequency bands in three levels (55, 70, and 85 dBA) of continuous and pulse pink noises. As expected, the speech level was higher in each frequency band in the higher noise level of 85 dBA relative to 70 dBA and 55 dBA. Across ear conditions, the highest speech levels over the frequency bands corresponded to the condition of "OpenEar" in each of three different noise levels. Further, speech frequency band levels generally decreased with greater attenuation (or lesser transmission gain) from "HPD-S4" to “HPD-S1” and then to "HPD-Off", mostly prominently in the highest noise levels of 70 dBA and 85dBA.

In Figure 6.9, male talkers produced higher speech levels than the female talkers in the lower frequency bands up to 200 Hz, especially at the lower noise levels of 55 dBA and 70 dBA. In 55 dBA of noise, there was not a big difference between speech levels of males and females over frequency bands higher than 200 Hz when comparing the same ear condition. When wearing the HPD at the lower noise level of 55 dBA, the talkers' speech frequency band levels remained close to each other for different surround settings on the HPD in the two noise types.

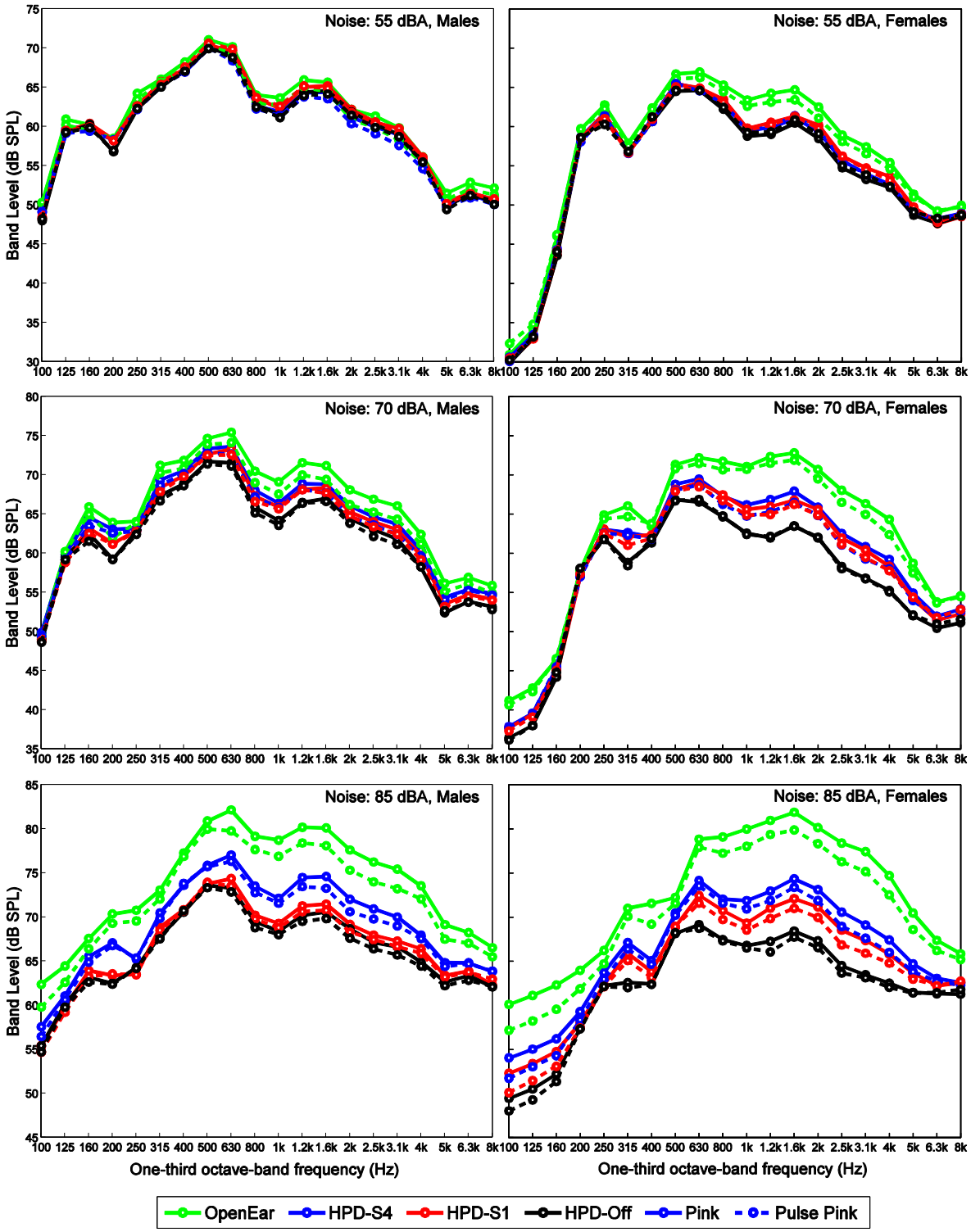


Figure 6.9: *The speech level distribution by one-third octave band (dB SPL) for males and females in 55 dBA, 70 dBA, and 85 dBA of continuous pink (solid lines) and pulse pink (dashed lines) noises for different talker ear conditions (open ears and with a level-dependent hearing protector).*

6.4 Discussion

6.4.1 Effects of level-dependent hearing protector on speech in quiet condition

The effect of level-dependent hearing protector on speech energy level in quiet was depended on the ear condition (Figures 6.3 and 6.4); however, the effects were small. At the surround setting with the highest transmission gain (S4), there was a small decrease (about 1.2 dB over both genders; 0.3 dB non-significant effect for males, 2.3 dB significant effect for females) in the speech level compared to the open ear condition. It can be hypothesized that the female voices were favorably transmitted in quiet at this high amplification setting and that female talkers wearing the protector at the surround setting S4 heard their own voice louder than that in open ears. Thus, combined with the occlusion effect, this may have promoted female talkers to talk at a lower level than in open ear condition.

At the lower surround setting (S1) and passive mode (Off), it is hypothesized that the occlusion and attenuation effects counteracted each other leading to no significant changes in speech levels compared to open ear condition for males and females. No comparable data was uncovered in the literature on the effect of level-dependent HPD on speech level in quiet.

The trend of pitch extracted from the talkers' speech (Figure 6.7) was seen to be similar to that of speech level (Figure 6.4), but less pronounced.

6.4.2 Effects of level-dependent hearing protector on speech in noise condition

The speech level of talkers, either with open ears or with hearing protectors, increased consistently in noise relative to quiet condition (Figure 6.5). The Lombard effect drove the talkers to raise the vocal output to overcome the masking effect of background noise. There was a higher increase in speech level as the noise level increased.

When wearing the level-dependent HPD in noise, the talkers' speech level reduced as much as 10 dB compared to open ear condition. The size of decrease depended on the surround setting chosen on the HPD. At the high surround setting S4, the protector is able to amplify the low level incident sounds (<60 dBA) while attenuating loud sounds (>60 dBA). In contrast, the HPD at low surround setting S1 attenuates the incident sounds, no matter low or high level, but the attenuation gain changes depending on the level of sound. The level-dependent protector equally attenuates the sounds at any levels when wearing at the passive mode ("HPD-Off"). It was found that the speech level had the smallest decrease with respect to open ears at "HPD-S4" condition compared to "HPD-S1" and "HPD-Off" conditions. This may be because the talkers had a better perception of the surrounding noise which weakened the dampening effect of HPD on the Lombard effect. At the low surround setting S1, apart from the occlusion effect which was also present at high surround setting S4, the HPD attenuation decreasing the amount of external noise heard reduced the talkers intention to speak loud enough against the noise. At the passive mode, the decrease in speech level arising from the attenuation effect is even higher since the HPD provided the largest size of attenuation at this mode compared to the surround S1. The results showed that the speech level produced when wearing the level-dependent HPD at the surround modes (S4 and S1) is

significantly different than that at the Off/passive mode of the level-dependent HPD, but the difference between the speech levels at the two surround modes of S4 and S1 was not significant (S4 v. S1). Moreover, the result of this study verified that the level-dependent HPD at higher surround setting can more presumably maintain speech communication in noise while simultaneously protecting the hearing against high noise level.

The effect of level-dependent HPD on speech level depended on the level of noise (Figure 6.6). At the low noise level of 55 dBA, there were no significant differences in speech levels when wearing the HPD at different operation modes ("HPD-S4", "HPD-S1", and "HPD-Off"). It was because the level of noise was too low to elicit a strong Lombard effect and when wearing the HPD the talkers did not respond to the noise but rather they adjusted their voice primarily due to the occlusion effect and/or perception of their own voice. When the level of noise increased, the effect of wearing the HPD and the influence of surround setting on speech level became more prominent. The largest differences in speech level were thus observed when wearing the HPD at the high noise level of 85 dBA. Overall, while none of the HPD conditions promoted speech levels as strong in open ear conditions, the level-dependent settings of the HPD produced higher speech levels in noise than in passive mode. Further, the benefit of level-dependent settings is stronger at higher noise levels and the higher the transmission gain setting (e.g. S4 vs. S1).

The talkers spoke at a slightly lower level in the fluctuating pulse pink noise compared to that in the continuous pink noise (Figure 6.5 and 6.6). This is likely related to the fact that fluctuating noises are less masking than continuous noises [97, 98, 99] due to the phenomenon of "listening in the gaps" [97, 99, 103]. Thus, the talkers who have a better perception of their own voice because of this "release from masking" effect speak not as loud in fluctuating than continuous noises. The talkers reaction to the noise type when

speaking is affected by the wearing of HPD and the level of noise. The effect of noise type on speech level reduced when wearing the HPD compared to the open ears, because of the decrease in the perception of background noise (Figure 6.6). The difference between speech levels across two types of noises became smaller when the surround setting of the HPD was lowered from S4 to S1 or turned to "Off". As expected, the smallest difference was observed at the condition of "HPD-Off" where the talkers did not react appreciably to the noise type because of the high attenuation of the HPD at this passive mode (Figure 6.6). The higher attenuation of external noise at low surround setting of the HPD causes the talkers to not appreciate the change in noise type. At the low noise level of 55 dBA, the differences between speech levels produced in the two types of noise at "HPD-S4", "HPD-S1", and "HPD-Off" conditions were smaller than at higher noise levels. No matter the surround setting, talkers reacted mostly to the HPD occlusion effect and/or their own voice transmitted through the HPD, not the noise, at this very low noise level. The effect of noise type was observed to be more prominent at the high noise levels of 70 dBA and 85 dBA where the talkers could more clearly perceive the noises even in the highest level of protection ("HPD-Off" condition).

The pitch increased in noise compared to quiet condition for both males and females (Table 6.5 vs. Table 6.4). Figure 6.8 shows that the pattern for pitch is the same as that for the speech level in Figure 6.7, especially at high noise levels of 70 dBA and 85 dBA. At very low noise level (55 dBA), the HPD setting selected only had a small non-significant effects in pitch across different ear conditions of "HPD-S4", "HPD-S1", and "HPD-Off" (Figure 6.8). At the higher noise level (85 dBA), the S4 and S1 conditions promoted a higher pitch than in the passive mode (Off), but not as much as that seen in the open ear condition.

The speech spectra show that the largest effect size of level-dependent HPD on speech level in each one-third octave frequency band was at high noise levels (70 dBA and 85 dBA) (Figure 6.9). The highest speech band levels were observed at the "OpenEar" condition while the lowest band levels were related to the "HPD-Off" condition. The smallest decrease in speech levels was observed when the HPD was worn at the high surround setting (S4 vs. S1 and passive mode). The effect of noise type on adjusting the talkers' speech level, higher band levels in the continuous noise and lower in the fluctuating noise, was larger at high noise levels (70 dBA and 85 dBA) than that at low noise level (55 dBA). At 55 dB of noise, the residual noise under the HPD is too low to affect the Lombard effect and that the occlusion effect dominates so there is no effect for pitch, speech level and spectra across the HPD ear conditions.

6.5 Conclusion

The level-dependent HPD has been designed to protect hearing against high noise levels while allowing the wearer to hear low level sounds in surrounding environment. This is accomplished by built-in electronics providing amplification for the external low level incident sounds (typically below 60-70 dBA) while attenuating higher level incident sounds. Although the level-dependent gain setting can maintain the users' auditory perception and situational awareness, the inevitable impacts of transmission gain and the compression circuitry of level-dependent HPD on speech production cannot be ignored. This chapter investigated the effects of a level-dependent HPD on the talkers' speech produced in two

noise types (continuous and fluctuating) at three different levels ranged from low to high (55 dBA to 85 dBA) in an external sound field.

In quiet, the effect of HPD on speech level and pitch was non-significant for males and only showed a small significant effect for females involving the higher surround gain setting. In noise, there was a consistently lower speech level when wearing the HPD than in open ears, except in low noise level of 55 dBA where the difference was not significant. At high noise levels (70 dBA and 85 dBA), the smallest decrease in speech level was related to the high surround setting on HPD while the largest decrease was observed at the low surround setting of HPD and at the passive mode. The effect of noise type on speech level was more prominent at higher noise levels but decreased at the low surround settings providing larger attenuation. The adverse effect of the level-dependent HPD on speech communication depending on the surround setting is found to be less than the conventional passive protector and so the information exchange can be less impeded.

The limitations of this study include: (1) there was no real interaction between the talker and the listener simulated by an acoustic manikin which could affect vocal adjustments by the talker (2) the participants were naive users of hearing protectors and had no experience of adjusting the voice output when wearing the HPDs (3) there were no other comparable studies in the literature evaluating the effects of wearing level-dependent HPD on speech production in noise (4) the effect of wearing a level-dependent HPD by a real target listener was not taken into account. In Chapter 5, it was verified that use of hearing protection by the target listener also affects the talker's speech level [15, Chapter 5]. Talkers increased vocal output when they had knowledge that the listener wore a hearing protector. Whether or not the talkers themselves wore passive hearing protectors, an increase of 3−5 dB in the talkers' speech levels was observed when the target listener wore the passive hearing protector

[Chapter 5]. Thus, there are further avenues which could be explored for future work on the topic. This study contributes to expand the knowledge about the changes in the way speech is produced in noise when the talkers are wearing a level-dependent HPD and about the effect of different surround settings of HPD on the changes in speech characteristics.

Chapter 7

General Discussion and Conclusion

This doctoral thesis aimed to evaluate the effects of wearing passive conventional and active level-dependent hearing protection devices on speech production in quiet and in different noise conditions of an external sound field. This chapter summarizes the content of the chapters including the assessment of different noise suppression methods, the recovery of speech signal from noisy vocal output in an external sound field, and the evaluation of a passive and a level-dependent HPD on speech produced by real participants under different types and levels of noises. This is followed by limitations and future research.

7.1 Summary

7.1.1 Study context

Exposure to high level occupational noises is one of the risk factors which adversely affect job performance, job satisfaction [4], well-being and safety of workers in many industrial and military workplaces. Further, noise-induced hearing loss (NIHL) is one of the most serious effects of exposure to high noise level [2]. An often used approach to protect the workers' hearing against the damaging effects of high noise levels is to use personal hearing protection devices (HPD) where engineering and administrative strategies may not be practical or possible to implement [1, 6] to reduce the noise hazards and exposure [8]. However, hearing protectors have inevitable effects on speech communication, not only on the auditory perception of speech and other desirable sounds but also on the way speech is produced. As a result, it has been found that wearing conventional hearing protectors makes communication more difficult [15] and impairs the perception of workplace sounds [16-19]. Conventional hearing protectors attenuate equally the level of unwanted sounds (noise) and wanted sounds such as speech and warning signals [22, 24, 31, 32, 37, 49, 119]. The adverse effect of conventional passive protectors' fixed attenuation on speech perception can be more challenging for workers who have pre-existing hearing loss [31, 40]. In contrast, level-dependent hearing protection devices (HPD) have been designed to attenuate high level noise and intense sounds while amplifying speech and other desirable low level sounds [24]. Thus, they are able to simultaneously protect hearing against noise and to maintain situational awareness (e.g. warning signal perception, sound localization, speech communication, and detection of distant events) [20, 21, 22, 49]. In spite of these

advantages, level-dependent protectors may have inevitable effects on speech production since they affect the user's perception of their own voice and surrounding noise.

The available studies in the literature have mostly investigated the effect of wearing conventional and level-dependent protectors on speech and auditory perception, but only a few focused on the talker's speech produced while wearing conventional passive protectors. Further, it has been found that there is a lack of knowledge on the impacts of wearing level-dependent protectors on the acoustical characteristics of speech produced by users in noisy environments. A limitation of most available studies on speech production is the methodology used to generate noisy environments. They collect the talkers' speech data in noise injected directly in the ears through headphones. This technique makes it difficult to evaluate the real impact of hearing protectors on speech production and impractical to investigate the interaction of hearing protectors, especially for level-dependent type working in nonlinear modes, or hearing aids with the speech produced in an external noise field.

In summary, the three main gaps in the literature include, 1) extracting the clean Lombard speech induced by an external noise field, rather than the conventional approach of in-ear noise exposure, to take into account the realistic effects of noise on vocal output and to facilitate the study of speech under condition of using hearing protectors and/or hearing aids, 2) evaluating the effects of wearing two types of conventional passive and advanced level-dependent hearing protection devices on speech characteristics produced in quiet and noise, 3) investigating the effects of different noise types (continuous and fluctuating) and levels on the talkers' speech under different protected and unprotected ear conditions.

7.1.2 Thesis contributions

This thesis contributed knowledge to account for the three main gaps in the literature regarding the effect of passive conventional and active level-dependent hearing protection devices on speech production in different conditions of noise types and noise levels by the works accomplished and provided in the four chapters of 3, 4, 5, and 6.

In Chapter 3 and 4, the two noise suppression methods of DWS and ANC were selected among several methods to recover the speech signal in an external noisy filed of a simulation room since they met the defined desirable criteria: applicable in a wider range of SNRs, not sensitive to nonlinear distortion by loudspeakers, insensitive to noise playback and audio recording settings, less sensitive to dynamic perturbation arising from talkers' motion in acoustics of recording room. The better the recovery of speech can be achieved when there is the least movement in the talkers' body during the recording of noises. At the very low SNRs, the ANC method could suppress more noise than that by DWS method. At the more favorable SNRs, although the performance of the ANC method gradually deteriorates due to the greater prominence of speech hindering the ability of the ANC filter, the speech spectra recovered are as good as those with the DWS method. The performance of the two noise suppression methods is also dependent on the type of noise. Greater noise reduction is achieved in the fluctuating and low-frequency noises, while the least reduction is achieved in the continuous and high-frequency noises. The recovery of speech in the external noise filed by the two noise suppression methods is also affected by the acquisition configurations used for recording the speech and noise signals. The omnidirectional microphone relative to the cardioid directional microphone at the shorter distance of 25 cm rather than 50 cm is the preferred acquisition configuration to provide good noise reduction for all noise types.

Overall, the pitch and energy values extracted from the enhanced speech verify that the two DWS and ANC methods are able to adequately suppress the noise in the vocal output recorded with different noise types at SNRs as low as −10 dB, suitable to successfully recover Lombard speech produced in an external noise field with open ears and when wearing hearing protectors.

In Chapter 5, it was observed that overall talkers' speech level is slightly higher when wearing the passive conventional HPD in quiet. In contrast, the same HPD consistently leads to a decrease (2.5−9.5 dB) in the talkers' speech level in noise. The speech energy, pitch, and spectrum also change consistent with the change in vocal effort when wearing a passive hearing protector compared to open ear condition. Depending on the level of noise, the damping effect of passive HPD on speech level is greater at the higher noise levels. The talkers speak louder in the continous noises compared to fluctuating ones. At the low noise level of 70 dBA, the talkers do not react appreciably to the differences in the noise type when wearing the passive HPD so that the speech levels in the different types of noises are very close to each other (within around 1 dB). In contrast, the effect of noise type on the speech level is much larger when the talker is wearing the HPD at the higher noise level of 85 dBA (by up to 3.2 dB) or when in open ears at both noise levels (by up to 4.4 dB). The effect of wearing a passive HPD by the listener in noisy conditions consistently led to an increase by 1−5 dB in the talkers' speech levels. On the other hand, when both the talker and listener wear HPDs compared to both open ears, then there is an interplay between the effects of the talker and listener HPDs. At the low noise level this interplay varies depending on the type of noise causing a raise or a drop, while at the high noise level there is always a drop in speech level. Thus, as the level of noise increases information exchange can be significantly impeded when both individuals wear hearing protection compared to open ears.

According to Chapter 6, the significant effect of level-dependent hearing protector on speech level in quiet depends on the chosen surround setting of the protector. The speech level slightly decreases (about 1.2 dB) when the level-dependent HPD is worn at the high surround setting S4 while it slightly increases at the low surround setting S1 and passive mode, compared to open ear condition. When wearing the level-dependent HPD in noise, the talkers' speech level reduced as much as 10 dB compared to open ear condition. The size of decrease depends on the surround setting chosen on the HPD, smallest decrease at S4 and the largest decrease at the passive "Off" mode. Thus, the level-dependent HPD at higher surround setting can presumably better maintain speech intelligibility in noise while simultaneously protecting hearing against high noise levels. The effect of wearing the HPD and the effect of surround setting on speech level become more prominent as the level of noise increases. The effect of noise type on speech level, leading to a lower level in fluctuating relative to continous noises, reduces when wearing the HPD compared to the open ears. The smallest difference between speech levels across the two noise types is observed at the passive mode of the HPD where the talkers do not react appreciably to the noise type. The effect of noise type is more prominent at the high noise levels of 70 dBA and 85 dBA than that at the low noise level of 55 dBA where the talkers react mostly to the HPD's occlusion effect not the noise. Overall, depending on the selected surround setting, the level-dependent HPD has an adverse effect on speech communication compared to open ears, but less than the conventional passive protector.

7.2 Limitations and future works

This research could achieve the aims and make several contributions to the field of speech production and hearing protection devices, but it has some limitations which should be addressed in future works.

In Chapter 3, some noise suppression methods known in the field of speech (e.g. Wavelet, spectral filtering) were chosen to be evaluated to extract speech from noisy vocal output recorded in the external noise field. There might be many other noise reduction techniques available which might have met the desirable criteria (e.g. largest size of noise reduction over a wide range of SNRs with less amount of artifact and distortion) defined for the purpose of research. Another limitation is that to use the selected noise suppression methods the noise signal needs to be sent twice in the room, one for recording the reference noise and one as target noise when the talker is speaking, which is possible only in some experimental situations where the noise can be exactly reproduced.

In Chapter 4, the two noise suppression methods selected through several tests and in-laboratory noise measurements from a set of available methods were validated to investigate their performance in the presence of talkers' body movements. A limitation of this study was that to simulate noisy speech the noise and speech signals recorded separately were mixed at different SNRs. The noise signals were recorded while there was a real participant with/without dynamic body movements in the room but speech was recorded as it was produced through the mouth of the manikin at the reference position in the room. This was done because it was necessary to have an identical baseline “clean” speech signal in all conditions with different acquisition set-ups in order to be able to directly compare the recovered speech following noise suppression. Therefore, the speech material could not be

recorded while it was read by a real participant in the noisy room that would result in no single comparator. It is expected that this recording methodology would not have any significant effect on the results, since the focus was on the size of noise reduced by two methods, which is not affected by the way speech is recorded.

In Chapters 5 and 6, two distinct experiments were carried out to investigate the effects of wearing a passive and an active level-dependent hearing protectors on speech produced by 24 participants in different noise conditions. The first limitation of these studies was that an acoustic manikin was used as target listener so that there was no real interaction between the talker and listener which might have resulted in dynamic vocal adjustment by the talker. Furthermore, speech material used in the studies was a list of sentences which the talkers read in such a manner to be understood by the target listener (manikin) while spontaneous speaking about a topic or a picture might be more representative of a real conversation. Thus, it remains possible for future research to evaluate speech signals recorded where two real conversation partners engage in a more challenging communicative settings. Another experimental condition limiting the studies (Chapters 5 and 6) was that the participants were naive users of hearing protectors while workers with experience of hearing protection use may have developed strategies to speak louder to compensate for wearing an HPD in noise. One of the criteria to recruit participants for the two experiments was that they had normal hearing thresholds, while future research can take into account the effects of hearing status of talker or listener or both. It can extend models developed for use of hearing protection devices by talkers with a wider range of hearing profiles. In Chapter 5, an earmuff was chosen as a passive hearing protector while an earplug worn either by the talker or the target listener might affect speech production differently. In the last study (Chapter 6), the talkers' speech was evaluated to reveal the effects of using a level-dependent hearing protector while

the target listener was with open ears. The ear condition of listener as a factor affecting talkers' speech can be investigated further by future research. The results provided in the last study (Chapter 6) could not be compared to other studies because the available research on the level-dependent HPDs is limited to the studies focussing on the listener's end and not on speech production in noise. To this end, more research conducted in this field is necessary and could help obtain more knowledge of speech production while wearing different types of level-dependent hearing protectors as well as earplugs and earmuffs.

7.3 Outlook

In conclusion, the main goal of the research presented in this thesis was to characterize speech production under the condition of wearing conventional passive and advanced level-dependent hearing protection devices in different types and levels of noises of an external sound field. The results of experimental measurements expand our understanding of the advantages of level-dependent protectors over passive protectors in terms of their effects on speech production An article published in the *Proceedings of Meetings on Acoustics* and an article presented in *National Hearing Conservation Association Conference (NHCA)* has arisen from this work, a full journal paper is under review, and some more publications will be submitted in the near future. Finally, the present research could contribute key data towards the development of comprehensive tools or quantitative models to assess the impact of different noise types, noise levels, and hearing protection devices on face-to-face speech communication in real noisy environments.

References

[1] Tak, S., Davis, R. and Calvert, G., Exposure to hazardous workplace noise and use of hearing protection devices among US workers-NHANES, 1999-2004, *American Journal of Industrial Medicine,* 52: 358−371, 2009.

[2] Nelson, D. I., Nelson, R. Y., Concha-Barrientos, M. and Fingerhut, M., The global burden of occupational noise-induced hearing loss, *American Journal of Industrial Medicine,* 48: 446−458, 2005.

[3] Parent-Thirion, A., Fernandes Macias, E., Hurley, J. and Vermeylen, G., Fourth European Working Conditions Survey. Dublin: European Foundation for the Improvement of Living and Working Conditions, Catalogue no: TJ-76-06-497-EN-C, ISBN 92-897-0974-X, 139p. Available online at http://www.eurofound.europa.eu/publications/htmlfiles/ef0698.htm]

[4] Leather, P., Beale, D. and Sullivan, L., Noise, psychosocial stress and their interaction in the workplace, *Journal of Environmental Psychology,* 23: 213−222, 2003.

[5] Morata, T. C., Themann, C. L., Randolp, R. F. and Verbsky, B. L., Working in noise with a hearing loss perceptions from workers, supervisors, and hearing conservation program managers, *Ear & Hearing,* 26(6): 529−545, 2005.

[6] Themann, C., Suter, A. and Stephenson, M., National research agenda for the prevention of occupational hearing loss, Part 1 and Part 2, *Seminar in Hearing,* 34(3):145−251, 2013.

[7] Abel, S. M., Barriers to hearing conservation programs in combat arms occupations, *Aviation, Space, and Environmental Medicine,* 79(6): 591−598, 2008.

[8] Suter, A. H., Engineering controls for occupational noise exposure, *Sound and Vibration,* 24−31, 2012.

[9] Gerges, S. and Casali, J., Hearing protectors. In: M.J. Crooker (ed.) *Handbook of Noise and Vibration Control,* New York: Wiley, 2007.

[10] Canetto, P., Hearing protectors: Topicality and research needs, *International Journal of Occupational Safety and Ergonomics,* 15: 141−153, 2009.

[11] EN 458: 2004. *Hearing protectors-recommendations for selection, use, care and maintenance - Guidance document*. Brussels: European Committee for Standardization

[12] ANSI/ASA S12.68 458, *American National Standard, Methods of estimating effective A-weighted sound pressure levels when hearing protectors are worn*, Acoustical Society of America, Melville, NY, 2007.

[13] ANSI/ASA S12.42 (2010). *American National Standard Methods for the measurement of insertion loss of hearing protection devices in continuous or impulsive noise using microphone-in-real-ear or acoustic test fixture procedures*. Melville NY: Acoustical Society of America.

[14] CAN/CSA Z94.2-02 R2011. *Hearing protection devices-performance, selection, care, and use*. Toronto: Canadian Standards Association.

[15] Hormann, H., Lazarus-Mainka, G., Schubeius, M. and Lazarus, H., The effects of noise and the wearing of ear protectors on verbal communication, *Noise Control Engineering Journal*, 23(2): 69−77, 1984.

[16] Giguère, C., Laroche, C., Vaillancourt, V. and D.Soli, S., Modelling speech intelligibility in the noisy workplace for normal-hearing and hearing-impaired listeners using hearing protectors, *International Journal of Acoustics and Vibration,* 15(4): 156−167, 2010.

[17] Abel, S. M., Tsang, S. and Boyne, S., Sound localization with communication headsets: comparison of passive and active systems, *Noise & Health,* 9: 101−107, 2007.

[18] Costa, N., Abreu, S. and Arezes, P. M., Alarm detection in noise work environments: the influence of hearing protection devices, in *Proceedings of the European Safety and Reliability Conference, ESREL*, London, 2013.

[19] Laroche, C., Giguère, C., Vaillancourt, V., Roy, K., Pageot, L., Nèlisse, H., Ellaham, N., and Nassrallah, F., Detection and reaction thresholds for reverse alarms in noise with and without passive hearing protection, *International Journal of Audiology*, 57:S51−S60, 2018.

[20] Giguère, C., Laroche, C., Brammer, A. J., Vaillancourt, V. and Yu, G., Advanced Hearing Protection and Communication:Progress and Challenges, in *10th International Congress on Noise as a Public Health Problem (ICBEN)*, London, 2011.

[21] Giguère, C., Laroche, C., Vaillancourt, V., Shmigol, E., Vaillancourt, T., Chiasson, J. and Rozen-Gauthier, V., A multidimensional evaluation of auditory performance in one powered electronic level-dependent hearing protector, in *19th International*

Congress on Sound and Vibration, Vilnius, 2012.

[22] Tufts, J. B., Hamilton, M. A., Ucci, A. J. and Rubas, J., Evaluation By Industrial Workers of Passive and Level-dependent Hearing Protection Devices, *Noise & Health*, 13(50): 26−36, January-February 2011.

[23] Giguère, C., Laroche, C. and Vaillancourt, V., The interaction of hearing loss and level-dependent hearing protection on speech recognition in noise, *International Journal of Audiology*, 54: S9−S18, 2015.

[24] Giguère, C., Laroche, C. and Vaillancourt, V., Advanced Hearing Protection and Auditory Awareness in Individuals with Hearing Loss, in *Proceedings of Meeting on Acoustics*, Montreal, 2013.

[25] Kryter, K., Effects of ear protective devices on the intelligibility of speech in noise, *The Journal of the Acoustical Society of America*, 18(2): 413−417, 1946.

[26] Pollack, I., Speech communication at high noise level: The roles of a noise-operated automatic gain control system and hearing protection, *The Journal of the Acoustical Society of America*, 29(12): 1324−1327, 1957.

[27] Howell, K. and Martin, A., An Investigation of the Effects of Hearing Protectors on Vocal Communication in Noise, *Sound and Vibration*, 41(2): 181−196, 1975.

[28] Abel, S. M., Alberti, P. W., Haythornthwaite, C. and Riko, K., Speech intelligibility in noise: effects of fluency and hearing protector type, *The Journal of the Acoustical Society of America*, 71(3): 708−715, 1982.

[29] Pekkarinen, E., Viljanen, V., Salmivalli, A. and Suonpaa, J., Speech recognition in a noisy and reverberant environment with and without earmuffs, *Audiology*, 29: 286−293, 1990.

[30] Wilde, G. and Humes, L. E., Application of the Articulation Index to the Speech Recognition of Normal and Impaired Listeners Wearing Hearing Protection, *Acoustical Society of America*, 87(3):1192−1199, March 1990.

[31] Abel, S. M., Armstrong, N. M. and Giguère, C., Auditory perception with level-dependent hearing protectors. The effects of age and hearing loss, *Scandinavian Audiology*, 22(2): 71−85, 1993.

[32] Simpson, B., Bolia, R., McKinley, R. and Brungart, D., The impact of hearing protection on sound localization and orienting behavior," *HUMAN FACTORS*,

47(1):1−11, 2005.

[33] Zimpfer, V., Sarafian, D., Buck, K. and Hamery, P., Spatial localization of sounds with hearing protection devices allowing speech communication, in *Proceedings Conference of the Acoustics*, Nantes, 2012.

[34] Brown, A., Beemer, B., Greene, N. T., Argo, T., IV, Meegan, D. and Tollin, D. J., Effects of active and passive hearing protection devices on sound source localization, speech recognition, and tone detection, *PLoS ONE,* 10(8): 1−21, 2015.

[35] Brungart, D. S., Hobbs, B. W. and Hamil, J. T., A comparison of acoustic and psychoacoustic measurements of pass-through hearing protection devices, in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, 2007.

[36] Zimpfer, V. and Sarafian, D., Sound-localization Performance with the Hearing Protectors, in *Proceedings of Meeting on Acoustics*, Montreal, 2013.

[37] Hiselius, P., Edvall, N. and Reimers, D., To measure the impact of hearing protectors on the perception of speech in noise, *International Journal of Audiology,* 54: S3−S8, 2015.

[38] Casali, J., Ahooran, W. and J.A., L., A field investigation of hearing protection and hearing enhancement in one device: for soldiers whose ears and lives depend upon it, *Noise Health*, 11(42): 69−90, 2009.

[39] Navarro, R., Effects of Ear Canal Occlusion and Masking on the Perception of Voice, *Perceptual and Motor Skills,* 82: 199−208, 1996.

[40] Tufts, J. B. and Frank, T., Speech Production in Noise with and without Hearing Protection, *The Journal of the Acoustical Society of America*, 114(2): 1069−1080, August 2003.

[41] Bouserhal, R., Macdonald, E. N., Falk, T. H. and Voix, J., Variations in voice level and fundamental frequency with changing background noise level and talker-to-listener distance while wearing hearing protectors: A pilot study, *International Journal of Audiology*, 55(1): S13−S20, 2016.

[42] Kent, R., *The speech sciences*, London: Singular Publishing Group, 1997.

[43] Lindblom, B., Explaining Phonetic Variation: A Sketch of the H&H Theory, In: Hardcastle W.J., Marchal A. (eds), *Speech Production and Speech Modeling*,

NATO ASI Series (Series D: Behavioural and Social Sciences), vol. 55, Springer, Dordrecht, 403−439, 1990.

[44] Tourville, J. A., Reilly, K. J. and Guenther, F. H., Neural Mechanisms Underlying Auditory Feedback Control of Speech, *Neuroimage*, 39(3): 1429−43, Feb. 2008.

[45] Raphael, L. J., Borden, G. J. and Harris, K. S., *Speech science primer: physiology, acoustics, and perception of speech*, Fifth ed., Lippincott Williams & Wilkins, 2007.

[46] Hardcastle, W., Marchal, A., *Speech production and speech modeling*. Springer, Dordrecht, Dordrecht, 1990.

[47] Giguère, C., Understanding Hearing loss: Implications for Speech Perception, Chapter 2 in E.M. Fitzpatrick and S.P. Doucet: *Audiologic Rehabilitation for Children with Hearing Loss: From Infancy to Adolescence*, Thieme Medical Publishers, New York, 18−28, 2013.

[48] Canadian Centre for Occupational Health and Safety, Noise, Available online: https://www.ccohs.ca/topics/hazards/physical/noise/.

[49] ANSI/ASA S3.5. American National Standard Methods for calculation of the speech intelligibility index. Acoustical Society of America, 1997.

[50] Yost, W. A., *Fundamentals of Hearing: An Introduction*, Fifth edition, San Diego: Academic Press Elsevier, 2007.

[51] Cook, M., Mayo, C., Valentini-Botinhao, Stylianou, C., Y., Sauert, B. and Tang, Y., Evaluating the intelligibility benefit of speech modifications in known noise conditions, *Speech Communication*, 55: 572−585, 2013.

[52] Lindblom, B., Explaining Phonetic Variation: A Sketch of the H&H Theory, In: Hardcastle W.J., Marchal A. (eds), *Speech Production and Speech Modeling*. NATO ASI Series (Series D: Behavioural and Social Sciences), vol. 55, Springer, Dordrecht, 403−439, 1990.

[53] Kent, R., *The speech sciences*, London: Singular Publishing Group, 1997.

[54] O'Shaughnessy, D., *Speech Communications:Human and Machine*, New York: IEEE Press, 2000.

[55] Ephraim, Y. and Cohen, I., *Recent Advancement in Speech Enhancement*, CRC Press, Boca Raton, 2004.

[56] Zeine, L. and Brandt, J. F., The Lombard Effect on Alaryngeal Speech, *Communication Disorder,* 21: 373−383, 1988.

[57] Lombard, E., Le signe de l'elevation de la voix, *Ann. Malad. Oreille, Larynx, Nez, Pharynx,* 37: 101–119, 1911.

[58] Bapineedu, G., Analysis of Lombard effect speech and its application in speaker verification for imposter detection, MS thesis, Language Technologies Research Centre International Institute of Information Technology, Hyderabad, India, 2010.

[59] Junqua, J., Finckle, S. and Field, K., The Lombard effect: A reflex to better communicate with others in noise, *IEEE Proceedings International Conference on Acoustics, Speech, and Signal Processing*, 1999.

[60] Garnier, M., Henrich, N. and Dubois, D., Influence of sound immersion and communicative interaction on the Lombard effect, *Journal of Speech, Language, and Hearing Research,* 53: 588−608, 2010.

[61] Zeine, L. and Brandt, J. F., This Lombard Effect On Alaryngeal Speech, *Communication Disorder,* 21: 373−383, 1988.

[62] Zollinger, S. A. and Brumm, H., The Lombard Effect, *Current Biology,* 21(16): 614−615, 2011.

[63] Junqua, J.-C., The Lombard Reflex and its Role on Human Listeners and Automatic Speech Recognizers, *The Journal of the Acoustical Society of America,* 93(1): 510−524, 1993.

[64] Lau, P., The Lombard Effect as a Communicative Phenomenon, *UC Berkeley Phonology Lab Annual Report*, Berkeley, 2008.

[65] Stowe, L. M. and Golob, E. J., Evidence that the Lombard effect is frequency-specific in humans, *The Journal of the Acoustical Society of America,* 134(1): 640−647, 2013.

[66] Hodgson, M., Steininger, G. and Razavi, Z., Measurement and Prediction of Speech and Noise Levels and The Lombard Effect in Eating Establishment, *The Journal of* the *Acoustical Society of America*, 121(4): 2023−2033, April 2007.

[67] Rindel, J. H., Acoustical capacity as a means of noise control in eating establishments, *Baltic-Nordic Acoustics Meeting*, Odense, Denmark, 2012.

[68] Rindel, J. H. and Gade, A. C., Dynamic sound source for simulating the Lombard effect modeling software, *Internoise*, New York, August 2012.

[69] Nijs, L., Saher, K. and Ouden, D. d., Effect of Room Absorption on Human Vocal Output in Multitalker Situations," *The Journal of the Acoustical Society of America*, 123(2): 803−813, February 2008.

[70] Lane, H. and Tranel, B., The Lombard sign and the role of hearing in speech, *Journal of Speech and Hearing Research*, 14(4): 677−709, 1971.

[71] Korn, T. S., Effect of psychological feedback on conversational noise reduction in rooms, *The Journal of the Acoustical Society of America*, 26(5): 793−794, 1954.

[72] Pickett, J. M., Limits of direct speech communication in noise, *The Journal of the Acoustical Society of America*, 30(4): 278−281, 1958.

[73] Webster, J. C. and Klumpp, R. G., Effects of ambient noise and nearby talkers on a face-to-face communication task, *The Journal of the Acoustical Society of America*, 34(7): 936−941, 1962.

[74] Whitlock, J. and Dodd, G., Classroom Acoustics-Controlling the Cafe Effect ... is the Lombard Effect the Key?, in *Proceedings of ACOUSTICS*, Christchurch, 2006.

[75] Sato, H. and Bradley, J. S., Evaluation of acoustical conditions for speech communication in working elementary school classrooms, *The Journal of the Acoustical Society of America*, 123(4): 2064−2077, 2008.

[76] Drugman, T. and T. Dutoit, Glottal-based Analysis of The Lombard Effect, in *11th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, Chiba, 2010.

[77] Garnier, M., Bailly, L., Dohen, M., Welby, P. and Loevenbruck, H., An Acoustic and Articulatory Study of Lombard Speech:Global Effect on the Utterance, in *9th International Conference on Spoken Language Processing (INTERSPEECH)*, Pittsburgh, 2006a.

[78] Davis, C., Kim, J., Grauwinkel, K. and Mixdorff, H., Lombard Speech:Auditory (A), Visual (V) and AV Effects, *3rd International Conference on Speech Prosody*, Dresden, 2006.

[79] Garnier, M. and Henrich, N., Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise?, *Computer Speech*

and Language, 28: 580−597, 2014.

[80] Junqua, J. C., The Influence of Acoustics on Speech Production: A Noise-induced Stree Phenomenon Known as the Lombard Effect, *Speech Communication,* 20: 13−22, 1996.

[81] Boril, H. and Hansen, J. H., Unsupervised Equalization of Lombard Effect for Speech Recognition in Noisy Adverse Environments, *IEEE Transaction on Audio,Speech, and Language Processing,* 18(6): 1379−1393, August 2010.

[82] Ogawa, T. and Kobayashi, T., Influence of Lombard effect: accuracy analysis of simulation-based assessments of noisy speech recognition systems for various recognition conditions, *IEICE Transactions on Information and Systems,* E.92-D(11): 2244−2252, 2009.

[83] Junqua, J. C. and Anglade, Y., Acoustic and Perceptual Studies of Lombard Speech: Application to Isolated Words Automatic Speech Recognition, *International Conference on Acoustics, Speech, and Signal Processing*, Albuquerque, 1990.

[84] Hansen, J. H. and Varadarajan, V., Analysis and Compensation of Lombard Speech Across Noise Type and Levels With Application to In-Set/Out-of-Set Speaker Recognition, *IEEE Transaction on Audio, Speech, and Language Processing,* 17(2): 366−378, February 2009.

[85] Lu, Y. and Cooke, M., Speech Production Modifications Produced by Competing Talkers, Babble, and Stationary Noise, *The Journal of the Acoustical Society of America,* 124(5): 3261−3275, November 2008.

[86] Hansen, J. H. and Cairns, D. A., ICARUS:Source Generator Based Real-time Recognition of Speech in Noisy Stressful and Lombard Effect Environments, *Speech Communication,* 16: 391−422, 1995.

[87] Giguère, C., Laroche, C., Brault, E., Ste-Marie, J. C., Brosseau-Villeneuve, Philippon, M., B. and Vaillancourt, V., Quantifying the Lombard Effect in Different Background Noises, *The Journal of the Acoustical Society of America,* 3378−3378, 2006.

[88] Huang, D.Y., Rahardja, S. and Ong, E. P., Lombard Effect Mimicking, in *International Conference on Spoken Language Processing (INTERSPEECH)*, Makuhari, 2010.

[89] Ho, A. K., Bradshaw, J. L., Iansek, R. and Alfredson, R., Speech Volume Regulation in Parkinson's Disease: Effects of Implicit Cues and Explicit Instructions, *Neuropsychologia,* 37: 1453−1460, 1999.

[90] Huang, D. Y., Rahardja, S. and Ong, E. P., Biologically Inspired Algorithm for Enhancement of Speech Intelligibility Over Telephone Channel, *International Workshop on Multimedia Signal Processing* , Rio De Janeiro, 2009.

[91] Goldenberg, R., Cohen, A. and Shallom, I., The Lombard Effect's Influence on Automatic Speaker Verification Systems and Methods for its Compensation, *International Conference on Information Technology: Rresearch and Education*, Telaviv, 2006.

[92] Skowronski, M. D. and Harris, J. G., Applied Principles of Clear and Lombard Speech for Automated Intelligibility Enhancement in Noisy Environment, *Speech Communication,* 48: 549−558, 2006.

[93] Mokbel, C. E. and Chollet, G. F. A., Automatic Word Recognition in Cars, *IEEE Transaction on Speech and Audio Processing,* 3(5): 346−356, September 1995.

[94] Levins, J., A Holistic Approach to Room Design for Hearing Impaired Populations, *Proceedings of Meeting on Acoustics*, Montreal, 2013.

[95] Wong, L. L. N., Ng, E. H. N. and Soli, S. D., Characterization of speech underestanding in various types of noise, *The Journal of the Acoustical Society of America,* 132(4): 2642−2651, 2012.

[96] Bacon, S. P., Opie, J. M. and Montoya, D. Y., The Effects of Hearing Loss and Noise Masking on the Masking Release for Speech in Temporally Complex Backgrounds, *Speech, Language, and Hearing Research,* 41: 549−563, June 1998.

[97] Laroche, C., Roveda, J. G., Levionnois, J., Giguere, C. and Vaillancourt, V., Binaural Speech Recognition in Continuous and Intermittent Noises in People with Hearing Loss, *165th Meeting of the Acoustical Society of America,* Montreal, 2013.

[98] Nie. Y., Svec, A., Nelson, P. and Munson, B., The Effects of Temporal Envelope Confusion on Listeners' Phoneme and Word Recognition, *Proceedings of Meeting on Acoustics*, Montreal, 2013.

[99] Oxenham, A. J. and Simonson, A. M., Masking Release for Low- and High-pass-filtered Speech in the Presence of Noise and Single-talker Interference, *The Journal*

of the Acoustical Society of America, 125(1): 457−468, January 2009.

[100] Crandell, C. C. and Smaldino, J. J., Classroom Acoustics for Children With Normal Hearing and Hearing Impairment, *Language, Speech, and Hearing Services in Schools,* 31: 362−370, October 2000.

[101] MacDonald, E. N. and Raufer, S., Speech Perception in Amplitude-modulated Noise, *Proceedings of Meeting on Acoustics*, Montreal, 2013.

[102] Velez, A. and Bee, M. A., Dip Listening and the Cocktail Party Problem in Grey Treefrogs: Signal Recognition in Temporally Fluctuating noise, *Animal Behaviour,* 82: 1319−1327, 2011.

[103] Phatak, S. A. and Grant, K. W., Phoneme Recognition in Modulated Maskers by Normal-hearing and Aided Hearing-impaired Listeners, *The Journal of the Acoustical Society of America,* 132(3): 1646−1654, September 2012.

[104] Moore, B. C. J., Chapter8: Speech Perception, *Cochlear Hearing Loss:Physiological,Psychological and Technical Issues*, John Wiley & Sons Ltd, 201−232, 2007.

[105] Castellanos, A., Benedi, J. M. and Casacuberta. F., An Analysis of General Acoustic-phonetic Features for Spanish Speech Produced with Lombard Effect, *Speech Communication,* 20: 23−35, 1996.

[106] Wakao, A., Takeda, K. and Itakura, F., Variability of Lombard effects under different noise, *Proceedings of the International Conference on Spoken Language Processing (ICSLP),* 1996.

[107] Ferrand, C. T., Relationship Between Masking Levels and Phonatory Stability in Normal-speaking Women, *Voice,* 20(2): 223−228, 2005.

[108] Pichora-Fuller, M. K., Goy, H. and Lieshout, P. v., Effect on Speech Intelligibility of Changes in Speech Production Influenced by Instructions and Communication Environments, *Seminars in Hearing*, 31(2): 77−94, 2010.

[109] Howard-Jones, P. and Rosen, S., The Perception of Speech in Fluctuating Noise, *ACUSTICA,* 78: 258−272, 1993.

[110] Nelson, P. B., Jin, S. H., Carney, A. E. and Nelson, D. A., Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners, *The Journal of the Acoustical Society of America,* 113(2): 961−968, 2003.

[111] Rhebergen, K. S., Versfeld, N. J. and Dreschler, W. A., Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise, *The Journal of the Acoustical Society of America,* 120(6): 3988−3997, 2006.

[112] George, E. L. J., Festen, J. M. and Goverts, S. T., Effects of reverberation and masker fluctuations on binaural unmasking of speech, *The Journal of the Acoustical Society of America,* 132(3): 1581−1591, 2012.

[113] Wagoner, L., McGlothlin, J., Chung, K., Strickland, E., Zimmerman, N. and Carlson, G., Evaluation of noise attenuation and verbal communication capabilities using three ear insert hearing protection systems among airport maintenance personnel, *Journal of Occupational and Environmental Hygiene,* 4: 114−122, 2007.

[114] Serhal, R. E. B., Falk, T. H. and Voix, J., Integration of a Distance Sensitive Wireless Communication Protocol to Hearing Protectors Equipped with in-ear Microphones, *Proceedings of Meeting on Acoustics*, Montreal, 2013.

[115] Berger, E. H. and Kerivan, J. E., Influence of physiological noise and the occlusion effect on the measurement of real-ear attenuation at threshold, *The Journal of the Acoustical Society of America,* 74(1): 81−94, 1983.

[116] Porschmann, C., Influence of bone conduction and air conduction on the sound of one's own voice, *Acta Acustica United with Acustica*, 86(6): 1038−1045, 2000.

[117] Brungart, D., Cord, M. T., Solomon, N. P., Dietrich-Burns, K. and Block, K., Evaluating the effects of hearing protection on speech production in noisy environments, Edinburgh,UK, 2012.

[118] Ternstrom, S., Sodersten, M. and Bohman, M., Cancellation of Simulated Environmental noise as a Tool for Measuring Vocal Performance During Noise Exposure, *Journal of Voice*, 16(2): 195−206, 2002.

[119] Wilde, G. and Humes, L. E., Application of the Articulation Index to the Speech Recognition of Normal and Impaired Listeners Wearing Hearing Protection, *The Journal of the Acoustical Society of America,* 1192−1199, 1990.

[120] Norin, J. A., Emanuel, D. C. and Letowski, T. R., Speech intelligibility and passive, level-dependent earplugs, *Ear & Hearing*, 32(5): 642−649, 2011.

[121] Williams, W., A qualitative assessment of the performance of electronic, level-dependent earmuffs when used on firing ranges, *Noise & Health,* 13(51): 189−194,

2011.

[122] Murphy, W. J., Flamme, G. A., Meinke, D. K., Sondergaard, J., Finan, D. S., Lankford, J. E., Khan, A., Vernon, J. and Stewart, M., Measurement of impulse peak insertion loss for four hearing protection devices in field conditions, *International Journal of Audiology*, 51: S31−S42, 2012.

[123] McKinley, R. L., Thompson, E. R. and Simpson, B. D., Measuring effective detection and localization performance of hearing protection devices, *The Journal of the Acoustical Society of America*, 136(4), 2014.

[124] Scharine, A. A. and Weatherless, R. A., U.S. marine corps level-dependent hearing protector assessment: objective measures of hearing protection devices, *Army Research Laboratory*, 2014.

[125] Nakashima, A., Comparison of different types of hearing protection devices for use during weapons firing, *Journal of Military, Veteran, and Family Health*, 1(2): 43−51, 2015.

[126] Vaziri, G., Giguère, C., Dajani, H. and Ellaham, N., A comparison of speech enhancement methods to extract Lombard speech in an external noise field, *The Journal of the Acoustical Society of America*, 138(3), 2015.

[127] Chen, S. H., Speech enhancement using perceptual wavelet packet decomposition teager energy operator, *Journal of VLSI Signal Processing*, 36: 125−139, 2004.

[128] Abd El-Fattah, M. A., Dessouky, M. I., Abbas, A. M., Alaa, M., Diab, S. M., El-Rabaie, E. M., Al-Nuaimy, W., Alshebeili, S. A. and Abd El-samie, F. E., Speech enhancement with an adaptive filter, *International Journal of Speech Technology*, 17: 53−64, 2014.

[129] Kaladharan, N., Speech enhancement by spectral subtraction, *International Journal of Computer Applications*, 96(13): 45−48, 2014.

[130] Fukane, A. R. and Sahare, S. L., Different approaches of spectral subtraction method for enhancing the speech signal in noisy environments, *International Journal of Scientific & Engineering Research*, 2(5), 2011.

[131] Ayat, S., Manzuri-Shalmani, M. T. and Dianat, R., An improved wavelet-based speech enhancement by using speech signal features, *Computer and Electrical Engineering*, 32: 411−425, 2006.

[132] Bahoura, M. and Rouat, J., Wavelet speech enhancement based on time-scale adaptation, *Speech Communication*, 48: 1620−1637, 2006.

[133] Ramli, R. M., Abid Noor, A. O. and Abdul Samad, S., A Review of Adaptive Line Enhancers for Noise Cancellation, *Australian Journal of Basic and Applied sciences*, 6(6): 337−352, 2012.

[134] Jia, H., Zhang, X. and Bai, J., A continuous differentiable wavelet threshold function for speech enhancement, *Journal of Central South University*, 20: 2219−2225, 2013.

[135] Holube, I., Fredelake, S., Vlaming, M. and Kollmeier, B., Development and analysis of an International Speech Test Signal (ISTS), *International Journal of Audiology*, 49: 891−903, 2010.

[136] Zhao, H. and Gan, W., A new pitch estimation method based on AMDF, *Journal of Multimedia*, 8(5): 618−625, 2013.

[137] Larm, P. and Hongisto, V., Experimental comparison between speech transmission index, rapid speech transmission index, and speech intelligibility index, *The Journal of the Acoustical Society of America*, 119(2): 1106−1117, 2006.

[138] Payton, K. L. and Shrestha, M., Evaluation of short-time speech-based intelligibility metrics, *The 9th International Congress on Noise as a Public Health Problem (ICBEN)*, Foxwoods, 2008.

[139] Payton, K. L. and Shrestha, M., Comparison of a short-time speech-based intelligibility metric to the speech transmission index and intelligibility data, *The Journal of the Acoustical Society of America*, 134(5): 3818−3827, 2013.

[140] Rhebergen, K. S., Versfeld, N. J. and Dreschler, W. A., Extended speech intelligibility index for the prediction of the speech reception threshold in fluctuating noise, *The Journal of the Acoustical Society of America*, 120(6): 3988−3997, 2006.

[141] Garnier, M., Dohen, M., Loevenbruck, H., Welby, P., and Bailly, L. The Lombard Effect:A Physiological Reflex or a Controlled Intelligibility Enhancement, *Proceedings of 7th International Seminar on Speech Production*, 255−262, Ubatuba, 2006b.

[142] Laugesen, S., Nielsen, C., Maas, P., and Jensen, N. S. Observations on hearing aid users' strategies for controlling the level of their own voice, *Journal of American Academy of Audiology*, 20(8): 503−13, 2009.

[143] Garnier, M., and Henrich, N. Speaking in noise; how does the Lombard effect improve acoustic contrasts between speech and ambient noise?, *Computer Speech and Language*, 28: 580−597, 2014.

[144] Luke, C., Theib, A., Schmidt, G., Niebuhr, O., and John, T. Creation of a Lombard speech Database Using an Acoustic ambiance Simulation with Loudspeakers, *6th Biennial Workshop on DSP for In-Vehicle Systems*, Seoul, korea, 1−8, 2013.

[145] Abd El-Fattah, M. A., Dessouky, M. I., Abbas, A. M., Alaa, M., Diab, S. M., El-Rabaie, E. M., Al-Nuaimy, W., Alshebeili, S. A., and Abd El-samie, F. E., Speech enhancement with an adaptive Wiener filter, *International Journal of Speech Technology*, 17: 53−64, 2014.

[146] Nymand, M., Directional vs. omnidirectional microphones. DPA MICROPHONES, https://www.dpamicrophones.com/mic-university/directional-vs-omnidirectional-microphones, 2015.

[147] Thompson, S. C., Directional microphone patterns: they also have disadvantages. Audiology Online: https://www.audiologyonline.com/articles/directional-microphone-patterns-they-also-1294.

[148] Vermiglio, A. J. The American English Hearing in Noise Test, *International Journal of Audiology*, 47: 386−387, 2008.

[149] Nassrallah, F. G., Measurement og occupational sound exposure from communication headsets, Doctorate Disertation, University of Ottawa, 2016.

[150] Gonzalez, S., and Brookes, M. A pitch estimation filter robust to high levels of noise (PEFAC), *IEEE/ACM Transactions on Audio, speech and Language Processing (TASLP)*, 22(2): 518−530, 2011.

[151] Zhao, H., and Gan, W., A new pitch estimation method based on AMDF, *Journal of Multimedia*, 8(5): 618−625, 2013.

[152] Tan, L., and Karnjanadecha, M., Pitch detection algorithm: autocorrelation method and AMDF, *Intelligent Signal Processing and Communication Systems (ISPACS)*, Bangkok, Thailand, 551−556, 2003.

[153] Liu, W. M., Jellyman, K. A., Evans, N. W. D., and Mason, J. S. D., Assessment of objective quality measures for speech intelligibility, *International Conference on Acoustics, Speech Processing (ICASSP)*, Toulouse, France, 1225−1228, 2006.

[154] Gomez, A. M., Schwerin, B., and Paliwal, K., Improving objective intelligibility prediction by combining correlation and coherence based methods with a measure based on the negative distortion ratio, *Speech Communication*, 54: 503−15, 2012.

[155] Goldworthy, R. L., and Greenberg, J. E., Analysis of speech-based speech transmission index methods with implications for nonlinear operations, *The Journal of the Acoustical Society of America*, 116(6): 3679−3689, 2004.

[156] Loizou, P., and Ma, J., Extending the articulation index to account for non-linear distortions introduced by noise-suppression algorithms, *The Journal of the Acoustical Society of America*, 130(2): 986−995, 2011.

[157] Payton, K. L., and Braida, L. D., A method to determine the speech transmission index from speech waveforms, *The Journal of the Acoustical Society of America*, 106(6): 3637−3648, 1999.

[158] Beerends, J. G., Helstra, A. P., Rix, A. W. and Hollier, M. P., Perceptual Evaluation of Speech Quality (PESQ), the new ITU standard for end-to-end speech quality assessment. Part II-Psychoacoustic model, *Journal of Audio Engineering Society*, 50(10): 765−778, 2002.

[159] Taal, C. H., Hendriks, R. C., Heusdens, R., Jensen, J., Kjems, U., An evaluation of objective quality measures for speech intelligibility prediction, *Proceeding of Interspeech*, Brighton, UK, 1947−1950, 2009.

[160] Pourmand, N., Objective and subjective evaluation of wideband speech quality, Doctorate Disertation, The University of Western Ontario, London, Ontario, Canada, 2012.

[161] Hu, Y., Loizou, P. C., Evaluation of objective quality measures for speech enhancement, *IEEE Transaction on Audio, Speech, and Language Processing*, 16(1): 229−238, 2008.

[162] Ma, J., Hu, Y., and C., L. P., Objective measures for predicting speech intelligibility in noisy condition based on new band-importance functions, *The Journal of the Acoustical Society of America*, 125(5): 3387−3405, 2009.

[163] Dittberner, A., Interpreting the Directivity Index (DI), *The Hearing Review*, 2003, Available at: http://www.hearingreview.com/2003/06/interpreting-the-directivity-index-di/.

[164] Berger, E.H., The effects of hearing protectors on auditory communications, *Occupational Health Nursing*, 28(1): 6−7, 1980.

[165] Stone, M. A., Fullgrabe, C., Mackinnon, R. C., and Moore, B. C. J., The importance for speech intelligibility of random fluctuations in "steady" background noise, *The Journal of the Acoustical Society of America*, 130(5): 2874−2881, 2011.

[166] Qin, M. K., and Oxenham, A., Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers, *The Journal of the Acoustical Society of America*, 114(1): 446−454, 2003.

[167] Bernstein, J. G. W. and Grant, K. W., Auditory and Auditory-visual Intelligibility of Speech in Fluctuating Maskers for Normal-hearing and Hearing-impaired Listeners, *The Journal of the Acoustical Society of America*, 125(50): 3358−3372, May 2009.

[168] Lindeman, H.E., Speech Intelligibility and the Use of Hearing Protectors, *Audiology*, 15(4): 348−356, 1976.

[169] Kohata S., Inoue J., Sakuma T., Nakagawa T., Kawanami, S., and Horie, S., Evaluation of the hearing of Japanese speech while wearing earplugs in a noisy environment, *Proceedings of Inter-Noise*, San Francisco, CA, 2015.

[170] Curran, J., A forgotten technique for resolving the occlusion effect, *Innovations*, 2(2), 2012.

[171] Casali, J.G., Horylev, M.J., and Grenell, J.F., A pilot study in the effects of hearing protection and ambient noise characteristics on intensity of uttered speech, *Trends in Ergonomics/Human Factors IV*, 303−310, Amsterdam: North-Holland, 1987.

[172] Abel, S. M., Boyne, S., and Roesler-Mulroney, H., Sound localization with an army helmet worn in combination with an in-ear advanced communication system, *Noise & Health*, 11: 199−205, 2009.

[173] Lee, K., Effects of Earplug Material, Insertion depth, and Measurement Technique on Hearing Occlusion Effect, Doctoral Dissertation, VirginiaTech, Blacksburg, 2011.

Appendix A

Auditory History Questionnaire

Auditory History Questionnaire

Title of research project:

Evaluation of changes in speech production induced by conventional and level-dependent hearing protectors and noise characteristics

Identification

Participant number:_____

Age:_____

Gender: ☐ Male ☐ Female ☐ You don't have an option that applies to me. I identify as (please specify) : _____

Date of birth:_____

Case History

Buzzing or ringing sounds in ears (tinnitus): ☐ yes ☐ no

If so, which ear: ☐ left ear ☐ right ear

Additional comments (intensity, frequency…): _____

Vertigo or dizziness: ☐ yes ☐ no

If so, when? _____

Have you ever had ear infections? ☐ yes ☐ no

If so, which ear: ☐ left ear ☐ right ear

Have you ever undergone ear surgery? ☐ yes ☐ no

If so, please specify? ☐ stapedectomy

☐ myringotomy

☐ tympanoplasty

☐ transtympanic tubes

☐ other: _____

Noise Exposure

Have you ever been exposed to loud noises? ☐ yes ☐ no

If so, in which situation? _____
_____

For how long? _____

Were you wearing hearing protectors? _____

Noisy Activities

- Use of firearms: ☐ yes ☐ no
  Firearm held on: ☐ left arm ☐ right arm

  Number of years: _____

- Snowmobile: ☐ more than 4 hrs/week ☐ less than 4hrs/week
- Motorcycle, CCV ("4 wheels"), motocross, car race: ☐ more than 4 hrs/week ☐ less than 4hrs/week
- Farm equipment: ☐ more than 4 hrs/week ☐ less than 4hrs/week
- Snow blower: ☐ more than 4 hrs/week ☐ less than 4hrs/week

- Workshop equipment: ☐ more than 4 hrs/week ☐ less than 4hrs/week
- Loud music exposure: ☐ more than 4 hrs/week ☐ less than 4hrs/week
- Chain saw: ☐ more than 4 hrs/week ☐ less than 4hrs/week
- Other noisy leisure activities: ☐ more than 4 hrs/week ☐ less than 4hrs/week
- (please specify): _____)

Appendix B

Consent form

Evaluation of changes in speech production induced by conventional and level-dependent hearing protectors and noise characteristics

Consent form

I, _____ hereby accept to participate in a research project funded by NSERC Discovery Canada. The goal of this project is to evaluate the effect of alterations in the characteristics of noise, hearing protectors, and sensorineural hearing loss on the Lombard effect and characteristics of speech produced in noise. I have carefully read the content of this letter and understand that I will be completing a questionnaire regarding my hearing history, after which I will be participating in a series of tests. The first consists of a basic hearing screening test, during which I will be required to press a button upon hearing soft sounds, while the others include measures of speech production in noise, during which I will be asked to read a list of sentences to a listener. These measures will be carried out and audio-recorded during a single testing session lasting approximately 120 minutes. I am also aware that I will receive a sum of $25 to cover the expenses related to my participation in this study, even if I choose to withdraw, and that I can request the results of my hearing test.

My participation is voluntary. The risks related to my participation in this study are minimal and have been explained to me. I have also received a detailed explanation of the expected results. I am aware that I will constantly have the opportunity to ask questions, to which I will receive appropriate answers. I understand that I may choose to withdraw from this research study at any time, for any reason.

This research is being conducted as part of PhD thesis under the supervisions of Professors Christian Giguère and Hilmi Dajani. By participating in this study, I will be contributing to efforts aimed at improving communications and safety in the workplace for all individuals wearing advanced hearing protection, including those with hearing loss.

If my candidacy cannot be retained because I do not meet the audiometric criteria, I will be referred, if necessary, to one of the following audiology services:

Clinique universitaire interprofessionnelle en soins de santé primaires
- Université d'Ottawa – Campus Alta Vista, 600 croissant Peter Moran, pièce 113, Ottawa, ON (613) 562-5800, ext. 8037

Ottawa Hospital
- General campus, 501, Smyth road (613) 737-8899
- Civic campus, 1053, Carling Avenue (613) 798-5555

Centre hospitalier de Gatineau
- 909 Boul. de la Vérendrye O., Gatineau (819) 561-8100

Centre hospitalier de Hull
- 116 Boul. Lionel Émond, Gatineau (Secteur Hull) (819) 595-6300

Polyclinique de l'oreille
- 120 Boul. de l'Hôpital, Gatineau (819) 561-0002
- 290 Boul. St-Joseph, Gatineau (Secteur Hull) (819) 776-0163

Helix Hearing Care Center

- 1224 promenade Place d'Orleans (613) 834-7655
- 595 chemin Montreal (613) 745-0862
- 1385 rue Bank (613) 737-3500

Hearing Solutions Clinic

- 202-1915 chemin Baseline (613) 288-0295
- The Sound Room Ottawa
- 1-1800 rue Bank (613) 627-4700

In order to ensure data confidentiality, only the researchers mentioned below will have access to the collected data. I understand that, under no circumstances, will my identity be revealed and that a number assigned to my file will ensure this anonymity. However, my identity may be known to the listener; hence, the protection of my identity cannot be entirely guaranteed. Finally, I am aware that the collected data will be locked away in the Audiology Research Laboratory, Roger Guindon Hall, room 1117, and will be destroyed five years after the publication of results.

For additional information regarding this research, I may contact any of the researchers involved in this study, Ghazaleh Vaziri, Christian Giguère, Hilmi Dajani, whose contact information is found below. For all information, requests or complaints regarding the ethical conduct of this study, my enquiries can be addressed to the Protocol Officer for Ethics in Research at the following contact information: Office of Research Ethics and Integrity, University of Ottawa, Tabaret Hall, 550 Cumberland St, Room 154, Ottawa, Ontario K1N 6N5, phone number: (613) 562-5387, e-mail: ethics@uottawa.ca. There are two copies of the consent form, one of which I may keep.

At any time, if I choose to withdraw from this study, I will be given the option of withdrawing or retaining my collected data.

I authorize the researchers to solicit my participation in future studies.

Researcher's signature: _____ Date: _____

Participant's signature:_____ Date: _____

Names of researchers: Ghazaleh Vaziri, Ph.D. candidate, Christian Giguère, P.Eng, Ph.D., Hilmi Dajani, Ph.D., University of Ottawa, Faculty of Health Sciences, School of Rehabilitation Sciences, 451 Smyth road, Ottawa, Ontario, K1H 8M5. Phone number: (613) 562-5800, ext. 4649, 6217. Fax number: (613) 562-5428.