LOMBARD EFFECT IN SPEECH PRODUCTION BY COCHLEAR IMPLANT USERS: ANALYSIS, ASSESSMENT AND IMPLICATIONS

by

Jaewook Lee

APPROVED BY SUPERVISORY COMMITTEE:

Dr. John H. L. Hansen, Chair

Dr. Peter F. Assmann

Dr. P. K. Rajasekaran

Dr. Carlos Busso

LOMBARD EFFECT IN SPEECH PRODUCTION BY COCHLEAR IMPLANT USERS: ANALYSIS, ASSESSMENT AND IMPLICATIONS

by

JAEWOOK LEE, BE, ME

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

May 2017

To my wife Yoojin.

ACKNOWLEDGMENTS

I would like to acknowledge my advising professor, Dr. John H. L. Hansen, for his guidance and support. He provided unconditional support and expert advice throughout the years of my dissertation research. This learning experience from him not only enhanced my research skills, but also encouraged me to grow as a independent researcher. I would like to express my sincere appreciation to my committee members: Professors Peter F. Assmann, P. K. Rajasekaran, and Carlos Busso for their time and comments on this work. I would like to acknowledge the National Institute on Deafness and other Communication Disorders, National Institutes of Health. My undertaking of research was made possible in part by them. I would like to acknowledge the cochlear implant and normal hearing participants for their time and dedication. I would like to thank the members of the Center for Robust Speech Systems - Cochlear Implant Laboratory at The University of Texas at Dallas. I am grateful to Hua Xing, Dongmei Wang and Chengzhu Yu who were always available whenever I needed help. I also thank Dr. Hussnain Ali and Dr. Ali Ziaei for their help and collaborative suggestions in this work. I am so grateful to my parents and parents-in-law for all their love and support. They always pray for me with devoted affection. I would like to give special thanks to my son, Donghae. He gave me a lot of pleasures during study in the United States. I would like to thank my wife, Yoojin, for her encouragement, patience and underlying support throughout my life. "Your love and trust in me made this work possible."

March 2017

LOMBARD EFFECT IN SPEECH PRODUCTION BY COCHLEAR IMPLANT USERS: ANALYSIS, ASSESSMENT AND IMPLICATIONS

Jaewook Lee, PhD
The University of Texas at Dallas, 2017

Supervising Professor: Dr. John H. L. Hansen, Chair

In daily communication, speakers aim to communicate their message in a manner that is intelligible to listeners. When individuals with normal hearing become aware of reduced auditory feedback due to environmental noise, they likely adopt a different speaking style called the 'Lombard effect'. The Lombard effect is the tendency of speakers to modify their speech production while speaking in the presence of loud noise. Increased levels of masking noise lead to increase in vocal effort including energy, fundamental frequency, and glottal spectral slope. Lombard speech modification is aimed at providing the listener with increased speech intelligibility in challenging listening environments. The Lombard effect is also known to degrade automatic speech systems such as automatic speech recognition (ASR) and speaker identification (SID). While well-studied for normal hearing listeners and automatic speech systems, the Lombard effect has received little, if any, attention in the field of cochlear implant users. To our knowledge, no study has examined whether cochlear implant users employ the Lombard effect during voice communication. This dissertation provided a comprehensive investigation of the research concerning Lombard effect for cochlear implant users with post-lingual deafness. We investigated the nature of the Lombard effect that was produced by cochlear implant users. A variety of acoustic and phonetic features including voice power, fundamental frequency, glottal spectral tilt, phoneme duration, and formant

frequencies were analyzed. Mobile personal audio recordings from continuous single-session audio streams collected over an individual's daily life were used for these analyses. Prior advancements in this domain include the "Prof-Life-Log"longitudinal study at UT-Dallas. The Lombard effect was observed in the speech production of all cochlear implant users. The results indicate that both suprasegmental (*e.g.,* F0, glottal spectral slope and vocal intensity) and segmental (*e.g.,* F1 for /i/ and /u/)) features were altered in such environments. Along with speech production characteristics, the research also focused on the effect of Lombard speaking style on intelligibility by cochlear implant users. A speech corpus for the perceptual experiments of Lombard speech was formulated with normal hearing speakers. A subjective listening test was performed to provide how cochlear implant users respond to Lombard speech in challenging listening environments. The results indicates that the Lombard speech yielded a significant improvement in intelligibility in both quiet and noisy listening conditions. The specific modification of speech production of cochlear implant users under the Lombard effect may contribute to some degree an intelligible communication in adverse noisy environments. Lastly, a practical implication of Lombard effect for developing speech enhancement algorithm for cochlear implant users was discussed. A previous proposed framework based on Source Generator theory was employed to perturb neutral speech production based on Lombard effect modification. Data from subjective evaluation demonstrated the effectiveness of the proposed speech enhancement algorithm. The results indicated improvement in intelligibility when providing neutral speech which was modified based on Lombard effect properties via the proposed algorithm. The specific variations due to Lombard effect can be leveraged for new algorithm development and further applications of speech technology to benefit cochlear implant users.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Speech Produced in Noisy Environments

Changes in speech production, *e.g.,* vocal effort based on auditory feedback, are an important research domain for improving human-to-human and human-to-machine communication. For example, in the presence of environmental noise, a speaker experiences the well-known phenomenon known as Lombard effect (Lombard, 1911; Lane and Tranel, 1971). Lombard effect is perceptually realized with changes in speech production while speaking in challenging listening environments. Increase in the levels of masking noise lead to increase in vocal effort such as altered vocal intensity, glottal spectral slope, fundamental frequency or formant location of speech (Hansen, 1988; Pisoni et al., 1985; Junqua, 1992; Sodersten et al., 2005). Lombard speech modification effects intelligibility, such that the resulting speech can be easily understood by the listeners or the speaker themselves in noisy listening environments (Dreher and O'Neill, 1957; Summers et al., 1988; Lu and Cooke, 2008; Garnier et al., 2010).

Lombard effect has been widely studied in automatic speech systems, where is is known to degrade the quality of automatic speech recognition (ASR) and speaker identification (SID) systems (Hansen, 1988; Hansen and Varadarajan, 2009; Junqua, 1992; Bořil and Hansen, 2010). Since fundamental differences from neutral speech, Lombard effect speech causes a breakdown in speech system performance when systems are modeled with neutral speech. A range of signal processing techniques has been proposed to compensate for the effect of Lombard in speech to improve the robustness of the speech systems (Hansen, 1996, 1994; Hansen and Cairns, 1995; Bou-Ghazale and Hansen, 1996). The knowledge of Lombard effect can be integrated within a recognition system to reduce the effect of speech under noise.

Lombard speech research has another important implication for developing speech enhancement algorithm (Schepker et al., 2013; Zorila et al., 2012; Godoy and Stylianou, 2013;

Cooke et al., 2013). The aim of this study is to control the acoustic parameters of neutral speech in order to present intelligibility-enhancing speaking styles, such as Lombard or clear speech. For example, a set of acoustic parameters, such as spectral shaping and dynamic range compression were employed to simulate Lombard speaking styles from neutral speech (Zorila et al., 2012; Jokinen et al., 2016). A modification based on clear-style speech was also proposed for the purpose of increased intelligibility in noisy environments (Godoy and Stylianou, 2013; Godoy et al., 2014). The proposed approaches mentioned above were highly successful at increasing speech intelligibility of normal hearing listeners.

1.2 Problem Statement

The Lombard effect has been well studied for normal hearing listeners, automatic recognition systems and speech enhancement systems (Hansen, 1988; Junqua, 1996; Lu and Cooke, 2008; Garnier et al., 2010). However, not all of the Lombard research goals have been achieved to date. Most Lombard speech research have focused on adults with normal hearing. Numerous acoustic-phonetic features analyses and its intelligibility benefits have been reported for this listener population across studies. Currently, we need to expand this investigation by including more varied talker groups and listener groups, *e.g.,* a new clinical population: cochlear implant users.

A cochlear implant is an electronic device that is surgically implanted in the inner ear which directly stimulates the auditory nerve to provide a partial sense of sound (Loizou, 1998; Zeng et al., 2008; Wilson et al., 1991). It can be used for children and adults who are deaf or who have extreme hearing loss, and helps them to communicate with other people. Many patients communicate with only minor difficulty, use the telephone, and have resumed careers interrupted by profound deafness. The field of cochlear implants has experienced a considerable growth over the past decade (NIDCD, 2012). An investigation of this new population further our understanding on speech produced under noisy environments.

In order to analyze speech under noise, previous studies regarding Lombard effect analysis mostly employed read speech elicited by laboratory conditions while the speakers listened to noise through headphones (Lane and Tranel, 1971; Pisoni et al., 1985; Sodersten et al., 2005). This traditional paradigm, however, does not reflect acoustic variability in real-life communicative situations. For example, conventional datasets do not emphasize communication because the speaker was reading a list (Lindblom, 1990). Headphones may also induce additional degradation of auditory feedback, which may influence on the magnitude of vocal effort (Garnier et al., 2010; Ikeno et al., 2007). A development of new sets of materials which is collected in realistic listening environments is necessary to understand speech under realistic scenarios.

Finally, there is a potential feasibility in algorithmic modification of neutral speech to increase intelligibility in noisy environments (Zorila et al., 2012; Godoy and Stylianou, 2013; Schepker et al., 2013). However, the abovementioned speech enhancement approaches were used either only for near-end or text-to-speech systems. Virtually, none of the approaches have ever been employed for cochlear implant areas. Historically, conventional signal processing for cochlear implants have focused on noise suppression strategies to improve performance under adverse noisy environments (Loizou, 1999; Zeng et al., 2008; Wilson et al., 1991). However, many of these techniques are perceptually motivated and might not provide benefits under both stationary and non-stationary maskers. Rather than the method suppressing background noise, it is more desirable to have an effective way to enhance the speech signal, *i.e.* speech feature modifications.

1.3 Dissertation Contributions

This study contributed to the knowledge concerning the Lombard effect for cochlear implant users in several ways. The primary focus of this dissertation was acoustic and perceptual analyses of speech produced in noisy environments. Another main goal was to find a new

algorithm to improve speech recognition in noise for cochlear implant users by exploring the characteristics of Lombard effect. The proposed research in this dissertation, thus, provided both theoretical insight and practical application of Lombard effect for cochlear implant users. The contributions of the dissertation are discussed in the next section in greater detail.

First, this study developed a new speech corpus that can be used to investigate speech production in naturalistic environments. Mobile personal audio recordings from continuous single-session audio streams were used to collect speech production over an individual's daily life. Speakers produced read and conversational speech in on-campus environments at UT-Dallas, *e.g.,* office, cafeteria. Additional audio recordings were collected in uncontrolled off-campus environments *e.g.,* home, restaurants. A set of acoustic and orthographic transcription labels were assigned by a human transcriber. A total of 100 hours of personal audio recording were collected from 6 cochlear implant and 18 normal hearing participants.

Second, this dissertation analyzed speech production of cochlear implant adults with respect to environmental changes. The analysis conducted was based on personal audio recordings collected in UT-Dallas on-campus environments. Several approaches were used to characterize real-world listening environments, including long-term averaged spectra and signal-to-noise ratios. The parametric variations in vowel, consonant and individual phoneme production were investigated as a function of varying environments. This involved speech power, fundamental frequency, glottal spectral tilt, phoneme duration, and formant frequencies. Data from these analyses showed that the relationship of speech production parameters depend upon varying type of environments.

Third, the study examined the perceptual analysis of Lombard effect by post-lingually deafened cochlear implant users. A speech corpus intended for the perceptual experiments of Lombard speech was developed for this analysis. Normal hearing speakers read sentences to their communication partner while listening to noise samples through open-air headphone.

Then, subjective intelligibility was measured with cochlear implant users across signal-to-noise ratios. Data from perceptual analysis showed that differences between perception performance of speech production in quiet and noisy environments.

Fourth, this dissertation proposed a Lombard effect-based speech enhancement algorithm for cochlear implant users. The proposed algorithm controlled the acoustic parameters of neutral speech to present Lombard speaking style to improve intelligibility in noisy environments. Acoustic variations for neutral and Lombard conditions were modeled and used to generate Lombard synthetic speech. The modification areas considered in the proposed algorithm were: (1) voice intensity, (2) overall spectral contour, and (3) sentence duration. The analysis and modeling conducted here will be based on a previously established framework, Source Generator theory (Hansen and Cairns, 1995; Hansen, 1994). Data from subjective listening evaluation with cochlear implant users was presented to demonstrate the effectiveness of the proposed speech enhancement algorithm.

1.4 Dissertation Outline

This dissertation is organized as follows. In Chapter 2, we discuss prior studies that examine speech perception and production of adult cochlear implants users. In addition, Chapter 2 offers a brief summary of existing speech processing techniques for cochlear implants to improve performance in challenging listening environments. Chapter 3 presents a development of a new speech corpus that can be used to investigate the speech production of cochlear implant users in daily-life environments. Mobile personal audio recordings from continuous single-session audio streams are employed to examine speech production under naturalistic listening environments. Chapter 4 presents an investigation of speech produced by adult cochlear implant users with respect to varying environments. In Chapter 4, data from acoustic-phonetic analysis including voice power, fundamental frequency, glottal spectral tilt, phoneme duration, and formant locations are presented. Chapter 5 provides a focused study

examining the influence of Lombard effect on speech perception of post-lingually deafened cochlear implant adults. In Chapter 5, data from subjective intelligibility test is presented to provide a perceptual characteristics of Lombard speech. In Chapter 6, we introduce a Lombard effect signal processing for cochlear implant users to enhance intelligibility under noise. Chapter 6 proposes an algorithm to modify speech parameters of neutral speech based on Lombard effect in order to produce Lombard speaking style. Finally, Chapter 7 summarizes our findings, and suggests directions for further research.

CHAPTER 2

BACKGROUND

In this chapter, we described cochlear implants and speech processing strategies. Also, we reviewed previous findings regarding speech perception and production of post-lingual deafened cochlear implant users. Lastly, a brief overview of the existing speech enhancement algorithms for cochlear implant users was reviewed and discussed.

2.1 Fundamentals of Cochlear Implants

A cochlear implant is an electrical device that is surgically implanted in the inner ear which directly stimulates the auditory nerve to provide a partial sense of sound (Loizou, 1999; Zeng et al., 2008; Wilson et al., 1991). Cochlear implants are used to provide auditory information to individuals experiencing profound hearing impairments. A cochlear implant device consists of three major parts: microphone, signal processor, and transmission system (see Figure 2.1). The signal processor breaks the input signal into different frequency bands and convert the filtered signal into electrical signals. The transmission system conveys the electrical signals to an electrode located at the cochlear. The electrode stimulates different auditory fibers at different places. For example, the electrode located at the cochlear base are stimulated with high frequency signal, while electrode at the apex are stimulated with low frequency signals (see bottom panel in Figure 2.1).

2.1.1 Coding Strategies

The type of auditory information transmitted to each electrode heavily depends on the type of implant and the signal processing strategy (Loizou, 1998; Zeng et al., 2008; Wilson et al., 1991). Loizou (Loizou, 1998) summarized three kinds of speech processing strategies: waveform strategies, feature extraction strategies, and "n-of-m"strategies. The waveform

strategies filtered the input signal into multiple frequency bands, and sent to different electrodes non-simultaneously. The waveform strategy is used in Ineraid processor (Eddington, 1980). The feature extraction strategies delivered spectral features such as F0, F1, and F2 to appropriate electrodes in the electrode array. This strategy is used in Nucleus multi-channel processor (Tong et al., 1980). The "n-of-m"strategy is a combination of waveform and fea-

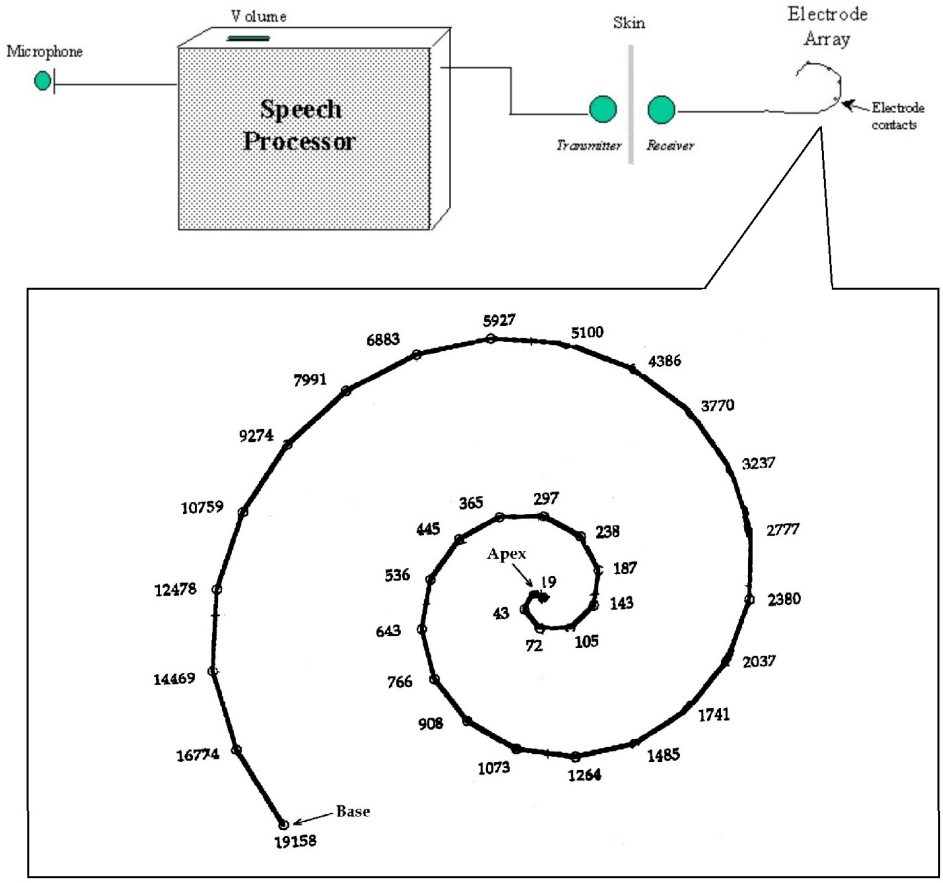


Figure 2.1. Diagram illustrating the basic operation of a cochlear implant. Sound is captured by a microphone and sent to a coding processor worn by the user. The sound is then signal processed, and electrical stimuli are sent to the electrodes through a RF link. Bottom figure shows diagram showing the basilar membrane indicating the base and the apex of cochlear. The location of displacement in response to audio signal of different frequency in Hz is presented (Loizou, 1998).

ture extraction strategies. The "n-of-m"strategy are used in Nucleus, Clarion and Med-El processors (Skinner et al., 1991; Friesen et al., 2001; Müller et al., 2002).

2.1.2 Populations

Young children and adults who have severe to profound hearing loss can be fitted with cochlear implants. For instance, children who born with hearing loss, or whose hearing loss occurred before language acquisition (*i.e.,* pre-lingual deafness) can take advantage of this device (Svirsky et al., 2000). Using a cochlear implant allows them to expose to sounds during an optimal period to develop speech and language skills. In addition to young children, adults who have lost their hearing after acquisition of language (*i.e.,* post-lingual deafness) can also be benefit from cochlear implants. They are able to learn how to associate the signal provided by an implant with sound they remember. This often provides them to identify speech to large extent without lip-reading or sign language. In this review, we focused on investigating post-lingually deafened cochlear implant users.

The field of cochlear implants has experienced a considerable growth over the past decade. Many patients communicate with only minor difficulty, use the telephone, and have resumed careers interrupted by profound deafness. According to the U.S Food and Drug Administration (FDA), roughly 58,000 and 38,000 devices have been implanted in adults and children respectively in the United States (NIDCD, 2012). As of December 2012, approximately 324,000 have been implanted worldwide (NIDCD, 2012).

2.2 Speech Perception with Cochlear Implants

Previous studies have considered how restoration of hearing with a cochlear implant may lead to changes in speech perception of post-lingual deafened adults. The speech perception presented through the cochlear implant is heavily influenced by the choice of speech processing strategies (Loizou, 1998; Wilson et al., 1991; Zeng et al., 2008). The first report

in speech perception for patients fitted with single-channel cochlear implants were not very promising (Doyle et al., 1964; Simmons, 1966). At best, speech signals were recognized as speech but were not intelligible. This was not surprising given the poor frequency resolution and poor detail in the time waveform.

2.2.1 Single-Channel Devices

As the number of patients increased to hundreds, different results were found in other studies (Bilger et al., 1977; Hochmair-Desoyer et al., 1981; Tyler, 1988). Bilger *et al.* (Bilger et al., 1977) reported that restored auditory feedback after cochlear implantation is a crucial factor for increased adult speech perception. In this study, significant improvement in speech intelligibility was found in House single channel cochlear implant users. Hochmair-Desoyer (Hochmair-Desoyer et al., 1981) found similar result for patient who use the Vienna single channel implant. Patient who use this device have shown good speech understanding, achieving a mean score of 45 percent points for words in sentences.

2.2.2 Multi-Channel Devices

Unlike single channel devices, multichannel implants provide electrical stimulation at multiple sites in the cochlea using an array of electrodes (see Figure 2.2). As the number of channel increased, reports of patient with good, even sometimes remarkably good, speech recognition abilities surfaced (Skinner et al., 1991; Müller et al., 2002; Friesen et al., 2001). For instance, Nucleus device from Cochlear Ltd. is in widespread use and use a 22-electrode array to stimulate the electrodes in a non-simultaneous manner. Patients with Nucleus device have achieved scores of 80 percent points on single-syllable words, and 100 percent points correct on test of words in sentences (Skinner et al., 1991) . In addition to the Nucleus device, the Med-El and Clarion processors are also in widely use, and provide the same level of speech understanding (Müller et al., 2002; Friesen et al., 2001). The increased number of channel

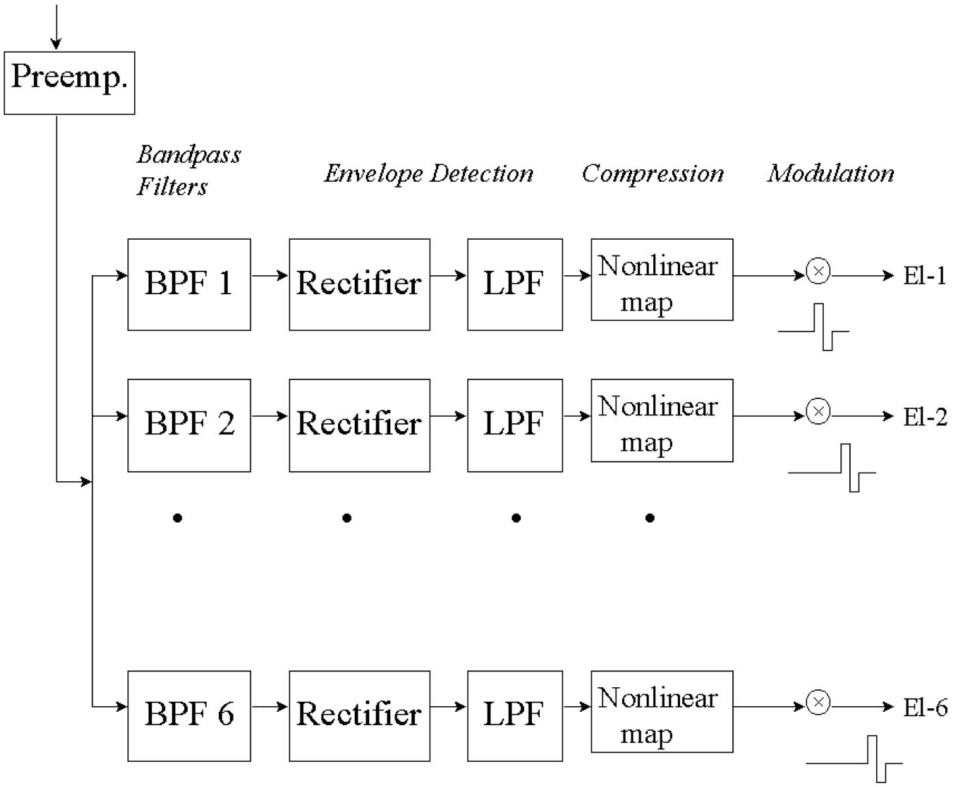


Figure 2.2. Diagram illustrating the signal processing of a multi-channel cochlear implant (*i.e.,* six-channel). Sound sent to the processor is first pre-emphasized, then band-pass filtered. The envelop of the sound are then estimated for each channel by rectification and low-pass filtering. A dynamic range compression function was used to reduce the volume of loud sounds or amplifies quiet sounds. Current pulses produced with the envelops of each channel are delivered to the electrodes (Loizou, 1998).

provided by these devices contributed to better transmission of both frequency and temporal information for perception.

2.2.3 Other Internal Factors

In addition to the external factors (*e.g.,* processor types or single- or multi-channel strategies), several internal factors appear to influence the performance of cochlear implants. Two major factors that have been found to affect perception performance are the amount of residual hearing and the duration of deafness. The amount of residual hearing is also re-

lated to speech recognition ability. It has been shown that the patient who had the most residual hearing were better at word recognition than other implant patients (Tye-Murray et al., 1988; Tyler, 1988). The duration of hearing loss prior to implantation has been found to have negative effect on auditory performance (Dorman et al., 1989; Dowell et al., 1986; Tyler et al., 1988). Individuals with shorter duration of hearing loss tend to achieve better performance than individuals with longer duration of hearing loss. Other factors that may affect auditory performance include the age of hearing loss onset, the duration of cochlear implant use, and age at implantation (Blamey et al., 1996; Dorman and Spahr, 2006).

2.3 Influence of Auditory Feedback on Speech Production

Auditory feedback refers to information received from the sense of hearing. This is internally monitored by the speaker and received by a listener (see Figure 2.3). Auditory feedback is closely related to the development, articulation and modification of speech production. During speech production, a person receives auditory feedback through two separate channels, air and bone conduction. The air conduction channel involves the transmission of acoustic signal from the outer ear to the inner ear. The bone conduction channel involves the transmission of acoustic energy from bones of the skull to the fluids of the inner ear.

Auditory feedback appear to interact with the control of some suprasegmental and segmental properties of human speech production system. Speech is commonly described as having two components: segmental and suprasegmental features. The term segmental features refer to vowel and consonants. The term suprasegmental features refer to larger units of linguistic organization, such as syllables, words, and sentences. The suprasegmental feature involves sound pressure level, fundamental frequency, spectral tilt, and speaking rate. The segmental features includes formant frequencies, phoneme duration, and vowel-consonant ratio. The role of auditory feedback on segmental and suprasegmental features has been investigated by studying speech produced by deaf individuals with cochlear implant.

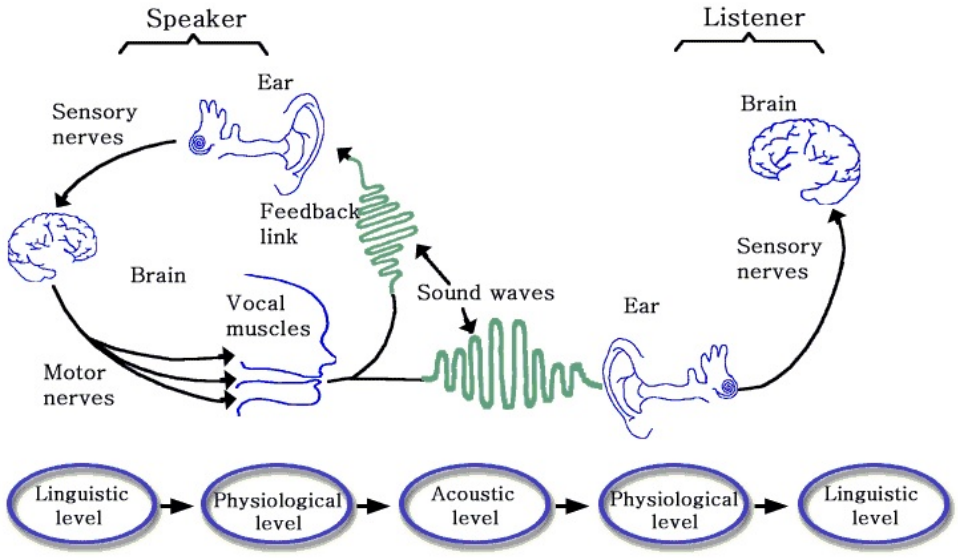


Figure 2.3. Diagram illustrating how message are communicated from speaker to listener through speech. The conveyance of this idea involves linguistic, physiologic and acoustic processes. Auditory feedback assist in regulating these complex interaction through two separate channels, air and bone conduction (Denes and Pinson, 1993).

2.3.1 Longitudinal Effect

While normal-hearing individuals receive auditory feedback via both air and bone conduction, individuals with profound hearing loss receive little or no feedback from either of the channels. Profound hearing loss may not hear speech even at 90 dB HL, and as consequence, they have little or no feedback from bone conduction channel. Elimination of auditory feedback after the acquisition of language results in poorer regulation of suprasegmental properties (Kirk and Edgerton, 1983; Lane and Webster, 1991; Leder et al., 1987). These include alteration of intensity, fundamental frequency, and temporal properties. Segmental performance is also adversely affected with reduced or eliminated auditory feedback (Lane and Webster, 1991). These features, however, are often coded in the processing strategies used in many cochlear implants. Thus, one might anticipate changes occurring to these parameters in speakers using cochlear implants.

Previous studies have considered how restoration of auditory feedback with a cochlear implant may lead to changes in speech production (Hochmair-Desoyer et al., 1981; Kirk and Edgerton, 1983). Some of them have been performed to examine long-term longitudinal effects on changes on speech. Hochmair-Desoyer *et al.* (Hochmair-Desoyer et al., 1981) suggested improved quality of speech production for post-lingually deafened adults who were fitted with cochlear implants. In that study, changes in vowel production and fundamental frequency have been found when compared to subject characteristics before implantation. Kirk and Edgerton (Kirk and Edgerton, 1983) also examined the suprasegmental properties of post-lingually deaf adults who received a single channel processor. They reported nearly normal production of fundamental frequency and intensity in speech, but the changes vary greatly from individual to individual.

2.3.2 Acute Effect

In addition to the longitudinal effect, some groups have investigated a more precise effect of auditory feedback on speech production by the speech processor turned on or off (Svirsky and Tobey, 1991; Svirsky et al., 1992). While the speech processor is turned off, auditory feedback is not available from either air or bone conduction channels. This strategy able to offer to explore more acute effect of auditory feedback on speech production (*e.g.,* from few seconds to minute). Svirsky and Tobey (Svirsky and Tobey, 1991), for example, suggested a brief absence of auditory feedback and its influences on speech production. According to the research, rapid changes in vowel formant frequencies were observed within a few seconds of turning the speech processor on or off. Svirsky *et al.* (Svirsky et al., 1992) also argued that many suprasegmental parameters, including pitch period and vowel duration, demonstrated instantaneous response to the short-time deprivation of auditory feedback.

Recent studies regarding phonetic analysis proposed more widely recognized views concerning the role of auditory feedback on speech production (Vick et al., 2001; Matthies et al.,

1996; Lane et al., 2007). Lane *et al.* (Lane et al., 2007), for example, suggested that auditory feedback played a dual role in speech production based on post-lingually deafened cochlear implant users. According to this research, auditory feedback has two important roles in speech production: regulation of suprasegmental features, and control of segmental features. These authors suggested that the suprasegmental features of speech change relatively rapidly with the change in hearing status. However, segmental features of speech, which are controlled by an internal model, are generally stable and change only in the long-term.

Perkell *et al.* (Perkell et al., 1997) also found similar results for the patients with cochlear implants. They reported that speech production appears to interact with auditory feedback within both short-time and long-term temporal windows. In that research, auditory feedback not only calibrates articulatory target over relatively long periods of time (*e.g.,* pre- and post-implantation), but also fine tunes articulatory movements within in short-time period (*e.g.,* from a few seconds to minutes). The study suggested that phonemic parameters do not change immediately following alterations to auditory feedback. Any immediate modifications found in phonemic variables are indirect consequences of modifications in suprasegmental variables.

2.4 Performance of Cochlear Implants in Noisy Environments

As previously discussed, cochlear implants contribute to reasonable speech understanding in quiet settings. Nonetheless, speech recognition in noisy situations is still a very challenging task for cochlear implant users. For example, while cochlear implants achieve open-set speech recognition scores of 80 percent or higher, their performance degrades significantly by environmental noise. Previous studies in cochlear implants have shown that the absence of fine spectral structure may contribute to the poorer performance under noise (Fu et al., 1998). It has been also known that reduced temporal fine structure which is important for pitch perception may exacerbate the speech perception faced by cochlear implant users in

noise (Moore, 2008). For the reason, cochlear implant listeners need higher signal-to-noise ratios to achieve similar levels of performance to that of normal hearing listeners.

Most commercially available cochlear implants do not perform well in complex listening situation for two reasons (Loizou, 1998; Zeng et al., 2008; Wilson et al., 1991). First, conventional speech processors do not offer the option to change speech coding parameters to different environmental types (Skinner et al., 1991; Müller et al., 2002; Friesen et al., 2001). They only allow users to select one of four (in most devices) prestored MAPs programmed during the initial cochlear implant fitting in the clinic. Next, the two processors fitted in the two ear run independently of one another. Having two ears allows listeners to take advantage of head-shadow effect, which provide access to binaural difference cues for localization (Doclo et al., 2015; Wouters and Berghe, 2001). This in general provides better speech perception in noise when the target signal is spatially separated from the masker. As two implants operate independently, the transmission of interaural time difference (ITD) cues is not done effectively.

2.4.1 Noise Reduction Strategies

In order to facilitate more intelligible speech, previous studies regarding front-end processing for cochlear implants have focused on noise suppression algorithm (Doclo et al., 2015; Hersbach et al., 2013; Wouters and Berghe, 2001; Loizou et al., 2005; Yang and Fu, 2005; Hu and Loizou, 2007). Numerous multi-channel and single-channel speech enhancement algorithms have been proposed to improve speech intelligibility for cochlear implant users. At one end, multi-channel methods (*e.g.,* beam forming and multi-channel Wiener filter techniques) have been shown to effective for improving speech intelligibility for cochlear implant recipients (Doclo et al., 2015; Hersbach et al., 2013; Wouters and Berghe, 2001). However, multi-channel methods are not effective when there is a high amount of reverberation or if the speech and noise sources are from the same direction.

For single-channel noise reduction, traditional signal processing approaches, such as subspace method (Loizou et al., 2005), spectral subtraction (Yang and Fu, 2005), Wiener filter, MMSE (Hu and Loizou, 2007) etc. have been employed. For example, Yang *et al.* (Yang and Fu, 2005) evaluated a spectral-subtractive algorithm and showed significant benefits in sentence recognition, particularly in speech-shaped noise. The study by Loizou (Loizou et al., 2005) evaluated the performance of a subspace noise reduction algorithm. Results showed significant benefits to cochlear implant users in regard to recognition of sentences corrupted by stationary noise. However, it was not clear whether such intelligibility benefits will sustain if the algorithm was tested in non-stationary environments.

2.4.2 Channel Selection

Some techniques have been used to attempt to integrate the algorithm into existing coding strategies, such as channel selection strategy (Kim and Loizou, 2010; Hazrati et al., 2013; Healy et al., 2013). It first examines each channel to see whether it is noise dominant or speech dominant, then, it eliminates the noise by gating the speech signal on and off. Speech-dominant channels are, thus, preserved. A number of techniques have been proposed to segregate the target from noise. These are mostly focused on auditory models and make extensive use of grouping principles, such as pitch continuity and sound localization cues. However, most of channel selection strategy assumed that the local signal-to-noise ratios are known (Hazrati and Loizou, 2012a,b). In practice, however, estimating local SNRs is a challenging task from noisy data, particularly in adverse noisy conditions.

2.4.3 Environmental Optimization

Current sound processing have included auditory scene analysis to optimize different environments (Hu and Loizou, 2010; Goehring et al., 2015; Hazrati et al., 2014). In brief, the speech processor continuously monitors the background noise to identify the type of noise, and then

engages the appropriate noise suppression approach. This environment customized paradigm have adjusted processor parameters (*e.g.,* compression function and pulse rate), and offer additional benefits of existing cochlear implant devices. A previous finding by Fu and Shannon (Fu et al., 1998) further support this position. They clearly demonstrated that maximum benefit in terms of intelligibility was obtained when cochlear implants are customized to the various listening situations. Although cochlear implant devices are adjustable, these techniques require the computational complexity, thus can not be easily integrated with existing devices.

2.5 Summary: Chapter 2

This chapter reviewed previous literature that address cochlear implants and the performance of adult cochlear implant users. Previous literature regarding the speech perception capabilities of individuals using cochlear implants were discussed. Several internal and external factors were also found to influence the effectiveness of cochlear implants across studies. However, not all of the factors for cochlear implant performance have been achieved to date. There are very limited data concerning how cochlear implant users respond to the speech produced under realistic communication scenarios, such as acoustic noise. This issue will be addressed in Chapter 5.

The linkage between speech production and perception was investigated by studying speech production by adults with cochlear implants. These studies suggested that speech production is heavily influenced by any form of auditory feedback. However, changes in speech production of cochlear implant users have been examined when auditory feedback was artificially distorted (*e.g.,* turning processor on/off or pre- or post-implantation). These experimental disruptions of auditory feedback, however, do not provide information about speech production in real communication scenarios, *i.e.,* noisy environments. Additional studies regarding acoustic and phonetic features in more practical listening environments

will be necessary. Findings from such study may cast light on the speech production of post-lingually deafened adults. This issue will be considered in Chapter 3 and Chapter 4.

Finally, the review introduced existing front-end signal processing for cochlear implant users to enhance the recognition under noise. Conventional sound processing techniques for cochlear implants have mostly focused on noise suppression algorithms. Although significant progress has been made, each of them has its own drawbacks and limitations. Specifically, computation complexity, estimating local signal-to-noise ratios, and requirement of second microphone are all considered to be the main drawbacks. The low benefit obtained in non-stationary noise poses another major limitation of the noise suppression approaches. For these reasons, an alternative way of enhancing speech signal is desirable to suppress background noise. This issue will be described in Chapter 6.

CHAPTER 3

UTD-CI-LENA: DEVELOPMENT OF NATURALISTIC SPEECH CORPUS

3.1 Introduction

In order to investigate speech production under noise, the most classic paradigm to simulate The Lombard effect has consisted of playing noise over headphones to individuals seated alone in a recording booth and reading out lists of words or sentences (Lu and Cooke, 2008; Junqua, 1992; Summers et al., 1988; Pisoni et al., 1985). Using such a method, a great variability in the increase of vocal effort has been observed. However, this paradigm may raise a number of concerns. One of concerns is that the speech task has been shown to affect increase of vocal intensity in noise that is greater in experimental situations (Garnier et al., 2010; Ikeno et al., 2007). We presumed that headphones may induce additional degradation of the auditory feedback, independently from the noise played into them. Another concern on the traditional paradigm is that talkers produce the speech that is more complex in natural communication settings (Lindblom, 1990). This involves producing utterances that are longer than syllables, words, or short sentences. Spontaneously produced speech, therefore, may show different pattern of acoustic and articulatory adjustments compared to speech produced in laboratory conditions. A development of new sets of materials, collected in realistic listening environments, is necessary to understand speech under realistic scenarios.

3.1.1 Objectives and Proposed Methods

In this chapter, we focused on the development of new sets of speech materials which is collected in realistic listening environments. The corpus established here was designed to capture audio from cochlear implant users to investigate any potential speech production modifications in naturalistic settings. The mobile personal audio recordings were used for

collecting naturalistic audio from cochlear implant users. The personal audio recordings are continuous single-session audio streams collected over an individual's daily life. These recordings contain an abundance of information regarding speaker, speech, environments, language, etc. This offers a unique opportunity to explore human-to-human interactions within the communication setting, as well as how the environment impact on the information exchange. Prior advancements in this domain include the Prof-Life-Log longitudinal study at UT-Dallas (Sangwan et al., 2012; Ziaei et al., 2012, 2013) which have explored speech communication in naturalistic daily life.

3.1.2 Application areas

Over the last few decades, analyzing personal audio recording has been considered to be a challenging task. This is because of their large data size (*i.e.,* typically 8-16 hours/day), diversity in communication style, and little prior knowledge concerning speaker and environments. Speech and language processing capability (*e.g.,* ASR, SID, etc.) in conjunction with personal mobile computing devices (*e.g.,* smartphone, Google Glass (Google Inc, 2014; Paxton et al., 2015)), however, opened new opportunities for data mining. Long-term personal audio recordings hold a wide range of potential applications for cochlear implant research.

One promising use of naturalistic audio streams is in analyzing language acquisition and development of infant and young children (Gilkerson and Richards, 2008; Hart and Risley, 1995; Xu et al., 2008). These analyses have been performed by measuring various metrics of interest, such as adult word count, adult-child turn taking, and child vocalizations. These indicators have helped to assess and monitor the natural language environment of children, especially the quantity of communicative input to or communicative exchange with children. Another application of personal audio recording is in the use of screening and diagnosis procedures for speech-related disorders in early childhood (Oller et al., 2010; Warren et al., 2010). Analysis of acoustic features from real-world audio recordings make it possible to

differentiate children with and without neurodevelopmental disorders such as Autism or language delay.

Finally, the capability of audio environment detection in conjunction with an appropriate environmental-optimized coding algorithm can also be used in personal audio recordings (Hu and Loizou, 2010; Goehring et al., 2015; Ali et al., 2013; Hazrati et al., 2014). These environment customized paradigms can adjust processor parameters (*e.g.* compression function or pulse rate), which are currently not optimized either for individuals or specific environments. The next generation of cochlear implant solutions need to be flexible and reconfigurable to learn and adapt to both the users as well as the corresponding environment. Specific knowledge provided in this chapter can be leveraged to form new algorithm development and further applications in speech systems to benefit cochlear implant users in diverse naturalistic noisy environments.

3.2 UTD-CI-LENA Corpus Development

In order to investigate the influence of auditory feedback on speech production, a corpus was developed. This corpus included audio streams of cochlear implant participants from their daily lives. The LENA (for "Language ENvironment Analysis") device was used for collecting naturalistic audio from cochlear implant users (Gilkerson and Richards, 2008; LENA Foundation, 2014). The LENA device is a lightweight compact digital audio recorder that is capable of capturing mono audio data continuously for up to 16 hours. The device was worn by each subject, and captured the participant's daily acoustics, including voice communication and interaction with other people, as well as environments (*e.g.,* noise level and type). Figure 3.1(a) demonstrates how the device was positioned for collecting the audio data using the LENA unit. A cross pack was used to hold the device inside a pocket made of meshed-material for secure and consistent placement. The device was positioned at the

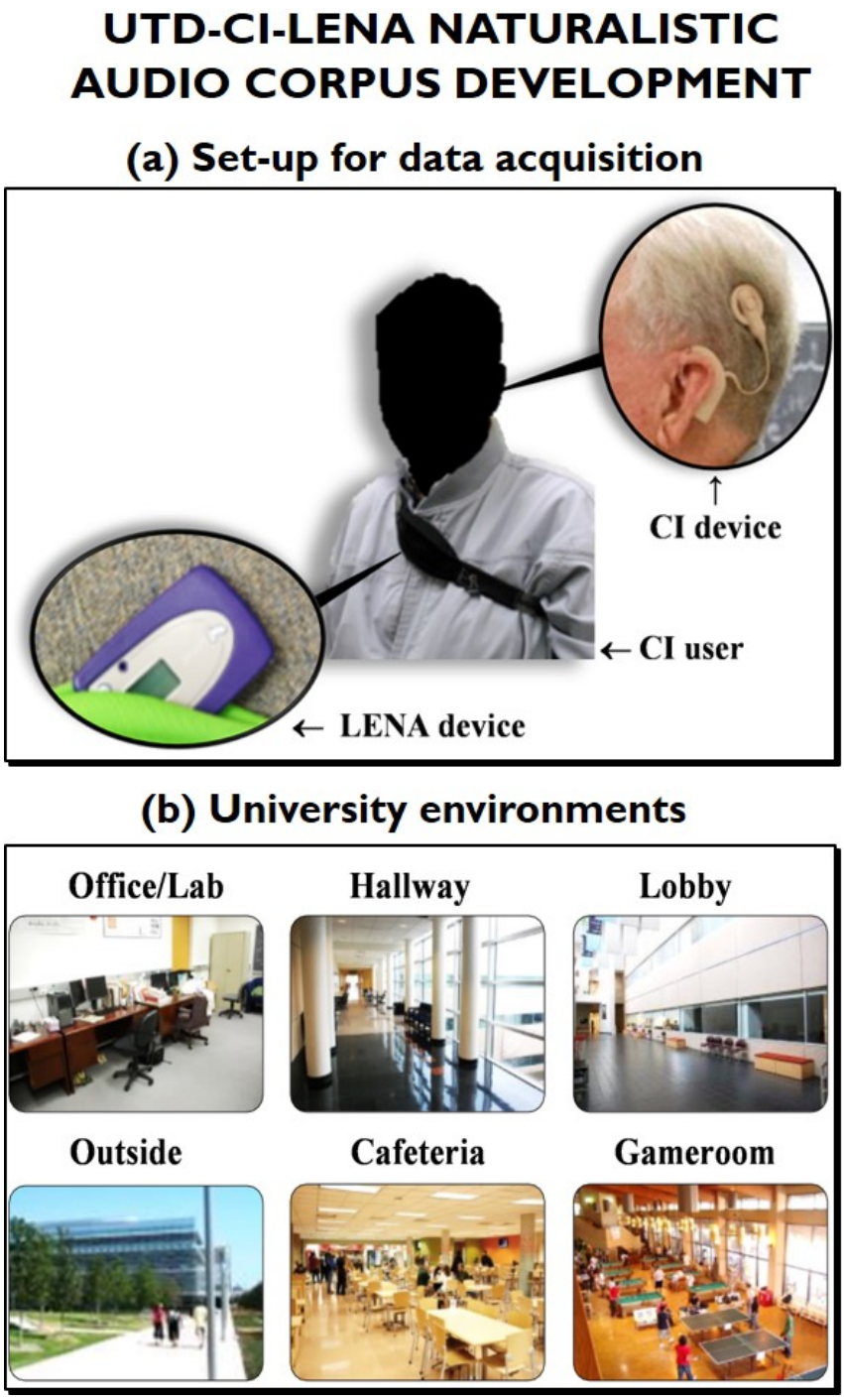


Figure 3.1. Naturalistic data collection for cochlear implant subjects: (a) set-up for data acquisition using the LENA unit, and (b) naturalistic environments on UT-Dallas college campus for data collection.

center of the chest where it was held stationary with respect to the subject's mouth (approximately 15 - 20 cm). Such a set-up made it possible for the unit's microphone to detect the acoustic signal more robustly and consistently (across subjects) against environmental noise/reverberation during data collection.

3.2.1 Subjects: CI-to-NH Pairs

A total of 6 cochlear implant users (mean age: 65 yrs.) participated by producing speech in various naturalistic environments. The cochlear implant speakers were fitted with the Nucleus devices from Cochlear Ltd. They were post-lingually deaf (lost hearing after the age of 18) and had been regularly using their cochlear implant devices for at least four years. Among the cochlear implant participants, five were bilateral, and one was a unilateral cochlear implant users. The unilateral cochlear implant user wore a hearing aid in the contralateral ear. It should be noted that bilateral listeners are expected to take advantage of head-shadow effect which offers improved sound localization versus a single implant (Hersbach et al., 2013; Wouters and Berghe, 2001). This in general, provides better speech perception in noise when the target signal is spatially separated from the masker. Detailed biographical information of the cochlear implant participants is presented in Table 3.1.

The same number of normal hearing speakers (mean age: 37 yrs.) participated in the study as a pair-wise conversational partner. The cochlear implant speakers in this research acted as the primary speaker, while the normal hearing listeners served as the secondary speaker/listener. Note that the objective of this study was to analyze the speech production of cochlear implant users in different noisy environments. Both cochlear implant and normal hearing subjects were native speakers of American English, and included an equal number of male and female participants in each group.

Table 3.1. Characteristic information of cochlear implant subjects who participated in UTD-CI-LENA corpus development.

Speaker	Gender	Age (yrs.)	Years of hearing loss	Years implanted	Etiology of hearing loss	Implant ear	Coding strategy
SPK 1	Female	61	56	11	Hereditary	Bilateral	ACE
SPK 2	Female	52	48	5	Hereditary	Bilateral	ACE
SPK 3	Female	61	14	4	Hereditary	Bilateral	ACE
SPK 4	Male	67	12	6	Hereditary	Left only	ACE
SPK 5	Male	81	55	9	Hereditary	Bilateral	ACE
SPK 6	Male	71	18	4	Unknown	Bilateral	ACE

3.2.2 Controlled On-Campus Environments

Naturalistic audio recordings were obtained in six environments on the UT-Dallas college campus. These included (i) office/lab, (ii) hallway, (iii) lobby, (iv) outside on campus, (v) college cafeteria, and (vi) college gameroom, as shown in Figure 3.1(b). The locations were chosen to provide a diverse range of noise conditions, for example, type, mixture, and the level varied greatly across environments. Here we present a summarization of the six naturalistic environments that were employed for data collection. This consists of (i) general room size, (ii) the number of people typically present during the day, (iii) average sound pressure level (SPL), (iv) degree of acoustic spectral stationarity, (v) reverberation time (RT_{60}) and (vi) room description. The "people" listing here refers to the typical number of subjects within that room/space. The "average SPL" was determined by calibrating the average noise intensity measured in Praat software (Boersma, 2002) and converted to dB SPL scale. For this, we have recorded an extra noise sound (*e.g.,* white Gaussian) with a known sound air pressure on the same recording device (*i.e.,* LENA). The "stationarity"refers to a relative

degree of stationarity for the background noise, with a 10-point scale (Hansen and Arslan, 1995), with 1/10 referring to stationary, and 10/10 referring to complete non-stationary (*i.e.,* multiple speaker babble noise would be 10/10 - see (Krishnamurthy and Hansen, 2009)). The "reverberation time (RT_{60})" refers to the length of time to decay by 60 dB from its initial level of impulsive excitation. The impulsive sound used in the calculation was created by a balloon burst recorded during off-hours in the respective locations:

- **Office/Lab**: 450 sq. ft. square room; 1-5 people; 1/10 stationarity; 43 dB SPL; RT_{60} = 0.3 sec.; An office/lab space where students and staff work; consists of desk with computers; office has windows; A/C vents present.

- **Hallway**: 2,250 sq. ft. long and narrow; 5-10 people; 3/10 stationarity; 52 dB SPL; RT_{60} = 0.6 sec.; A long passage in an engineering building; with doors into rooms/labs some open/some closed.

- **Lobby**: 11,500 sq. ft.; 5-50 people; 5/10 stationarity; 55 dB SPL; RT_{60} = 1.8 sec.; A high ceiling (3 story high) space near the entrance of an engineering building; corridors and staircases leading off it; classrooms lead into space; number of subjects much greater during time change between classes.

- **Outside**: Open-air; 30-50 people; 5/10 stationarity; 61 dB SPL; An outside location of UT-Dallas Richardson campus; surrounded by the main buildings and trees; sunny day, no rain.

- **Cafeteria**: 4,500 sq. ft.; 50-200 people; 9/10 stationarity; 65 dB SPL; RT_{60} = 0.8 sec.; A restaurant located in UT-Dallas student center; contains over 40 tables with 2-6 people/table; sounds from TVs and a music jukebox, discussion from students throughout.

- **Gameroom**: 9,500 sq. ft.; 30-75 people; 9/10 stationarity; 67 dB SPL; RT_{60} = 1.3 sec.; A public shared space in UT-Dallas student center; consists of numerous table tennis and billiard tables; music played from loud speakers, sounds from video arcade game machines, etc.

Sample time versus spectral estimate across 10 seconds of each noise data is shown in Figure 3.2. In this figure, the main noise sources were identified subjectively based on human transcribers listening to audio files. This shows an example of time varying nature of six naturalistic environments used in this study on UT-Dallas campus. Note the differences in their spectral content and spectral variations over time.

All audio recordings were collected when on campus population was expected to be consistent with daily conditions. This included data collection on weekdays (Monday - Friday) during normal working hours (10 am - 1 pm or 1 pm - 4 pm). In order to perform data collection, 3 minutes of background noise was first recorded prior to the subject's speech production in each location. None of the samples contain speech. These noise-only samples were used to assess subject's listening environments in subsequent data analyses (see Section 4.3.1). Following the background noise recording, subjects were asked to perform free conversation between each other for 5 minutes in each location. A list of topics were provided to participants as a suggestion before the test, which included general topics, such as sports, news, weather, movies, etc. The overall data collection period for each subject was about 2 hours each.

3.2.3 Uncontrolled Everyday Environments

In addition to the controlled environments, subsequent audio recordings were performed in cochlear implant user's real life. Cochlear implant users were asked to select one day, which was a normal workday, to provide 6 to 8 hours of additional audio recording. In this recording, no designated locations or conversation topics were given to subjects. However,

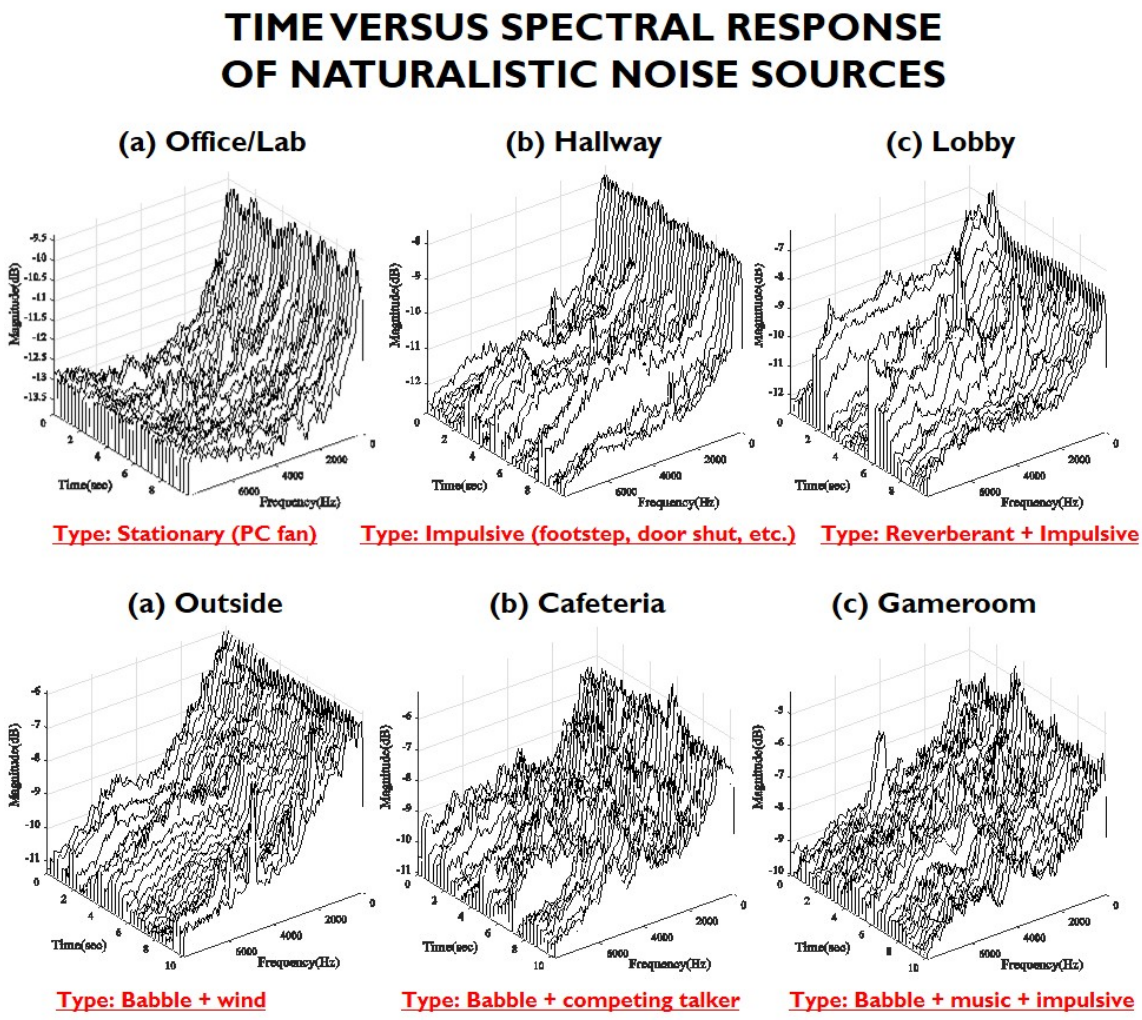


Figure 3.2. Spectral waterfall plots of six naturalistic environments used in this study on UT-Dallas campus. Sample time versus spectral estimate were computed using 10 seconds of each noise segments with a 5 seconds skip rate. The main noise sources were identified subjectively. Note the differences in their spectral content and spectral variations over time.

subjects were encouraged to select locations where the subject might have more interaction with other people so that diverse listening environments could be captured; for example, in quiet or noise, in places like their home, restaurants, stores, car, etc. All subjects were informed in both on- and off-campus sessions that they always have the option to pause the audio recording anytime if they might be in a situation where privacy or confidentiality

concerns arise during the recording session. However, no interruptions were experienced by any participants.

3.2.4 NH-to-NH Pairs for Comparison

For comparison of cochlear implant users versus normal hearing listeners, we repeated the same data collection with normal hearing speakers. Figure 3.3 demonstrates how the naturalistic audio data was collected with normal-hearing subjects. Six speakers who reported no history of speech or hearing related problems participated. The same number of normal hearing conversation partners participated in this collection as a secondary speaker. In order to minimize the age effect in speech production, we recruited age-matched subjects (mean age: 57 yrs.). They produced read and conversational speech in the same locations at UT-Dallas college campus. Following the on-campus environments, additional 6 to 8 hours audio recordings were performed in normal hearing speakers' daily life. In this way, we had a baseline data to compare with the cochlear implant users.

Figure 3.3. Naturalistic data collection with normal hearing listeners using the LENA unit. Naturalistic environments on UT-Dallas college campus for data collection are shown.

3.3 Post-processing

A set of acoustic and orthographic transcription labels were assigned to the collected audio data. The audio streams consisted of 2 acoustic categories, namely silence and spontaneous speech in each location. Labeling tasks were first performed by a human annotator based on events in that space. For example, sound events in the office space were different than outside in the public area. In order to produce orthographic transcription, every single isolated utterance (*e.g.,* sentence, phrase, word, and syllable) was first identified manually. Sentence level transcripts were then applied to each identified utterance based on listening to each individual audio file. Additional acoustic labels such as environments (office, hallway, outside, etc.), and speech types (silence, spontaneous) were applied manually to all recordings. An example of orthographic transcription labels was shown in Figure 3.4(a). In order to reduce inter-labeler variability, only a single annotator, a native speaker of American English, performed all data transcription collected from different speakers.

Following the manual labeling tasks, phoneme-level transcription labels were assigned. This task was done automatically by forced speech recognition alignment. Forced alignment is the process of taking the audio file and its orthographic transcription as input to produce word and phone boundary labels. Several other recent studies have successfully used forced alignment as a tool in phonetic research (Lu and Cooke, 2008, 2009; Yu et al., 2014). In this study, an open-source software P2FA was employed for this procedure (Yuan and Liberman, 2008). An example of word- and phoneme-level time-aligned labels was shown in Figure 3.4(b) and 3.4(c). Following the automatic alignment process, words and phonemes that were shorter than 40 msec in duration, and had fewer than 200 instances were excluded from analysis. From based on the efforts, a total of approximately 36,000 words including 38,000 vowel and 54,000 consonant nuclei were identified to be analyzed. It is important to note that due to the limited number of contexts, individual phoneme instances should not be regarded as prototypical.

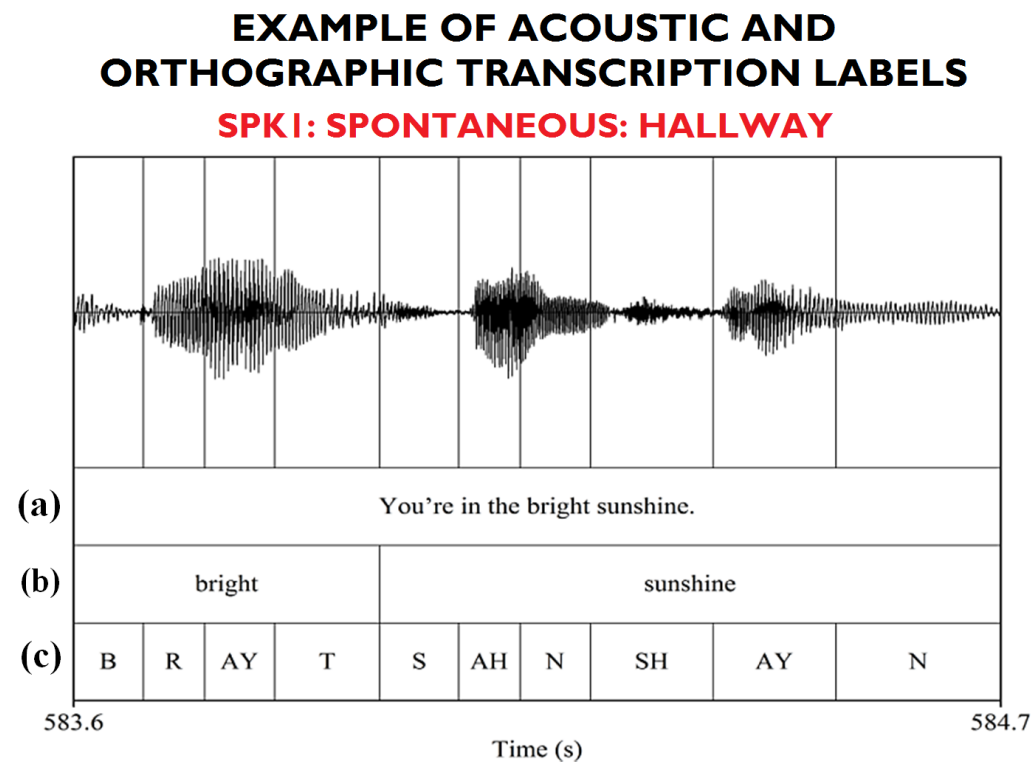


Figure 3.4. An example of transcription labels used in the speech analysis: (a) manual orthographic transcripts with corresponding waveform. Automatic (b) word-level, and (c) phoneme-level transcripts via forced alignment were shown with its boundary detection.

3.4 Summary: Chapter 3

This chapter provided a new speech corpus that can investigate the effect of auditory feedback on speech production in naturalistic daily environments. Mobile personal audio recordings from continuous single-session audio streams were used to observe this effect over an individual's daily life. Speakers produced read and conversational speech over various UT-Dallas on-campus environments (*e.g.,* office and cafeteria). Following the controlled locations, additional audio recordings were collected in uncontrolled environments (outside campus, *e.g.,* home and restaurants). A set of acoustic and orthographic transcription labels were assigned by a human transcriber. Phoneme-level transcription labels were also assigned automatically by forced alignment procedures. A total of 25 hours of personal audio recordings

were collected from 24 speakers (6 cochlear implant and 18 normal hearing speakers) while participated in college campus. Additional 75 hours of naturalistic audio were collected from subjects' everyday uncontrolled situations (*e.g.,* home, restaurant, store, etc.). This collection of speech production provides a unique and unprecedented opportunity to explore real-world listening and speech in diverse environments for CI-to-NH communications.

CHAPTER 4

EFFECT OF ENVIRONMENTAL NOISE ON SPEECH PRODUCTION OF COCHLEAR IMPLANT USERS: A NATURALISTIC STUDY

4.1 Introduction

Lombard effect has been shown to alter the normal behavior of human speech production and resulting speech feature characteristics (Lane and Tranel, 1971; Hansen, 1988; Junqua, 1996). Lombard speech typically involves a wide range of acoustic and articulatory adjustments, for example, altered vocal intensity, glottal spectral slope, fundamental frequency or formant location of speech (Garnier et al., 2010; Pisoni et al., 1985). Lombard effect helps to maintain speech intelligibility during human-to-human communication in challenging listening environments (Dreher and O'Neill, 1957; Summers et al., 1988; Lu and Cooke, 2009). Also, the variability introduced by a speaker under noise influences on performance in human-to-machine communication as well (Hansen, 1994; Bou-Ghazale and Hansen, 2000; Bořil and Hansen, 2010). Although well studied for normal hearing listeners as well as speech systems, Lombard effect has received little, if any, attention in the field of cochlear implant users.

A number of reports have established whether speech production of cochlear implant users is modified after the implantation. Some groups have performed to examine long-term longitudinal effects of auditory feedback (Bilger et al., 1977; Hochmair-Desoyer et al., 1981; Kirk and Edgerton, 1983), while some groups have investigated short-time effect on speech production (Svirsky and Tobey, 1991; Vick et al., 2001). These studies, as either acute or in chronic study conditions, established that speech production is heavily influenced by any form of auditory feedback. However, changes in speech production of cochlear implant users have been examined when only auditory feedback was artificially distorted (*e.g.,* turning processor on/off). These experimental disruption do not explain the nature of practical listening environments, such as noisy environments. To our knowledge, no study has examined

whether cochlear implant users employ Lombard effect during their voice communication after implantation.

4.1.1 Summary of Acoustic and Phonetic Features

The focus of this chapter was on finding acoustic features that characterize the Lombard effect of cochlear implant users. Before proceeding to investigate parametric changes in their speech production, it is important to identify which parameters are valid for Lombard effect. A discussion of Lombard effect in normal hearing subjects will be helpful in interpreting data from investigation of the speech production in cochlear implant users. The accumulated analyses have shown that naturally produced Lombard speech typically involves a wide range of acoustic and articulatory adjustment. While it would be impossible to include all the measurements, we can consider two broad categories of measurements, namely global and segmental measurements. The first category includes intensity, speaking rate, fundamental frequency, spectral tilt, and temporal envelop modulations. The second includes consonant-vowel ratio, vowel formant frequencies, phoneme duration, and vowel space. A summary of previous studies on acoustic and segmental adjustment is presented as follow.

The acoustic parameters of speech are likely to be influenced by limited auditory feedback due to noise. The average word and sentence intensities for Lombard speech were significantly increased from neutral speech (Hansen, 1988; Junqua, 1992; Lu and Cooke, 2008). Also, the average fundamental frequency and variance were significantly increased for Lombard effect speech (Summers et al., 1988; Pisoni et al., 1985; Garnier et al., 2010). The spectral energy for Lombard speech increased significantly, and there was more energy at higher frequencies (Sodersten et al., 2005; Hansen, 1988; Junqua, 1992). Lombard speech possessed more energy at higher frequencies than neutral speech. Acoustic analysis of Lombard speech revealed that a majority of talkers produced slower speech (Hanley and Steer, 1949).

An extensive analysis has been conducted on phoneme-level parameters. One feature that has consistently been shown in naturally produced Lombard speech is shift in formant frequency location (Pisoni et al., 1985; Hansen, 1988; Krause and Braida, 2004; Huber and Chandrasekaran, 2006). Lombard speech analyses have shown significant shift in first formant frequency location. No significant effect was found for second formant frequency location. Another finding in the phoneme-level measurement was consonant-vowel ratio. Normal hearing speaker emphasized intensity within consonant phoneme class with respect to vowel class (House et al., 1965; Hansen, 1988; Sodersten et al., 2005). Duration analysis indicated that average vowel duration increased for Lombard speech (Junqua, 1992; Lu and Cooke, 2008; Garnier et al., 2010). Little change, or a slight increase were observed for consonant duration under Lombard conditions.

4.1.2 Objectives and Proposed Methods

The objective of this study was to analyze and model the speech production of cochlear implant users with respect to environmental noise conditions. In addition, the study present to investigate the effect of auditory feedback on speech production in naturalistic daily environments. The analysis conducted in this chapter were based on a UTD-CI-LENA naturalistic speech corpus, which was described in Chapter 3. The collection of naturalistic data was performed in six realistic environments on UT-Dallas college campus (*e.g.,* office, lobby, cafeteria). Six cochlear implant users who were all post-lingually deafened adults participated and produced spontaneous speech in each environments.

Analysis of speech production was accomplished using: (i) characteristics of acoustic environments, (ii) evaluation of subject's listening environment, and (iii) acoustic and phonetic properties of speech production in relation to the listening environment. In the first analysis, various approaches were used to characterize real-world environments, including long-term averaged spectra, modulation spectra, and noise sound pressure level (SPL). The second

part of the analysis focused on objective metrics to predict speech quality, namely signal-to-noise ratios with and without Lombard effect. Lastly, the parametric variations in vowel, consonant and individual phoneme production were investigated as a function of varying environments. This involved speech SPL, fundamental frequency, glottal spectral tilt, phoneme duration, and formant frequencies. The analyses outlined here will explore relationships of speech production parameters of cochlear implant users upon varying noise/environment types.

4.2 Methods

4.2.1 Database Used

The database used for the analysis in this chapter were from a part of the UTD-CI-LENA naturalistic speech corpus described in Chapter 3. We have chosen a total of 25 hours of naturalistic audio recordings from the dataset, that were obtained in six environments on the UT-Dallas college campus. Table 4.1 summarizes the six naturalistic environments employed for data collection. These audio included 13 hours of cochlear implant-to-normal hearing (CI-to-NH) pairs as well as 12 hours of normal hearing-to-normal hearing (NH-to-NH) pairs. The CI-to-NH dataset was used to analyze the speech production of cochlear implant users, while NH-to-NH pair audio was used for pairwise comparison between cochlear implant user group and normal hearing listener group. Detailed biographical information of the cochlear implant participants as well as procedures are presented in Table 3.1 and Section 3.2.2 respectively.

4.2.2 Signal Processing: Features and Metrics

Acoustic characteristics of the background noise were considered by investigating the (i) long-term average spectrum, (ii) average modulation spectrum, (iii) noise SPL, (iv) spectral centroid, and (v) average modulation spectrum energy. It is important to mention here that

the noise analysis was carried out on "noise-alone audio segments" which were collected in each environment prior to conversation. None of these samples contained any speech. The long-term average spectrum was obtained by averaging short-time power spectral estimated by the Welch's method. The noise SPL was determined by calibrating the average noise intensity measured in Praat software (Boersma, 2002) and converted to dB SPL scale. For this, we have recorded an extra noise sound (*e.g.,* white Gaussian) with a known sound air pressure on the same recording device (*i.e.,* LENA). The spectral centroid was calculated based on the average frequency weighted by amplitudes, divided by the sum of the amplitudes. The duration of the analysis window was set at 100 ms with a 50 % overlap for each measurement. While average noise SPL shows change in signal strength over time-domain, spectral centroid represents where spectral energy was concentrated in frequency-domain.

We calculated the average modulation spectra to obtain a better understanding of overall room acoustics. The modulation spectrum represents the slowly varying temporal envelope components of signal, thereby providing a degree of acoustic spectral stationarity. Noise

Table 4.1. Summary of naturalistic environments used in this study on UT-Dallas campus. The stationarity was identified subjectively based on listening to audio file (1: wide-sense stationary to 10: non-stationary).

Location	**Room size (sq. ft.)**	**Number of people**	**Stationarity(1-10)**	**Avg. SPL (dB)**	**Room description**
Office	450	1-5	1	45	An office/lab space with PCs
Hallway	2,250	5-10	3	54	A long passage with windows
Lobby	11,500	5-50	5	60	A 3-story high ceiling space
Outside	Open-air	30-50	5	64	An on-campus outside location
Cafeteria	4,500	50-200	9	70	A restaurant location with tables
Gameroom	9,500	30-75	9	74	A public space with billiard tables

samples from each environment were divided into 2-second segments with 1 second time interval. The modulation spectrum was computed by taking Fourier transform of the Hilbert envelope for each segments. The 0 - 20 Hz components were then added together across all segments. Finally, the average between 0 - 20 Hz was computed and considered as the average modulation spectrum energy.

Following the noise characteristics, the individual's listening environments were evaluated by estimating signal-to-noise ratios. In this study, two signal-to-noise ratios approaches were employed, which include signal-to-noise ratio (i) with neutral speech (SNRN), and (ii) with Lombard speech (SNRL). In these measurements, the assumption is that the *office environment* is a quiet baseline, and speech produced in this location will be *neutral*. SNRN was defined as the energy ratio of neutral speech to noise energy in each environment, which is assumed to be without the Lombard effect as follows:

$$SNRN = 10 \cdot log_{10} \left(\frac{E_{Neutral}}{E_{noise}} \right), \tag{4.1}$$

and SNRL was calculated from the energy ratio of Lombard speech to the corresponding background noise for each environment as follows:

$$SNRL = 10 \cdot log_{10} \left(\frac{E_{Lombard}}{E_{noise}} \right), \tag{4.2}$$

where E is the average energy. For these calculations, acoustic boundary detection which was marked by hand was employed for separating speech from background noise. The leading and trailing silent intervals derived from each audio stream served as noise samples in each location for computing signal-to-noise ratios. These metrics shows the following two observations: (i) there is a the change in signal-to-noise ratio reduction due to the effect of noise, and (ii) the level to which the decreased signal-to-noise ratios recover by the presence of Lombard speech is clear and measurable.

In addition to environmental characteristics, various acoustic and acoustic-phonetic parameters for speech production were analyzed. These include (i) average speech SPL, (ii) fundamental frequency (F0), (iii) overall spectral tilt, (iv) phoneme duration, (v) first formant frequency (F1) location, and (vi) second formant frequency (F2) location. All measurements except phoneme duration and overall spectral tilt were computed using PRAAT software (Boersma, 2002). The average speech SPL measurements used here was similar to the metrics employed in the noise analysis, so to ensure connected scales between the measurements when reporting power. Phoneme duration was obtained from analysis of the phoneme-level transcripts. Spectral tilt was calculated from the difference between the magnitudes of the first spectral harmonic (H1) and that of the third formant peak (A3), *i.e.,* H1-A3 (Hanson, 1997; Iseli et al., 2007) via PRAAT capability. It should be noted here that, the focus was to compute an overall spectral slope. A related study by Hansen (Hansen, 1988) demonstrated changes in glottal spectral slope for various types of speech under stress by averaging individual frame spectral slopes of voiced speech over multiple utterances. Since the focus here is on overall speech content, that approach was not employed.

Each acoustic-phonetic features were extracted based on phoneme nuclei boundaries marked by a forced phoneme alignment process in Section 3.3. The beginning and ending markers of each phoneme were reduced by 20% to eliminate any transitional effects across phoneme classes. The duration of the analysis window used here for both speech and noise measurements was set to 20 ms with a 10 ms skip rate. After the feature extraction, normalization procedures were applied to reduce *speaker-particular* effects from the data (*e.g.,* baseline F0 differences across male and female talkers). This was achieved by scaling each parameter to have the same overall level across speakers.

Lastly, a repeated-measures analysis of variance (ANOVA) was performed to assess the effect of environment type on noise/speech parameters, and determine statistical significance of differences between speech produced in neutral and Lombard conditions. Subjects were

considered as random (blocked) factors, while environment conditions were used as the main analysis factors. Following the ANOVA, a *post-hoc* pairwise comparison test was performed to determine if the noisy conditions were significantly different from the quiet baseline. Bonferrroni adjustment was used to control for family-wise error in the pairwise test. In this study, a difference in means between two or more groups was considered significant if the significance level fell below 5.0% ($p < 0.05$).

4.3 Results

4.3.1 Noise and Environment Analysis

Prior to any analysis, it is important to understand the characteristics of the acoustic and listening environments. This section offers some level of baseline knowledge regarding each environment's acoustic characteristics as well as how cochlear implant users perceive their speech in these particular noisy environments. This may relate to speech perception by cochlear implant users in these environments.

Noise Characteristics

Fig. 4.1(a) shows the long-term averaged spectra of various maskers. It can be seen that the office environment has the least spectral impact in terms of overall spectral energy as compared to other noisy environments, which is why it was chosen as a baseline in this study. Spectral energies in general were highly concentrated in the lower frequency range (< 2 kHz) for all environments. When compared to the office baseline, a progressive increase in spectral energies (from hallway to gameroom) was observed in all noisy environments based on the increasing complexity of acoustic space.

This can be better visualized from Fig. 4.2(a) and 4.2(b) which present the distribution of average noise SPL and spectral centroid for each environment. The results indicate that

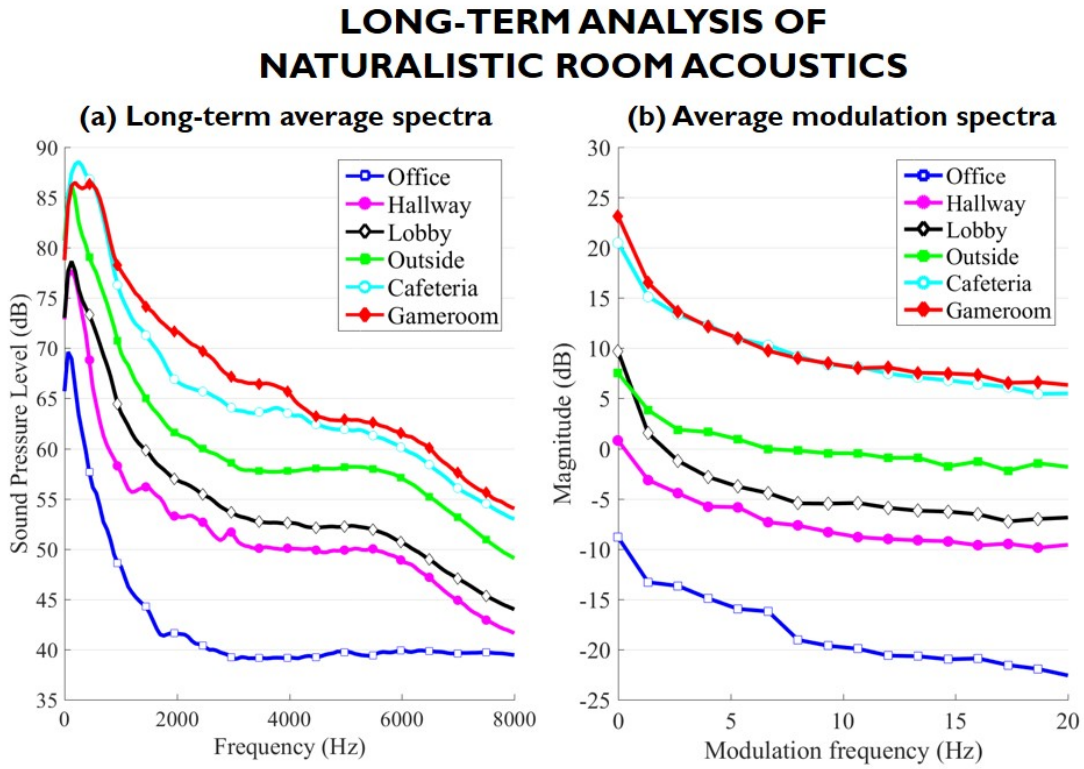


Figure 4.1. *Long-term analysis for six naturalistic environments used in this study on UT-Dallas campus: (a) long-term average spectra, and (b) average modulation spectra. Each line corresponds to the average long-term features for each naturalistic environments. The main noise sources were recorded prior to subject's speech production for 3 minutes in each environment.*

both features increased monotonically when switched from office to gameroom. The range of average noise SPL extended from approximately 42 dB (for office) to 67 dB (for gameroom), and all noisy environments had mean values which were significantly different from the office baseline ($p < 0.05$). Spectral centroid was almost always under 500 Hz for all conditions. With the exception of hallway condition, all noisy environments had mean SPL and spectral centroid values significantly different from the office baseline ($p < 0.05$). Hallway remained almost constant in terms of mean of the spectral centroid ($p > 0.05$).

Fig. 4.1(b) shows the change of the average modulation spectrum as a function of environments. The modulation spectrum energy between the modulation frequencies of 0 - 20

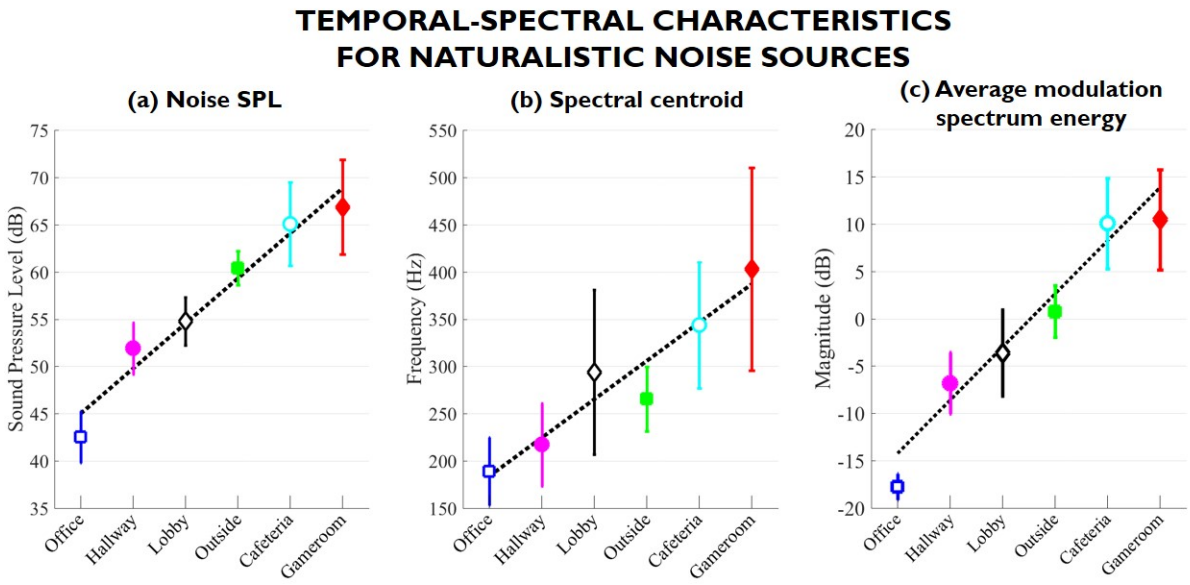


Figure 4.2. *Acoustic characteristics of background noises: average (a) noise SPL, (b) spectral centroid, and (c) average modulation spectrum energy with respect to different environments. While average SPL shows changes in signal strength over time-domain, spectral centroid represents where spectral energy was concentrated in frequency-domain. Average modulation spectrum energy estimate the relative degree of stationarity for the noise signal.*

Hz are presented. All the noise signal has a distinct modulation spectrum with a peak at 0 Hz. All the conditions, with the exception of office environment, had a similar modulation pattern. The modulation spectrum energy between 2 and 20 Hz increased when complexity of noise increased (*i.e.,* with an increase in the number of people, noise SPL). This effect can be better visualized from Fig. 4.2(c), which shows the average modulation spectrum energy of the tested environments. From this figure, we confirmed the increasing tendency in an explicit way. This suggest that such modulation analysis provides a useful measure of stationarity.

Signal-to-Noise Ratios

Fig. 4.3 illustrates the average SNR levels with respect to each environment. For each environment type, the bar on the left (hatched) indicates average SNR without the Lombard ef-

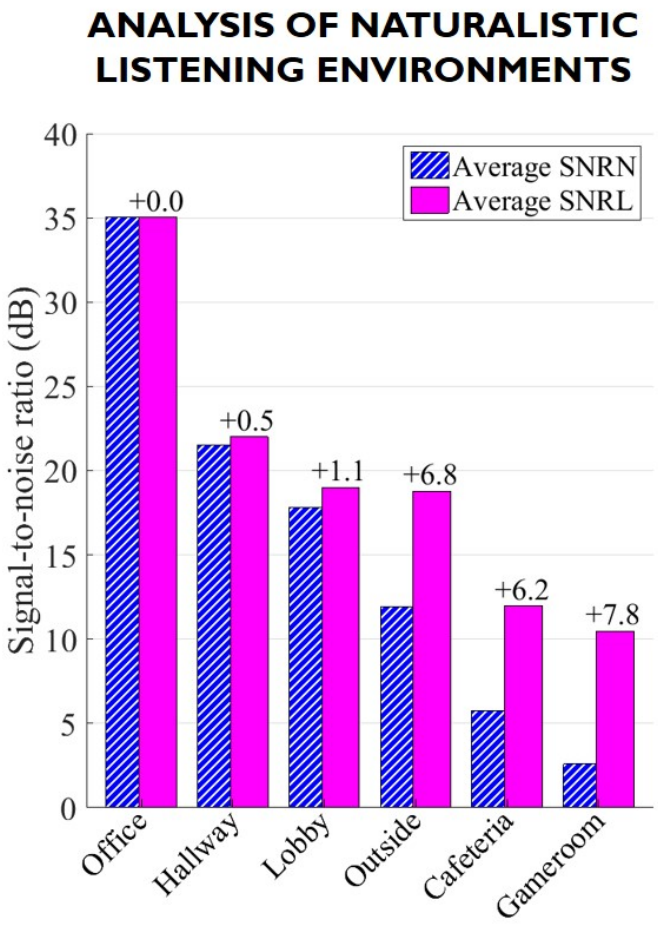


Figure 4.3. Evaluation of subject's listening environment using average signal-to-noise ratios as a function of varying environments. Two bars on the left (hatched) and right (full-colored) for each condition correspond to the SNR with Neutral speech (SNRN) and with Lombard speech (SNRL) respectively.

fect (SNRN), and the bar on the right (full-colored) indicates average SNR with the Lombard effect (SNRL). With an increase in average SPL of noise at various environments, average SNRN decreased (left-side bars) (baseline office = 34 dB to gameroom = 3 dB), indicating deterioration in speech signal intensity/quality. On the other hand, deterioration in SNRL was not as steep as SNRN (right-side bars) indicating that CI users tend to speak louder in noisy environments to effectively improve the SNR of their speech. This phenomenon is a key characteristic of the Lombard effect. For example, consider the gameroom environment

where SNRN was 3 dB. The Lombard effect here helped to increase overall SNR up to 11 dB, thus the corresponding benefit to include the Lombard effect in this environment (gameroom) was + 7.9 dB. The Lombard effect demonstrated here could boost the perceived SNR levels, and thereby facilitate in auditory decoding for the two-way conversations of CI users in noisy conditions.

4.3.2 Speech Production Analysis: CI Users

In this section, we consider methods for analyzing speech production characteristics of conversation as a function of varying environment. Note that in this section, we again established the *office environment* as the quiet baseline (< 45 dB SPL), assuming speech production in this location to be *neutral*.

Vowel and Consonant Analysis

For the following analyses, six acoustic parameters were investigated: (i) vowel SPL, (ii) fundamental frequency F0, (iii) glottal spectral tilt, (iv) vowel duration, (v) consonant SPL, and (vi) consonant duration. While the trajectory of vowel and consonant SPL indicate the variation in signal strength over time, pitch and spectral tilt deliver the pattern of shift in spectral cue/energy in the frequency-domain. Here, phoneme duration represents the temporal aspect of the vowel and consonant production.

Fig. 4.4(a) and 4.4(b) show the average vowel SPL and F0 as a function of varying noise SPL respectively. The data presented here were averaged across all subjects and all vowel phonemes. The noise SPL presented here were shown/derived from in Figure 4.2(a). Note that the asterisk marked at respective data points indicate statistical significance as compared to the office baseline ($p < 0.05$). The results indicate that both vowel SPL and F0 varied across conditions. Average values of both features increased significantly for outside, cafeteria, and gameroom environments ($p < 0.05$). However, only little changes occurred in

vowel SPL and F0 for hallway and lobby environments ($p > 0.05$). Two groups emerged for both features: a high-value group which was significantly different from office baseline ($p < 0.05$) and a low-value group with no statistically significant difference from baseline ($p > 0.05$). The low value group comprised of the hallway and lobby conditions, while the high value group included outside, cafeteria and gameroom environments. No significant differences were found between the conditions which belonged to the same group.

Fig. 4.4(c) and 4.4(d) present the variation of spectral tilt and vowel duration respectively with respect to each environment. Spectral tilt was found to progressively reduce with increasing noise SPL. From the baseline office to the gameroom environment, the mean spectral tilt fell from 19 dB to 14 dB; with other environments falling within this range. A significant effect of environment type on spectral tilt was observed between gameroom and office baseline ($p < 0.05$). However, no statistical significant differences were found for any combinations of the remaining four conditions (hallway, lobby, outside, cafeteria) ($p > 0.05$). For vowel duration, variations in mean were found in the presence of noise. As shown in Fig. 4.4(d), average vowel duration decreased progressively with noise complexity; however, it was only significantly different from the baseline condition for the gameroom environment ($p < 0.05$) only. Hallway, lobby, outside, and cafeteria environments did not result in a statistically significant change in the vowel duration as compared to the office baseline.

Fig. 4.5(a) and 4.5(b) display the variation in consonant SPL and duration respectively with respect to noise SPL. In general, both features were altered by speakers under noisy environments. For consonant SPL, the results followed the similar pattern to vowel SPL. Two distinct groups were found, the high value (outside, cafeteria, and gameroom) and low value group (hallway and lobby). The high value group increased significantly from the office baseline ($p < 0.05$), while the low value group resulted in little or no change across both environments ($p > 0.05$). No significant pairwise differences were found between the conditions which belonged to the same group. For consonant duration, mean values

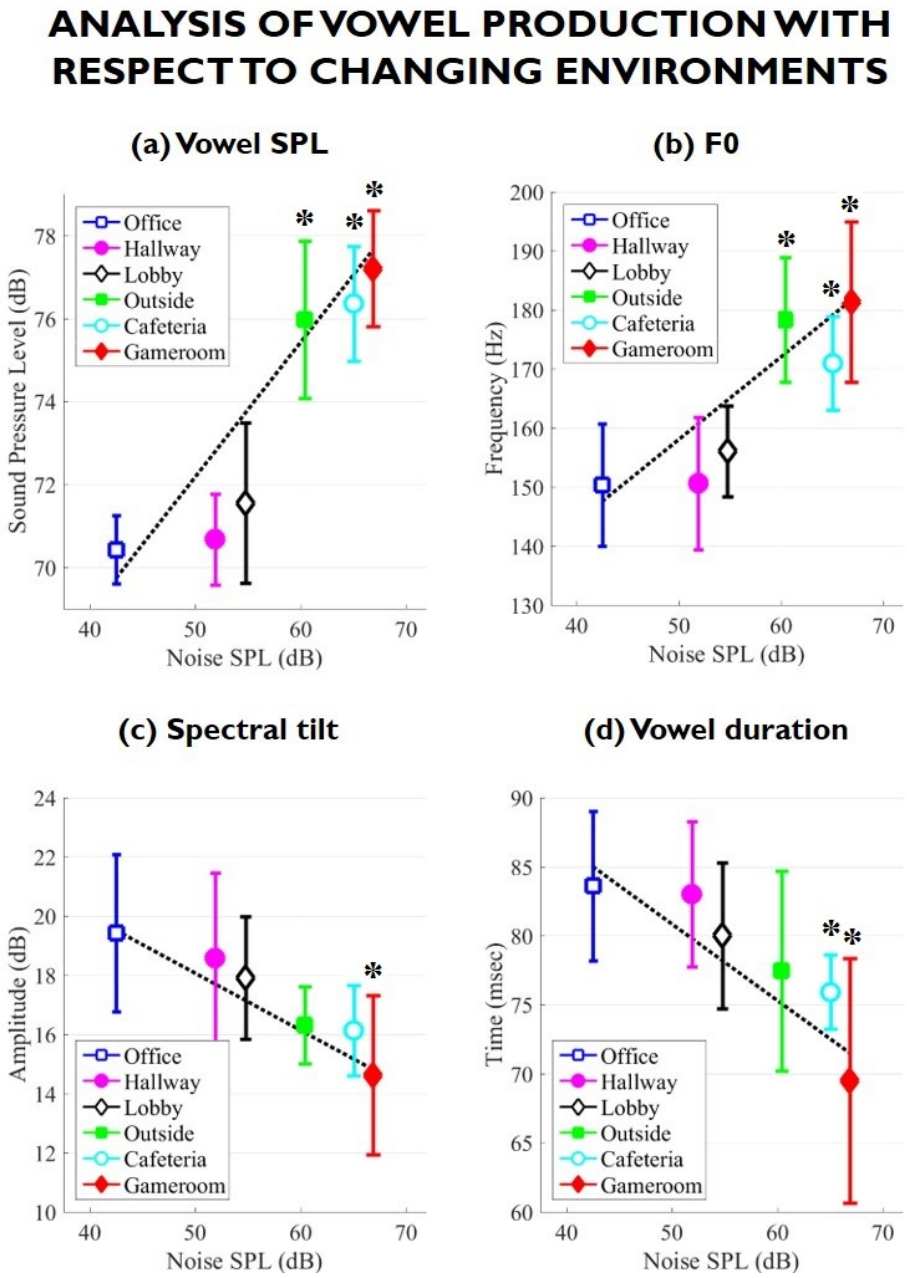


Figure 4.4. Acoustic analysis of vowel productions: individual variations of (a) vowel SPL, (b) fundamental frequency (F0), (c) spectral tilt, and (d) vowel duration, as a function of noise SPLs. Asterisk indicates statistical significance ($p < 0.05$) from natural speech.

monotonically decreased for most noisy conditions. However, there were no statistically significant difference in mean between quiet and all noisy conditions ($p > 0.05$). Speech

produced in all noise environments had slightly shorter consonant duration than that of the baseline office environment.

Vowel-Consonant Ratios

Additional analyses were conducted on global shifts in acoustic features between individual phoneme classes. It has been previously demonstrated with normal hearing listeners that a talker could maintain overall intensity, yet emphasize consonant phoneme class with respect to vowel class (House et al., 1965; Hansen, 1988). Hansen (Hansen, 1988, 1996) also suggested that consonant duration increases at the expense of vowel duration in an effort to increase speech intelligibility under noise. In the present analysis, two different ratios were considered: (i) a consonant versus vowel intensity ratio (CVIR), and (ii) a consonant versus

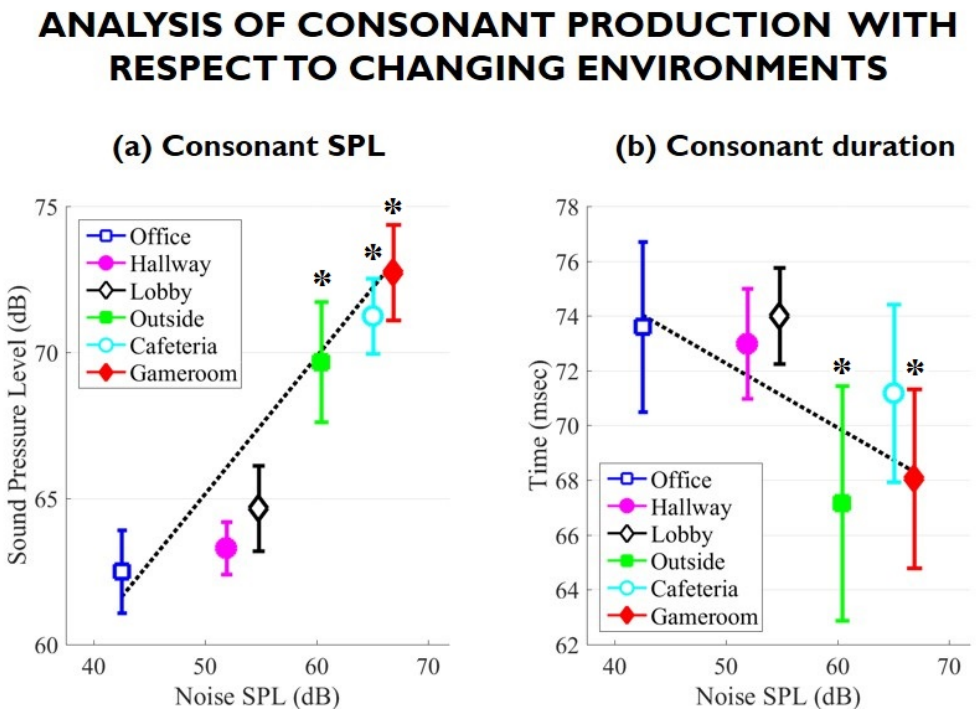


Figure 4.5. Acoustic analysis of consonant productions: individual variations of (a) consonant SPL, and (b) consonant duration, as a function of varying noise SPL. Asterisk indicates statistical significance ($p < 0.05$) from natural speech.

vowel duration ratio (CVDR) (House et al., 1965; Hansen, 1996). Vowel and consonant intensities were computed using PRAAT software (Boersma, 2002), and phoneme duration was obtained from analysis of the phoneme-level transcripts. These ratios indicate how energy or duration between vowel and consonant speech classes changes under noisy conditions.

A pictorial representation of global shifts between individual phoneme classes is presented in Figure 4.6(a) and 4.6(b). For each figure, shaded regions within each bar graph indicate average intensity/duration values for vowel and consonant phoneme classes. The asterisk indicates statistically significant shifts in mean based on the measures of CVIR and CVDR. Consider the CVIR first, where increased CVIRs resulted for most noisy conditions.

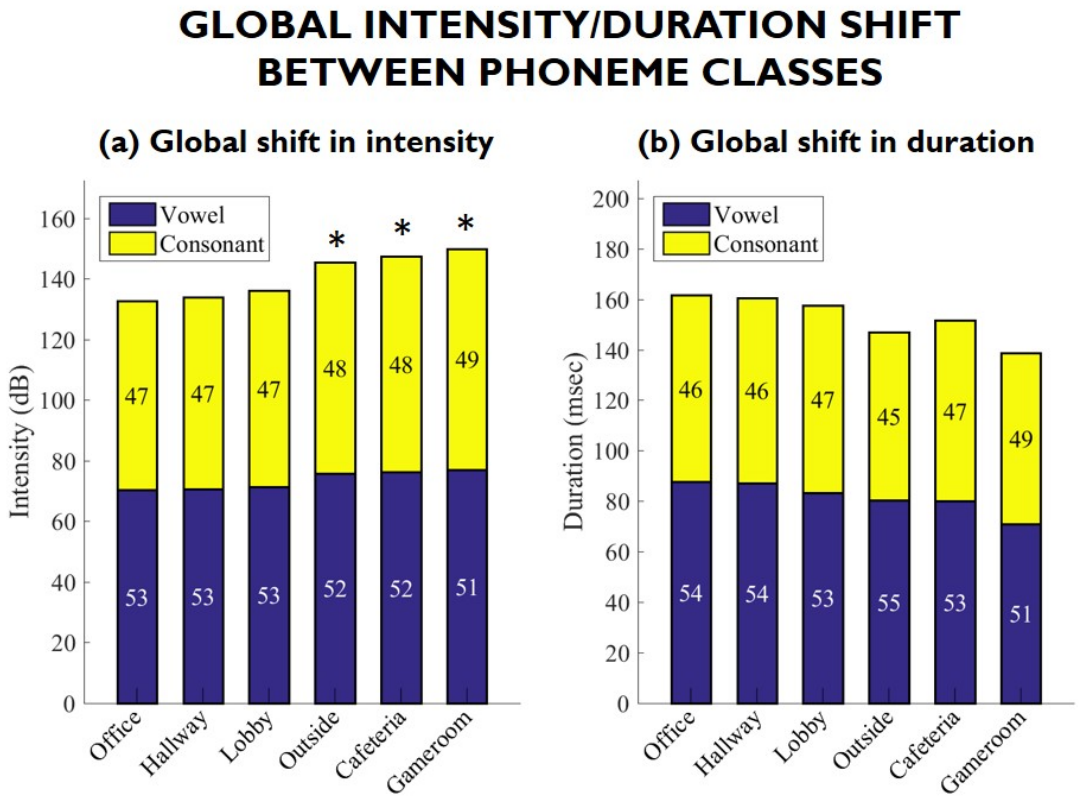


Figure 4.6. Pictorial representations of global shift in (a) intensity, and (b) duration, between vowel and consonant phoneme class. The speech class percentage is shown for each environment. Asterisk indicate significant shift in intensity/duration based on phoneme ratios.

The outside, cafeteria and gameroom conditions were significantly different from the office environment ($p < 0.05$). These particular changes in CVIR demonstrate increased consonant intensity as compared to vowel intensity in noisy condition. No significant differences were found for hallway and lobby as compared to office location ($p > 0.05$). For CVDR, no statistically significant shift in duration between vowel and consonant phoneme classes was observed in any locations ($p > 0.05$). It should be noted that consonant duration with respect to vowel duration plays a crucial factor in listener's ability to perceive the speech in the presence of noise (Hansen, 1988, 1996).

Vocal Tract Characteristics

Thus far, the investigation of the Lombard effect has focused on analysis of source excitation parameters. Power, pitch, spectral tilt and duration are all controlled in some manner by the supra-glottal and glottal systems. It is reasonable to hypothesize that noise also affects the articulators that configure vocal tract shape. In order to investigate vocal tract shape, we considered only the vowel space, since it controls the analysis to fixed articulator positioning versus the more complex time varying requirements for liquids, glides, and diphthongs. For statistical reliability, phonemes with sufficiently large sample sets were considered. Four cardinal vowel nuclei, /a/, /æ/, /i/, and /u/ were chosen for this analysis. Phoneme-level transcription labels identified boundary information of each phoneme. On average, more than 3,000 instances of each phoneme were employed.

Figure 4.7 illustrates the vowel space for phoneme nuclei, /a/, /æ/, /i/, and /u/, under various noisy conditions. Figure 4.7(a) includes hallway and lobby, while Figure 4.7(b) contains outside and cafeteria. Gameroom condition is shown in Figure 4.7(c). Office baseline results are also provided in each figure for comparison. The abscissa in each figure denotes F1 formant locations, while the ordinate denotes F2 formant locations. Overall, F1 formant locations changed with respect to environment, but F2 formant location did not. Consider

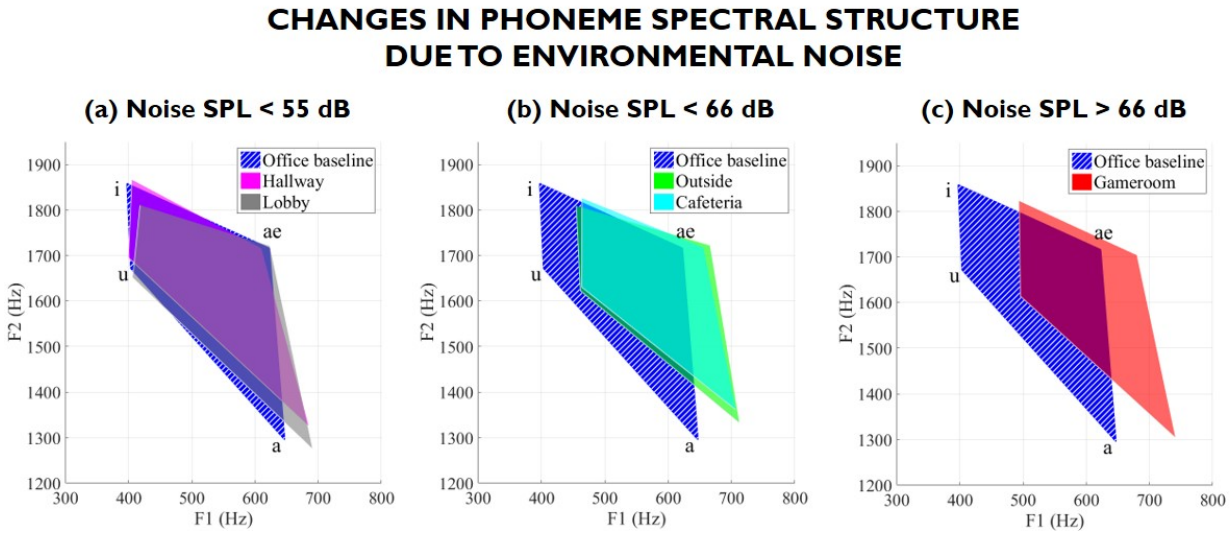


Figure 4.7. Spectral characteristics of vocal-tract: plots of first formant frequency F1 versus second formant frequency F2 for vowel phonemes /a/, /æ/, /i/, and /u/ with respect to (a) hallway and lobby, (b) outside and cafeteria, and (c) gameroom environments. Office result was given in each plot for comparison.

the front vowel /i/ first. F1 formant frequencies for /i/ phoneme significantly increased for outside, cafeteria and gameroom speech ($p < 0.05$). The pattern of the results was consistent with many other acoustic parameters (*e.g.,* vowel/consonant SPL and F0). However, this was not the case for the F2 parameter. For the phoneme /i/, no significant changes occurred for F2 formant locations across all environments for all speakers ($p > 0.05$).

Similar changes were observed for the remaining three vowel phonemes, /a/, /æ/, and /u/. In these phonemes, shifts in formant locations generally followed the trend observed for the /i/ phoneme. For example, significant increases in F1 formant location for /u/ were found for environments: outside, cafeteria and gameroom conditions versus baseline office condition ($p < 0.05$). F1 formant location for /a/, and /æ/ phonemes were changed significantly for gameroom condition ($p < 0.05$). However, similar to the /i/ phoneme, F2 formant location also resulted in relatively small changes for /a/, /æ/, and /u/ phonemes across all environments ($p > 0.05$). Thus, F2 locations were not a major factor for changes in speech production in noisy environments. It should be noted that F2 formant frequency

plays a critical role in speech comprehension of normal hearing as well as cochlear implant users in noisy environments (Loizou, 2013).

Summary of Lombard Speech Parameters

In this study, analysis of speech production in six naturalistic environments was presented along with the Lombard effect observed in speech produced by CI users. Statistical analysis techniques were employed to determine if any changes in speech features are reliable Lombard relayers. The results indicate that many speech features were used by the CI users in demonstrating Lombard effect stress condition. However, due to limited dataset and environments, it may be difficult to identify which specific parameters are sufficiently sensitive, and statistically reliable indicators of the Lombard effect. In order to identify these features, we grouped similar listening conditions together. Here, the environmental conditions were grouped into two areas: low and high noise groups. Low noise group includes office, hallway, and lobby, and high noise group includes outside, cafeteria, and gameroom. The decision was made based on average noise SPL (Figure 4.2(a)) and average speech SPLs (and 4.4(a) and 4.5(a)). Next, ANOVA was performed on the two groups. Therefore, this section allowed for a manageable summary of important Lombard relaying parameters from naturalistic audio recordings by fixing two conditions.

The analysis indicates that energy and fundamental frequency (F0) were the primary stress relayers for varying vocal effort. Vowel ($F[1, 40] = 125.76, p < 0.001$) and consonant ($F[1, 40] = 116.43, p < 0.001$) classes resulted in an increased SPL as vocal effort increased. In addition to individual phoneme classes, the phonetic class intensity ratio (CVIR) identified a strong shift in energy from vowels toward consonants in Lombard speech ($F[1, 40] = 52.63, p < 0.001$). Thus, CVIR can be used as a reliable indicator for speech under noise. The average F0 value was found to be statistically different from natural styles ($F[1, 40] = 50.00, p < 0.001$). Spectral tilt was also found to be a reliable perturbation of Lombard

speech ($F[1, 40] = 15.37, p < 0.001$). Vowel ($F[1, 40] = 14.16, p < 0.001$) and consonant ($F[1, 40] = 9.00, p < 0.005$) durations were found to be valid discriminating parameters for Lombard speech as well. However, duration ratio between vowel and consonant (CVDR) could not be considered as a discriminating Lombard indicator ($F[1, 40] = 1.99, p > 0.05$) at least on the environments tested. CVDR indicated only a slight shift in vowel duration towards an increase in consonant duration which was not statistically significant. Average first formant location F1 for the selected vowels /a/, /æ/, /i/, and /u/ showed statistically significant shifts from neutral speech ($F[1, 40] > 12.36, p < 0.001$). F2 formant location for /a/, and /u/ phonemes were significantly different from neutral speech as well ($F[1, 40] > 4.33, p < 0.05$). However, other phonetic areas such as second formant locations F2 for /i/, and /æ/ phoneme were not significantly different ($F[1, 40] < 1.42, p > 0.05$).

4.3.3 Pairwise Comparison of CI versus NH

The experiments described in the previous section have shown clear acoustic-phonetic variations in speech of cochlear implant users when produced in noisy environments. While many speech production parameters were found to show Lombard effect, it is still unknown how these are different from normal hearing listeners in noisy conditions. Thus, in this section, we investigated whether the nature of the extent of these shifts differ between the two speaker groups: cochlear implant users and normal hearing subjects. For this purpose, we employed the naturalistic data from NH-to-NH pairs, so we have a baseline to compare with CI-to-NH pairs. We measured the same speech parameters used in the cochlear implant users' analysis. An ANOVA test was performed to determine statistical difference between the two speaker groups.

Figure 4.8 shows pairwise comparison of cochlear implant users versus normal hearing listeners. The asterisk marked for each speech parameters indicates a statistical significance when compared to the normal hearing speech baseline ($***$:< 0.001, $**$:< 0.01, $*$:< 0.05).

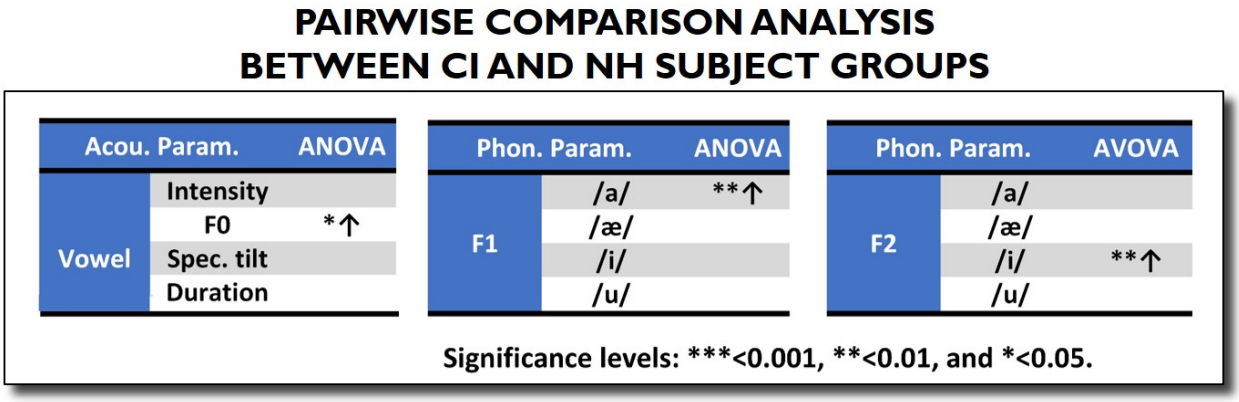


Figure 4.8. The nature of Lombard effect parameters differ between the two speaker groups, cochlear implant users and normal hearing listener groups. The asterisk marked for each speech parameters indicates a statistical significance when compared to the normal hearing speech baseline. A speech parameter without an asterisk means there is no significant difference between the two speaker groups.

A speech parameter without an asterisk means there is no significant difference between the two speaker groups. The results indicate that most speech production parameters for cochlear implant group have similar pattern changes with normal hearing groups. However, some parameters, *e.g.,* F0, F1 for /a/, and F2 for /i/, were significantly different from normal hearing listeners. These difference may due to the partial restoration of auditory feedback by cochlear implant devices. It has been known that reduced temporal fine structure provided by cochlear implants results in poorer performance for pitch perception (Moore, 2008). Previous studies in cochlear implants have also shown that the absence of fine spectral structure may contribute to formant perception (Fu et al., 1998).

4.4 Discussion

Many speech production parameters identified in this study were observed to change with limited auditory feedback due to acoustic noise. It was found that postlingually deafened cochlear implant users modify both segmental and suprasegmental properties of their speech in different listening environments. These parameters serves to calibrate speech production,

(*i.e.,* the speaker monitors their relations between his/her own phonemic intention and their acoustic output in the presence of noise). Moreover, it also influences speech production along an instantaneous basis, thus speakers modulate at least some suprasegmental features of their ongoing speech gesture. The results from this study indicate that cochlear implant users exhibit the Lombard effect in adverse listening environments.

Many investigators have suggested that auditory information for normal hearing listeners may be used to modulate at least some suprasegmental parameters of speech production under noise (Pisoni et al., 1985; Hansen, 1988; Summers et al., 1988; Junqua, 1992; Hansen, 1996; Lu and Cooke, 2008; Garnier et al., 2010). In these studies, results showed an increase in overall amplitude of vocalic sections, increased duration, increased average F0, and a decreased spectral tilt. This was presumed to be the result of increased subglottal pressure, and vocal-fold tension as a response to the reduced auditory feedback due to noise. Existing studies suggest that the most widely considered area of the Lombard effect involves vocal intensity and F0 (Hansen, 1996; Lu and Cooke, 2009; Garnier et al., 2010). Spectral balance of vowels was affected by the higher vocal effort for Lombard speech, resulting in relatively greater intensity in the higher frequency bands of the spectrum (Hansen, 1988; Junqua, 1992; Lu and Cooke, 2008).

Interestingly, the results showed that vowel and consonant duration (Figure 4.4(d) and 4.5(b)) decreased in most noisy environments in contrast to the earlier studies. It is well established in normal hearing research that Lombard speech has a generally increased phoneme duration in comparison to normal speech, and speech intelligibility in noise is associated with lengthening the phoneme duration (Junqua, 1996; Lu and Cooke, 2008; Garnier et al., 2010). We suggest that the main contribution to this difference is the conversational speaking style which was used in the data analysis. Participants in this study produced solely conversational speech in realistic scenarios, while the previous studies primarily focused on reading speech style with given sentences. Another possibility is that increasing phoneme duration

is not necessary for maintaining the high intelligibility. Several studies confirmed that other inherent temporal properties, such as temporal amplitude modulations and vowel-consonant duration ratios, may directly contribute to enhanced intelligibility rather than phoneme durations (Payton et al., 1994; Hansen, 1996; Krause and Braida, 2004).

In addition to suprasegmental variables, there has been a general consensus concerning the control of segmental parameters (Pisoni et al., 1985; Hansen, 1988; Summers et al., 1988). The rise in subglottal pressure needed to increase vocal effort leads to an increase in formant locations. For example, the wider jaw opening in order to increase sound amplitude causes an increase in F1 Frequency (Huber and Chandrasekaran, 2006). It has also been suggested that under noise conditions, speakers vary their speech characteristics so that speech segments rich in information are emphasized, while those less important to intelligibility are de-emphasized (House et al., 1965; Hansen, 1988; Sodersten et al., 2005). For example, consonant SPL increased at the expanse of vowel energy under noisy conditions in an effort to increase speech intelligibility (House et al., 1965; Hansen, 1988). This is a useful characteristic, as consonants carry more speech information in the presence of noise.

The consistency between the two speaker groups (CI versus NH) indicated above could be mainly due to the presence of auditory feedback provided by the cochlear implant device. Long-term absence of auditory feedback could potentially result in poor regulation of acoustic, phonetic features of adventitiously deafened adults, such as F0, intensity, duration, etc. (Leder et al., 1987; Lane and Webster, 1991). Cochlear implant users, however, may demonstrate useful Lombard perturbation for regulating speech production parameters in noisy environments, which thereby assist in the development of more nearly neutral/typical acoustic, phonetic and temporal patterns under noise (Hochmair-Desoyer et al., 1981; Kirk and Edgerton, 1983; Svirsky and Tobey, 1991; Svirsky et al., 1992; Perkell, 2012). This is possible for post-lingually deaf cochlear implant users because they have learned Lombard effect in their hearing years.

The modification of speech production parameters under the Lombard effect may contribute to ensure intelligible communication in adverse noisy environments. The data from this study indicates that cochlear implant users respond to varying background noise types, and change their speech production accordingly. This articulatory modification allows speakers to avoid speech masked by the acoustic noise, so compensate for the decreased signal-to-noise ratio levels. Previous studies have reported that Lombard effect speech affect the intelligibility of speech (Summers et al., 1988; Junqua, 1996; Lu and Cooke, 2008; Garnier et al., 2010). These studies showed that the intelligibility of Lombard speech is higher than that of speech spoken in quiet environments. The acoustic and phonetic changes between speech produced in quiet and in noise may contributed to an improvement in speech intelligibility over speech produced in quiet. The intelligibility gain increased as the environment became more severe (*e.g.,* increased noise level).

The specific variations in speech production features due to the Lombard effect investigated here can be used to formulate new algorithms for improved intelligibility in noisy conditions. For example, strategies that exploit the impact of particular acoustic features of speech with respect to Lombard speaking style (Zorila et al., 2012; Godoy and Stylianou, 2013). Historically, it is known that different environments will have specific noise types and levels. Traditional front-end processing for hearing aids and cochlear implants, for example, have focused on noise suppression to minimize the impact of noise. Algorithmic advancements which modify neutral speech based on Lombard effect properties offers a unique opportunity to improve the listening/decoding experiences of cochlear implant users.

4.5 Summary: Chapter 4

In this study, we analyzed the speech production of cochlear implant users with respect to environmental context. Naturalistic human-to-human voice interactions were captured using mobile personal audio recordings from continuous single-session audio streams collected

over various realistic environments. An analysis of speech produced in noise and Lombard effect observed in the speech of cochlear implant users was presented. The results indicated that Lombard effect was found in speech of post-lingually deafened cochlear implant adults. Speakers demonstrated increased vocal effort, including F0 and speech SPL, as well as altered glottal spectral slope, and phoneme duration in response to challenging noisy environments. Statistically significant shifts in intensity between vowel and consonant phoneme classes were observed in noisy environments. Segmental articulatory movements, for example, F1 for specific phonemes, /a/, /æ/, /i/, and /u/, and F2 for /a/, and /u/, also appeared to play an important role in relaying Lombard perturbations for speech produced in the presence of noise. The significance of the results is that the Lombard effect could potentially be helping cochlear implant users to ensure/maintain intelligible communication by compensating for the reduced signal-to-noise ratio. The specific variations due to the Lombard effect can be leveraged for new algorithm development and further applications of speech technology to benefit cochlear implant users.

CHAPTER 5

INFLUENCES OF LOMBARD EFFECT ON SPEECH INTELLIGIBILITY IN COCHLEAR IMPLANT USERS

5.1 Introduction

Speech recognition in noise is an extremely challenging task, particularly for hearing impaired listeners using cochlear implant. One way to alleviate this difficulty is to employ Lombard effect to these individuals. Chapter 4 examined the speech production of cochlear implant users with respect to environmental context. In that chapter, naturalistic human voice communication was collected by six post-lingually deafened cochlear implant speakers. Parametric variation in vowel production were investigated as a function of noise. The result indicated that Lombard effect has been found in speech of cochlear implant user. Many speech parameters including intensity, pitch structure, glottal spectral slope and first formant characteristics were shown to increase in response to noisy environments. No significant effect was found for second formant frequency location. However, no study has examined if the effect of Lombard speech influence on speech perception of cochlear implant users. Thus, to extend the previous acoustic study, this chapter focused on perceptual analysis of Lombard effect.

5.1.1 Previous Studies on Speech Intelligibility

A number of studies of the Lombard effect have found that acoustic difference between speech produced in quiet and in noise affect speech the intelligibility (Dreher and O'Neill, 1957; Pittman and Wiley, 2001; Summers et al., 1988; Pickett, 1956; Rostolland and Parant, 1973). Dreher and O'Neill (Dreher and O'Neill, 1957), and Pittman and Wiley (Pittman and Wiley, 2001) reported that speech produced in noise is more intelligible than speech produced in quiet. This was also confirmed by Summers *et al.* (Summers et al., 1988) for the same

signal-to-noise ratio with isolated words or continuous speech. The abovementioned research suggested that the magnitude of this effect increased as the environment became more severe (*e.g.,* increased noise level). However, Pickett (Pickett, 1956) showed that when a speaker increases his vocal effort to a level that corresponds to shouted speech, the intelligibility decreased. In the case of shouted speech, the vocal effort increases the energy but decreases the phonetic information (Rostolland and Parant, 1973).

In addition to Lombard speech materials, several studies have assessed how individual modification affect intelligibility (Skinner et al., 1997; Donaldson and Allen, 2003; Bradlow et al., 2003; Hazan and Markham, 2004; Krause and Braida, 2004; Picheny et al., 1986; Ferguson and Kewley-Port, 2002; Uchanski et al., 1996). These studies used signal processing to assess the role of individual acoustic features on intelligibility. The overall energy of the test stimuli is one of the factors that significantly affects intelligibility (Skinner et al., 1997; Donaldson and Allen, 2003). Bradlow *et al.* (Bradlow et al., 2003) found that there was no correlation between mean F0 and sentence intelligibility. Instead, the range of F0 was found in one case to be significantly correlated with sentence intelligibility. The spectral balance is another factor that affect speech intelligibility (Hazan and Markham, 2004; Krause and Braida, 2004). Vowel space (F1 and F2 range) is a significant feature for increased speech intelligibility, and that formant transition may also play an important role (Picheny et al., 1986; Ferguson and Kewley-Port, 2002; Bradlow et al., 2003). While many studies have not been all in agreement, duration is at least one component that is important to speech intelligibility (Bradlow et al., 2003; Hazan and Markham, 2004; Uchanski et al., 1996; Krause and Braida, 2004).

Although intelligibility benefit is evident for Lombard effect, most perceptual analysis of Lombard effect have focused on adults with normal hearing. Now, Lombard speech research need to expand to include more varied listener groups, for example, cochlear implant users. We believe an evaluation of listener judgments of speech intelligibility is critical. Specifically,

if it can be shown that some of the changes in speech production actually act to increase intelligibility, we may look forward to a device that will exploit this. An investigation of this new clinical populations further our understanding on speech intelligibility under challenging listening environments.

5.1.2 Objectives and Proposed Methods

The primary goal of this study was to examine the influence of Lombard effect on speech intelligibility of cochlear implant users in noisy environments. In addition, the present research also focused on how the perception performance differ from speech produced in different noisy conditions (*i.e.,* the level of vocal effort). The overarching hypothesis of this research was that perceptual benefit can be obtained by cochlear implant users when providing Lombard effect speech. This could provided a complete characterization of the Lombard speech advantage over a range of signal-to-noise ratios.

For these purposes, we developed a corpus that is intended for the perceptual experiment of speech under noise. Normal hearing listeners participated in the corpus to produce the general characteristics of acoustic modification in quiet and under noisy environments. Analysis of speech perception was accomplished in two ways: (i) acoustic properties of speech production in relation to the listening environment, and (ii) perceptual characteristics of Lombard effect speech by cochlear implant users. In the first analysis, the parametric variations in vowel production were investigated as a function of varying environments. This involved vocal intensity, fundamental frequency, glottal spectral tilt, vowel duration, and first and second formant frequencies. The second part of the analysis focused on subjective intelligibility measured with cochlear implant users across signal-to-noise ratios. Five postlingually deaf cochlear implant users participated and responded to the Lombard materials mixed with large-crowd noise. Data from acoustic analysis and its influences on intelligibility showed how perception performance differs from speech produced in quiet and noisy environments.

5.2 Methods

5.2.1 Database Formulation of Speech under Noisy Environments

NH Subjects

The corpus here was developed for the analysis and modeling of Lombard effect and its impact on speech intelligibility. Two normal hearing speakers (one male and one female) (mean age: 25 yrs.) participated. The same number of normal hearing listeners participate as a communication partner. All normal hearing subjects were native speakers of American English. None of them were reported any history of speech and hearing related problems. All participants were recruited from the campus population from the University of Texas at Dallas.

Stimuli

Speech under a single noise type at three noise levels were recorded. The present study concentrated on a large crowd noise. While performance vary according to noise type, the modification of Lombard speech primarily depend on noise level. Noise levels presented in the test were 70 dB, 80 dB, and 90 dB SPL. The noise samples were recorded inside UT-Dallas college cafeteria using LENA device (LENA Foundation, 2014) (see Section 3.2.2 and Figure 3.1). The sentences used in the test were chosen from an AzBio dataset (Spahr et al., 2012). The sentence consisted of 33 lists, each containing 20 sentences (660 sentences). Additional spontaneous speech were collected using collaborative works between two speakers.

Procedures

Data collection was performed in a sound recording booth. Noise source was presented monaurally at different level into open-air headphone worn by speakers. The open-air headphones provided a direct acoustic path, thus speakers were able to hear their own voice

when speaking. Closed-talk headset microphone was used to capture the speech signal. This headset microphone was firmly attached to the head keeping placed 5 cm away from the mouth. The signal was then pre-amplified and recorded at 16 kHz sampling rate and 16 bits quantization level. This way, we recorded a noise-free Lombard effect speech data sequences.

Speakers produced read and spontaneous speech to their partner. The partners were seated 1 meter away in front of speakers. The speech materials were displayed to speakers using LCD screen. The screen was positioned at eye level, thus speakers are able to look at their partner's face across the display. Speech partners always tried to response to a speaker if the sentence/message are recognizable/understandable. They were able to ask speakers repeat the sentences if needed. In this way, the read and spontaneous speech were collected in an inter-personal way.

The audio recording consisted of two consecutive sessions. First session consisted of 660 read sentences in quiet and Lombard conditions. Three hundred and sixty sentences were read in quiet condition, and 300 sentences were read under 3 Lombard effect conditions. In quiet condition, speakers read the sentences naturally without sounds played into their headphone. In Lombard effect condition, noise sample was presented into headphone at three different levels. None of sentences were repeated in this session. The three noise and a quiet conditions were randomly provided to speakers.

In the second session, we collected spontaneous speech in the same four (1 quiet and 3 noisy) environments. Two speakers collaborated to use a map to find the best way. A geographical road map was given to each speaker. A starting point and a destination point were informed (*e.g.,* from university's visitor center to remote parking at Dallas Fort-Worth airport). Both participants communicated to find the best way to get a target location. Only verbal communication was allowed in this session. A maximum of 5 minutes was given for each condition.

A 30-minute break in between each recording session allowed speakers to rest. Speakers were invited to drink warm water in regular intervals to reduce vocal fatigue. Figure 5.1

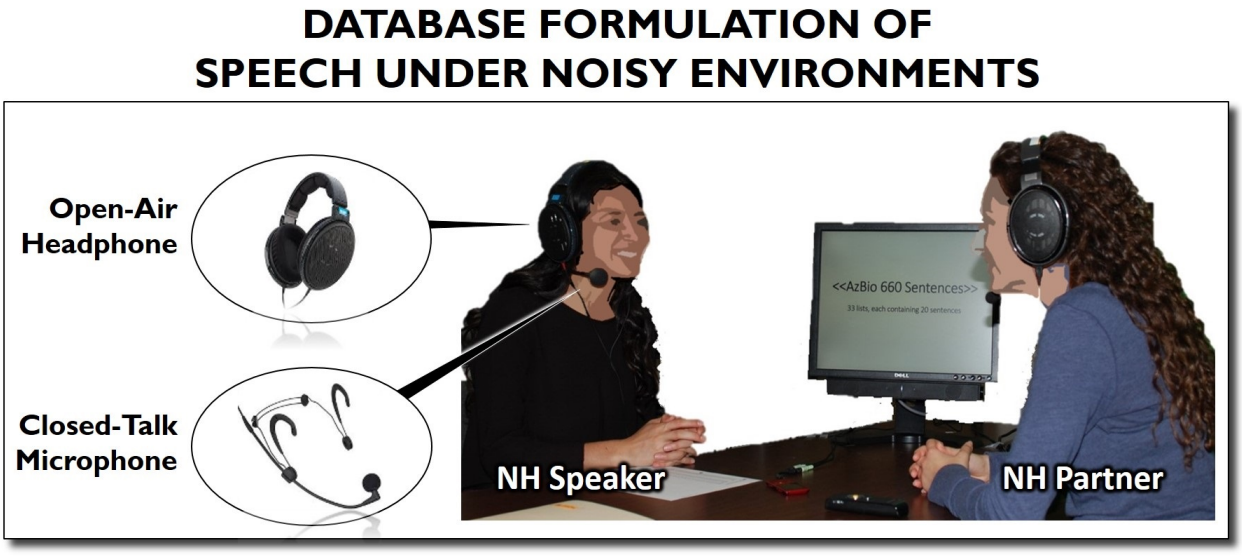


Figure 5.1. Lombard effect data collection with normal hearing participants. Data collection was performed in a sound recording booth. A set-up for data acquisition using open-air headphone and close-talk microphone was demonstrated.

demonstrates how the Lombard speech corpus was collected with normal hearing participants.

5.2.2 Signal Processing: Features and Metrics

Following the data collection, a number of acoustic parameters that are known to be sensitive to Lombard speech were measured. These included (i) vowel intensity, (ii) fundamental frequency (F0), (iii) overall glottal spectral tilt (H1-A3), (iv) phoneme duration, (v) first formant frequency (F1) location, and (vi) second formant frequency (F2) location. All measurements except overall spectral tilt and phoneme duration were computed using PRAAT software (Boersma, 2002). Spectral tilt was calculated from the difference between the magnitudes of the first spectral harmonic (H1) and that of the third formant peak (A3) (Hanson, 1997; Iseli et al., 2007) (See Section 4.2.2). Phoneme duration was obtained from analysis of the phoneme-level transcripts. Each acoustic-phonetic feature was extracted based on phoneme nuclei boundaries marked by a forced phoneme alignment procedures described

in Section 3.3. The beginning and ending markers of each phoneme were reduced by 20% to eliminate any transitional effects across phoneme classes. The duration of the analysis window used here was set to 20 ms with a 10 ms skip rate.

Several statistical analyses were employed to determine if any specific parameters that are significantly different to neutral speech baseline. A repeated-measures analysis of variance (ANOVA) was performed to assess the effect of noise type on speech parameters. Subjects were considered as random (blocked) factors, while noise conditions were used as the main analysis factors. Following the ANOVA, a *post-hoc* pairwise comparison test was performed to determine (1) if each of the noisy environments are potentially significantly different from the neutral baseline, and (2) if each noisy environment could potentially be significantly different with the other noisy environments. Bonferrroni adjustment was used to control for family-wise error in the pairwise test. In this study, a difference in means between two or more groups was considered significant if the significance level fell below 5.0% ($p < 0.05$).

5.2.3 Evaluation with Cochlear Implant Users

CI Subjects

A subjective listening test was performed to quantify the effect of Lombard on speech perception in cochlear implant users. Five cochlear implant listeners (two male and three female) (mean age: 64 years) who were post-lingually deaf adults participated. They were all fitted with the Nucleus device from Cochlear Ltd. with advanced combined encoder (ACE) strategy. They all used their cochlear implant device for at least four years. All cochlear implant subjects were native speakers of American English. Detailed biographical information of the cochlear implant participants is presented in Table 5.1.

Table 5.1. Characteristic information of cochlear implant subjects who participated in the perceptual evaluation of Lombard effect.

Listener	**Gender**	**Age (yrs.)**	**Years of hearing loss**	**Years implanted**	**Etiology of hearing loss**	**Implant ear**	**Coding strategy**
Listener 1	Female	62	56	11	Hereditary	Bilateral	ACE
Listener 2	Female	56	30	5	Hereditary	Bilateral	ACE
Listener 3	Female	59	30	6	Hereditary	Bilateral	ACE
Listener 4	Male	69	13	7	Hereditary	Left only	ACE
Listener 5	Male	66	55	4	Hereditary	Bilateral	ACE

Stimuli

The performance of the Lombard effect was tested in twelve conditions. These consisted of 4 speaking styles (speech in quiet and large-crowd noise at 70 dB, 80 dB, and 90 dB SPLs) in 3 noisy environments (quiet, 15 dB, and 10 dB SNRs). Each condition was composed 20 read sentences read by a male and a female normal hearing talkers (10 sentences from each talkers). The original clean sentences were individually mixed with a large crowd noise at 10 dB and 15 dB SNRs (160 degraded sentences). The same noise samples used in data collection was used in the listening test (*i.e.,* large-crowd noise). For comparative purpose, we presented to cochlear implant users clean Lombard sentences without background noise (80 clean sentences).

Procedures

The listening test was conducted in anechoic sound booth, using an interactive computer user interface. The cochlear implant subjects listened to the stimuli monaurally presented via loud speakers. The speech presentation level was fixed at 60 dB SPL. The noise level

was varied to produce different signal-to-noise ratios. The cochlear implant subjects listened to the stimuli monaurally presented via loudspeakers. Listeners were allowed listen to each test file only once to recognize the sentences. Recognition accuracy was scored based on the number of words correctly identified.

To avoid a sentence repetition effect, sentences were used only once for a given subject over the entire experiment. To avoid presentation order effect, Lombard and neutral speech sentences were mixed together and presented in random order. All subjects followed the same random order of list. In order to familiarize with the materials, a short session with 5 sentences in quiet and noise was conducted for practice prior to testing. The entire test session were limited to two hours to minimize subject fatigue. The subjects were asked to take a break at regular interval. The proposed subjective listening evaluation with cochlear implant patients is illustrated in Figure 5.2.

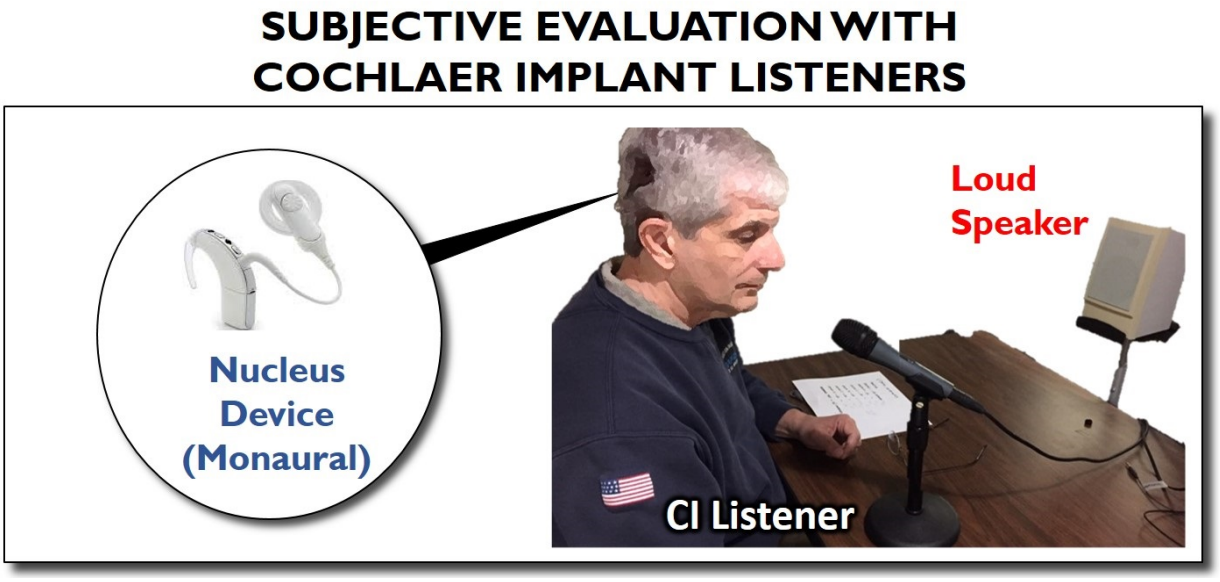


Figure 5.2. Proposed subjective listening test with cochlear implant patients. Data collection was performed in an anechoic sound proof booth. A set-up for data acquisition using loud speaker is demonstrated.

5.3 Results

5.3.1 Acoustic Analysis of Lombard Effect

An analysis of vowel speech production under noisy environment for individual speakers was presented in Figure 5.3. Six acoustic parameters (Vowel intensity, fundamental frequency (F0), overall spectral tilt, vowel duration, and first (F1) and second (F2) formant frequencies) were examined. While vowel intensity indicates the variation in signal strength over time, pitch, spectral tilt and formant frequency deliver the pattern of shift in spectral cue/energy in the frequency-domain. Phoneme duration represents the temporal aspect of the vowel production. Note that in this work, we established the speech produced in *quiet environment* as *neutral baseline*.

Figure 5.3 (a) and 5.3 (b) show the average vowel intensity and F0 for individual speaker as a function of speaking conditions. The test condition were anechoic quiet and large crowd noise at 70 dB, 80 dB, and 90 dB SPLs, represented as Neutral, Lom70, Lom80, and Lom90 respectively. The asterisk marked at each data point indicates a statistically significant change when compared to the neutral baseline ($p < 0.05$). The result indicated that the average vowel intensity and F0 were statistically different from neutral speech style. Average vocal intensity increased significantly for all Lombard effect conditions (Lom70, Lom80, Lom90). Average F0 increased significantly for all Lombard effect conditions except Lom70. Only little change occurred in F0 for speech under large crowd noise at 70 dB SPL condition ($p > 0.05$).

Figure 5.3 (c) and 5.3 (d) present the variation of spectral tilt and vowel duration with respect to each speaking style. Spectral tilt was found to progressively decrease with increasing noise SPL. Spectral balance of vowel was affected by the higher vocal effort, resulting in relatively more intensity on higher frequency band of the spectrum. A significant effect of environment type on spectral tilt was observed for all Lombard conditions ($p < 0.05$). For

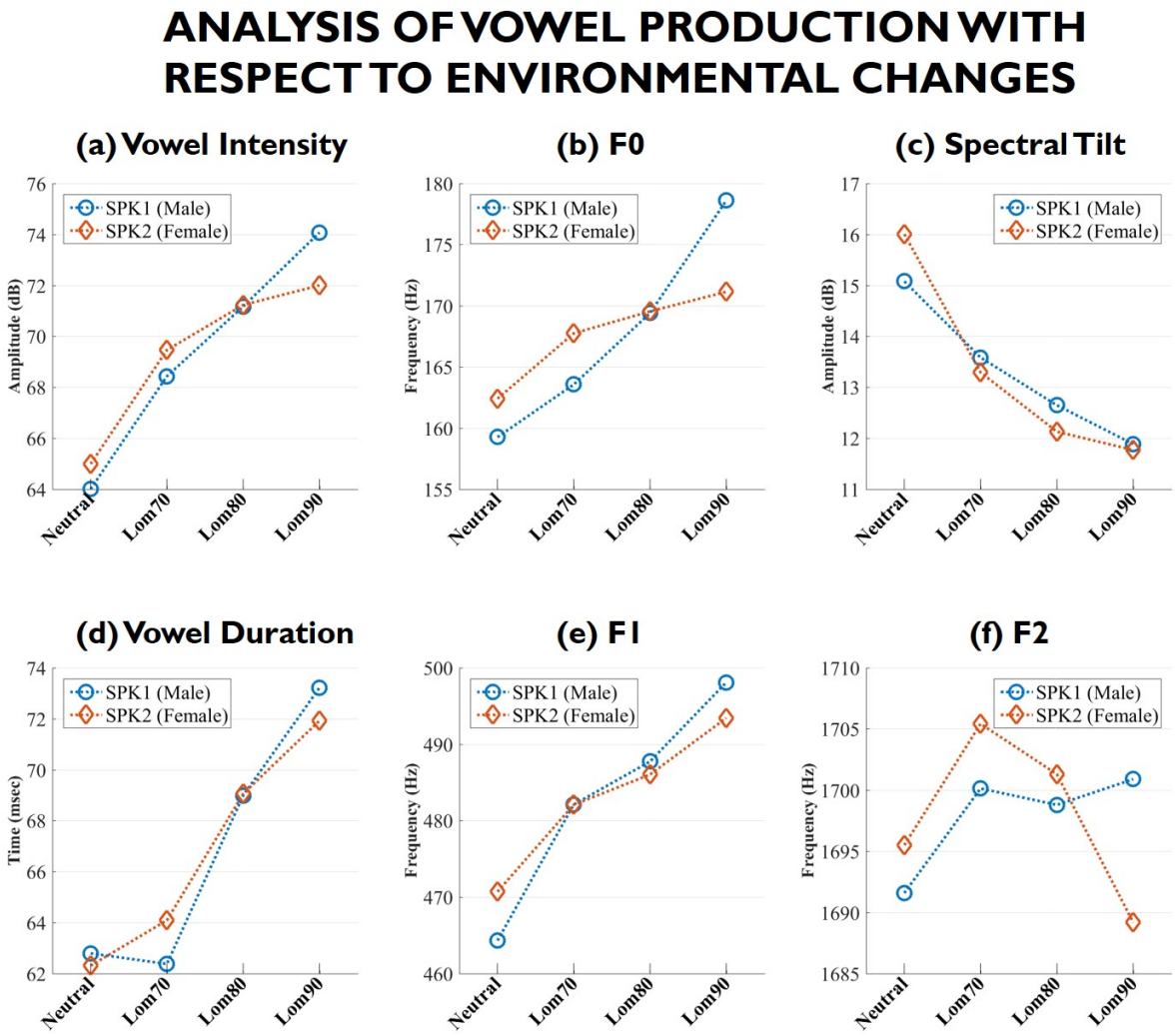


Figure 5.3. Acoustic analysis of vowel production: individual variations of (a) vowel intensity, (b) fundamental frequency (F0), (c) spectral tilt, (d) vowel duration, (e) first formant frequency (F1), and (f) second formant frequency (F2), as a function of varying listening environments. Asterisk indicates statistical significance ($p < 0.05$) from natural speech.

vowel duration, variations in mean increased with noise SPL. Speech under noise at 80 dB and 90 dB SPL result in a statistical significant change in the vowel duration as compared to the neutral baseline. No significant effect was observed for 70 dB SPL condition.

Figure 5.3 (e) and 5.3 (f) display the variation in F1 formant location and F2 formant location with respect to listening environments. Overall, F1 formant location changed with respect to speaking condition, but F2 formant location did not. F1 formant frequency

significantly increased for all Lombard speaking conditions. The pattern of the results was consistent with many other acoustic parameters (*e.g.,* vocal intensity and spectral tilt). However, this was not the case for the F2 parameter. As shown in Figure 5.3 (f), no significant changes occurred for F2 formant location across all Lombard environments for all speakers ($p < 0.05$)

In order to summarize important Lombard relaying parameters, we grouped three Lombard conditions into single condition. This allowed to perform statistically significant test (ANOVA) on the two fixed groups: neutral and Lombard speech. The analysis indicated that vowel duration ($F[3,4] = 75.85, p < 0.01$) was the primary stress relayer for varying vocal effort. Average vowel intensity ($F[3,4] = 34.25, p < 0.01$) and spectral tilt ($F[3,4] = 34.98, p < 0.01$) were also found to be statistically different from neutral style speech. Average first formant location F1 ($F[3,4] = 34.04, p < 0.01$) showed statistically significant shift from neutral speech. The average pitch ($F[3,4] = 6.8, p < 0.05$) was found to be statistically different from neutral style speech. However, second formant location F2 ($F[3,4] = 1.58, p > 0.05$) were not significantly different from neutral speech.

5.3.2 Perceptual Analysis of Lombard Effect

Figure 5.4 and Table 5.2 shows percent correct scores for Lombard and neutral speech as a function of signal-to-noise ratio in five cochlear implant subjects. Asterisk indicates statistical significance ($p < 0.05$). Error bars indicate the standard error of the mean (SEM). The results obtained from the Lombard speech are compared against those obtained by testing the subjects with the unmodified neutral stimuli. Scores obtained with the anechoic quiet condition are also given for comparison to provide the upper bound in performance. The most significant finding was that Lombard speech produced higher intelligibility than conversational speech. The average intelligibility score obtained in anechoic quiet condition was 67.3 % for the five tested cochlear implant listeners. However, this intelligibility scores

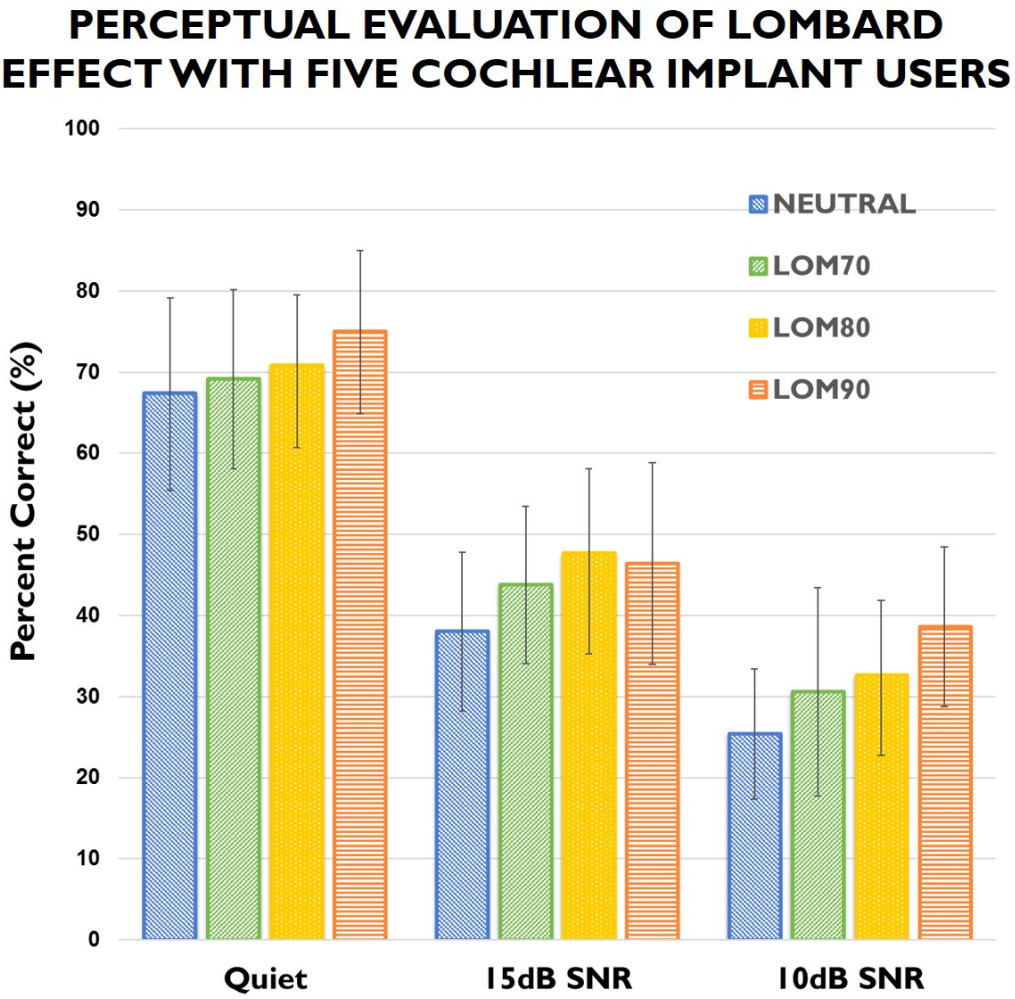


Figure 5.4. Average intelligibility scores of five cochlear implant users for neutral and Lombard speech as a function of signal-to-noise ratio. The test conditions in each environment were speech produced in quiet and large crowd noise at 70 dB, 80 dB, and 90 dB SPLs (represented as Neutral, LOM70, LOM80, and LOM90 respectively). Asterisk indicates statistical significance ($p < 0.05$) from natural speech. Error bars indicate the standard error of the mean (SEM).

dropped down to 38 % and 25.4 % in SNR = 15 dB and 10 dB conditions. The decreased scores, however, improved by employing Lombard effect up to 47.7 %, and 38.6 % in SNR = 15 dB and 10 dB conditions respectively. In addition to noisy environments, intelligibility advantage of Lombard speech was also evident in anechoic quiet condition. The average intelligibility score obtained in anechoic quiet condition, 67.3 % was improved up to 74.9 %.

Table 5.2. Average word recognition scores (%) of five cochlear implant users.

Condition	Neutral	Lom70	Lom80	Lom90
Quiet	67.3	69.1	70.8	74.9
15 dB SNR	38.0	43.8	47.7	46.4
10 dB SNR	25.4	30.6	32.6	38.6

As expected, the percent correct score increased as a function of signal-to-noise ratio. For example, the difference in intelligibility score between neutral and Lombard effect speech are by up to + 7.6 %, + 9.7 %, and + 13.2 % for anechoic quiet, SNR = 10 dB and 15 dB conditions respectively. An analysis of variance confirmed a significant interaction between listening environment and speech intelligibility was observed. The improvements are found to be statistically significant ($p < 0.05$) in relatively adverse listening environment (SNR = 10 dB). For SNR = 15 dB, no significant difference was found.

It is interesting to notice that there is a great correlation between the increase of the intelligibility and the level of vocal effort. As shown in Figure 5.4, higher vocal effort speech was more intelligible than the speech produced with lower vocal effort. For example, for speech produced with lower vocal effort (*i.e.,* Lom70), listener obtained a mean identification score of 43.8 % and 30.6 % for SNR = 15 dB and 10 dB. However, for the sentences produced with higher vocal effort, intelligibility scores were improved by up to + 3.9 % and + 8.0 % for Lom80 and Lom90 conditions respectively. The similar pattern of the results was obtained for the speech in quiet condition as well.

5.4 Discussion

In this chapter, analysis of speech production with respect to environmental noise was presented with Lombard effect in normal hearing listeners. Statistical significant test were em-

ployed to determine if any changes in speech parameters are reliable Lombard relayers. The results indicated that many speech parameters were used in conveying the Lombard effect stress conditions. Talkers demonstrated significant changes in vocal intensity, fundamental frequency, glottal spectral slope, vowel duration, and first formant frequency. No significant effect between neutral and Lombard speech was found for second formant frequency. These was presumed to be the result of increased vocal effort as a response to the reduced auditory feedback due to noise.

The current results from the acoustic analysis generally confirm the previous studies for normal hearing listeners. Many existing studies suggested that Lombard effect resulted in the modulation of acoustic parameters of speech production under noise (Pisoni et al., 1985; Hansen, 1988; Summers et al., 1988; Junqua, 1992; Hansen, 1996; Lu and Cooke, 2008; Garnier et al., 2010). In these studies, results showed an increase in overall amplitude of vocalic sections, increased duration, increased average F0, and a decreased spectral tilt. This was presumed to be the result of increased subglottal pressure, and vocal-fold tension as a response to the reduced auditory feedback due to noise. Existing studies suggest that the most widely considered area of the Lombard effect involves vocal intensity and F0 (Hansen, 1996; Lu and Cooke, 2009; Garnier et al., 2010). Spectral balance of vowels was affected by the higher vocal effort for Lombard speech, resulting in relatively greater intensity in the higher frequency band of the spectrum (Hansen, 1988; Junqua, 1992; Lu and Cooke, 2008). Prior works indicated that vowel duration was considered to as a crucial factor for exhibiting Lombard effect (Pisoni et al., 1985). There has been a general consensus concerning the control of first formant frequency (Pisoni et al., 1985; Hansen, 1988; Summers et al., 1988). The rise in subglottal pressure needed to increase vocal effort leads to an increase in first formant locations.

In addition to speech production, the perceptual performance of the Lombard effect was analyzed using intelligibility listening test with cochlear implant users. Perceptual difference

between speech produced in quiet and noisy environments presented across signal-to-noise ratios. The results indicated that the Lombard speech yielded a significant improvement in intelligibility in both quiet and noisy conditions. The improvement of the intelligibility was larger in challenging listening environments. Higher vocal effort speech resulted in more intelligible speech than the speech produced with lower vocal effort. This improvement was attributed to the modification of speech production parameters under the Lombard effect.

Our findings on Lombard speech perception by cochlear implant users were in agreement with those of previous studies with normal hearing listeners. Previous studies have reported that intelligibility of Lombard speech was found to be higher than that of speech spoken in quiet environments (Dreher and O'Neill, 1957; Summers et al., 1988; Pittman and Wiley, 2001; Pickett, 1956; Junqua, 1992; Lu and Cooke, 2008). In these studies, the acoustic changes between speech produced in quiet and in noise may contributed to an improvement in speech intelligibility over speech produced in quiet. The intelligibility gain increased as the environment became more severe (*e.g.,* increased noise level) (Pickett, 1956). It was also reported that the intelligibility of speech increased with increased in vocal effort, but dropped with increasingly forceful shouting (Picheny et al., 1986; Rostolland and Parant, 1973).

Example stimulus output patterns (electrodograms) was used here in order to visually assess the effectiveness of the Lombard speech in enhancing intelligibility. Electrodograms of neutral and Lombard speech signal are plotted in Figure 5.5. In all panels shown, the vertical axes represent the electrode position corresponding to a specific frequency, while the horizontal axes show time progression. The speech samples used to create the plot was derived from the UT-SCOPE database (Ikeno et al., 2007). The sentence "Basketball can be an entertaining sport"was spoken by a male normal hearing speaker in quiet and in large crowd noise at 90 dB SPL. These speech were then normalized to have the same root-mean-square level, and mixed with a speech shape noise masker at 10 dB SNR. The resulting

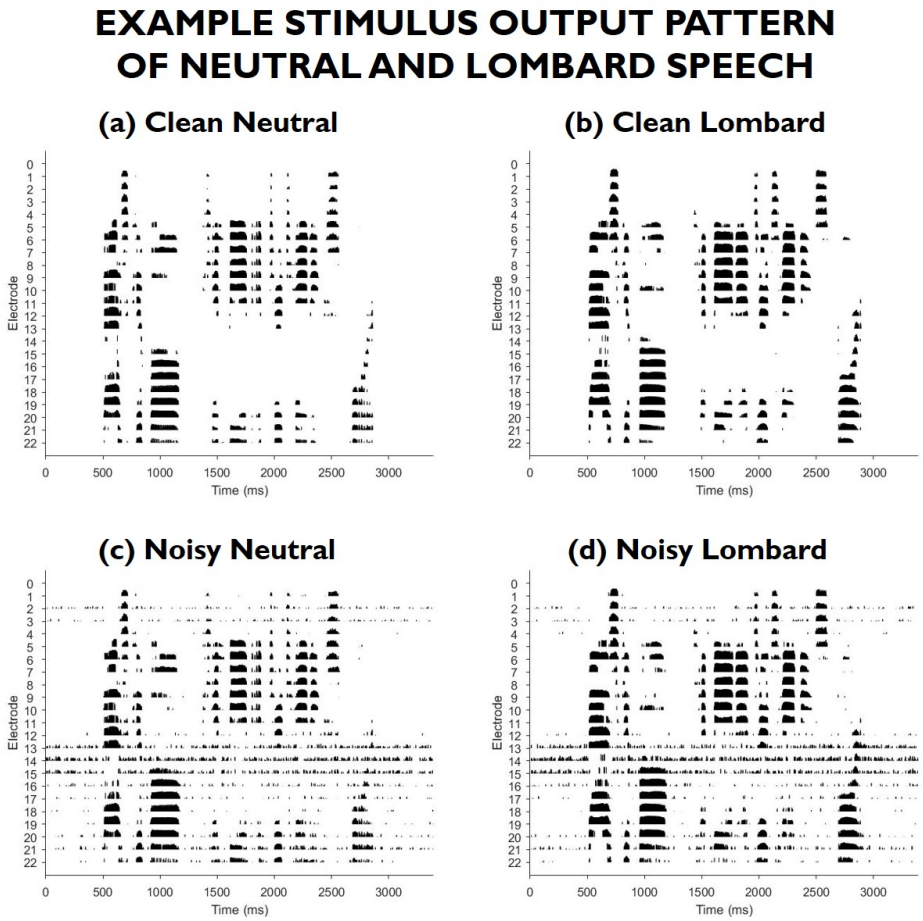


Figure 5.5. Stimulus output patterns (electrodograms) of the sentence "Basketball can be an entertaining sport" from UT-SCOPE database: (a) original neutral speech reference, (b) Lombard speech, (c) original neutral speech mixed with speech-shaped noise at 10 dB SNR, and (d) Lombard speech mixed with speech-shaped noise at 10 dB SNR. Neutral and Lombard speech used here was produced by a normal hearing speaker in quiet and under large crowd noise at 90 dB SPL.

stimulus patterns were obtained using a 22-band Advanced Combination Encoder (ACE) (Loizou, 1999; Vandali et al., 2000) sound processing strategy.

From the result, it was clear that Lombard speech modified the spectral envelop more flattened over frequency. The figure shows an increase in electrical activity on high-frequency channels (1 - 16) in Lombard speech (Figure 5.5(b)) when compare to the neutral baseline (Figure 5.5(a)). This was more evident by comparing Figure 5.5 (c) and 5.5 (d). The noise

added to the speech signal distorts the lower frequencies more than the higher frequencies. However, we noticed that Lombard stimuli provided more spectral energy in higher frequency region, which cause consonant and formant signal tend to be emphasized. This can provide more favorable in terms of intelligibility against background noise. The data presented here indicate a potential perceptual benefit of the Lombard effect for cochlear implant users.

5.5 Summary: Chapter 5

This chapter provided an investigation how listeners with cochlear implant would respond to Lombard effect speech under noisy listening environments. We analyzed acoustic properties of Lombard and neutral speech from two normal hearing speakers. Along with acoustic analysis, we measured their speech intelligibility with cochlear implant users in quiet and over signal-to-noise ratios. Consistent with previous studies, Lombard speech produced by normal hearing speakers had a higher vocal intensity, increased F0, decreased spectral slope, longer vowel duration, and shift in first formant location. This modification of speech production influenced intelligibility for electric hearing in both quiet and noisy environments. The results indicated an advantage of $+ 7.6 \sim + 13.2$ percentage points in intelligibility for Lombard speech over neutral speech. Larger improvement in intelligibility was found in challenging listening environments. Higher vocal effort speech are more intelligible than the speech produced with lower vocal effort. These results suggested that the modification of speech production parameters obtained under noise contribute to the Lombard speech intelligibility. The findings presented in this chapter further our understanding of the speech perception of cochlear implant users.

CHAPTER 6

DEVELOPMENT OF AN INTELLIGIBILITY ENHANCEMENT ALGORITHM BASED ON LOMBARD EFFECT PROPERTIES

6.1 Introduction

Although most cochlear implant users are able to achieve open-set speech recognition scores of 80 % or higher in quiet and anechoic environments, their performance degrades significantly in the presence of noise. Previous studies in cochlear implants have shown the absence of fine spectral structure may contribute to the poorer performance under noisy environments (Fu et al., 1998). It has been also known that reduced temporal fine structure may exacerbate the speech perception faced by cochlear implant users in noise (Moore, 2008). To ensure a high communication quality, a high speech intelligibility must be provided in challenging listening conditions.

In order to enhance speech intelligibility, speakers employs "Lombard effect" within their speech production. The Chapter 4 confirmed Lombard perturbation in the speech production of post-lingually deafened cochlear implant individuals. In addition to acoustic study, the Chapter 5 also found improvement in intelligibility when Lombard speech is presented to cochlear implant users in noisy conditions. In this study, perceptual difference between speech produced in quiet and noisy environments were measured across signal-to-noise ratios. The result indicated that Lombard speech yielded up to +13.2 % improvement in intelligibility over neutral speech in noisy environments. This improvement was larger in challenging listening environments. The results suggested that the modification of speech production parameters obtained under noise contribute to the Lombard speech intelligibility.

Motivated by these findings above, we will utilize Lombard speech perturbation processing paradigms to enhance sound coding of cochlear implant processors to potentially expand speech comprehension in noisy environments. This paradigm will allow more natural and

desirable approach of enhancing the speech signal itself; much like normal human communication in adverse listening environments to help maintain speech intelligibility. This will be achieved by artificially imparting Lombard effect perturbation in speech signal prior to cochlear implant encoding. Before proceeding to proposed a Lombard perturbation processing algorithm, we summarize previous studies on intelligibility/quality enhancing speech modification strategies in the following section.

6.1.1 Past Speech Modification Techniques

There has been a growing interest in algorithmic modification that aim to increase the intelligibility of neutral speech when presented in noise. One obvious solution to the speech enhancement problem is to increase the level of the speech (Skinner et al., 1997; Donaldson and Allen, 2003). Intelligibility can be maintained simply by raising the speech level to increase the signal-to-noise ratio. However, at a certain point, increasing the speech level may not be possible due to processor limitation or unpleasant levels. Therefore, a common approach is to fix the speech energy and redistribute speech energy over time or frequency domains. An example of time domain enhancement approaches was found to slow down speech. On the other hand, an effective and simple way to improve speech intelligibility is by changing the spectrum of the speech (Niederjohn and Grotelueschen, 1976; Schepker et al., 2013; Jokinen et al., 2016). For example, speech understanding increased when high frequencies are amplified at a cost of low frequencies. Experiments with the abovementioned approaches showed large intelligibility improvements over the unprocessed noisy speech.

In addition to individual modification, a limited number of previous studies on simulating emotions in synthetic speech has been established (Bou-Ghazale and Hansen, 1996; Hansen and Cairns, 1995). The aim of these studies have been to control the speech parameters of a synthesizer in order to present different emotions or speaking-styles. Bou-Ghazale and Hansen (Bou-Ghazale and Hansen, 1996) proposed a mathematical models for representing

speech parameter variations under angry, loud, and Lombard effect speaking. These models were used to modify the speaking style of neutral speech and to enhance the naturalness of text-to-speech system, many of which might be described as sounding Mechanical. Hansen and Cairns (Hansen and Cairns, 1995) proposed a speaker-independent perturbation algorithm based on differences between neutral and Lombard speaking conditions. The study modeled amplification, duration, and overall spectral structure to characterize speech parameter variation under noise. The knowledge from these model can be integrated within an automatic speech recognition system to improve recognition under noise.

Lastly, certain techniques that exploit the impact of particular acoustic characteristics with respect to a given speech style have been developed (Zorila et al., 2012; Godoy and Stylianou, 2013). In order to make speech more intelligible, various modification strategies have been proposed. Zoria *et al.* (Zorila et al., 2012) attempted to mimic acoustic trends observed in human "Lombard speech"in order to increase intelligibility of neutral speech. Spectral shaping and Dynamic Range Compression (DRC) methods were employed to modified Lombard-style speech. Moreover, a modification based on generally on observations from "clear-style speech"was performed by Godoy and Stylianou (Godoy and Stylianou, 2013). Specifically, the proposed approach included uniform time-stretching, and vowel space expansion in order to enhance intelligibility. The proposed methods mentioned above were highly successful at increasing speech intelligibility of normal hearing listeners in near-end (*e.g.,* telephony and public address) environments.

In summary, there is a potential feasibility in algorithmic modification of neutral speech to increase intelligibility/quality in noisy environments. However, the abovementioned speech enhancement techniques have mostly focused on text-to-speech, automatic speech recognition, or near-end systems. Virtually, none of these have ever been employed for cochlear implant areas. Historically, conventional signal processing for cochlear implants have focused on noise suppression strategies to improve performance under adverse noisy environments (Hochberg et al., 1992; Weiss et al., 1993; Loizou et al., 2005; Yang and Fu, 2005).

However, many of these techniques are perceptually motivated, and might not provide benefits under both stationary and non-stationary maskers. In order to address the above limitations, it would be desirable to have an entire new approach for speech enhancement techniques based on speech production constraints.

6.1.2 Objectives and Proposed Methods

The proposed study addressed the analysis, development, implementation, and assessment of Lombard effect signal processing. The aim of this chapter was to propose a novel intelligibility enhancing speech modification algorithm based on Lombard effect properties. This algorithm was compatible with cochlear implant users. This chapter also proposed to examine how cochlear implant users perceive the artificially modified speech in challenging listening environments. Note that the proposed algorithm attempt to modify the speech signal to increase intelligibility rather than suppressing background noise. This paradigm allows ultimately increased effectiveness and relevance in signal processing for cochlear implants. The common assumption to develop the proposed algorithm was that a clean version of the speech signal is available, *i.e.,* the potential noise is assumed to be successfully suppressed.

In order to develop an effective modification scheme, a previous proposed framework based on Source Generator theory (Hansen, 1994; Hansen and Cairns, 1995) was employed. This theory presumes that speech under noisy environments can be modified by mapping neutral speech parameters. Based on this assumption, we modeled speech parametric variations of neutral and Lombard effect conditions. The modification areas considered in the proposed algorithm were: (1) voiced intensity, (2) overall spectral contour, and (3) sentence duration. The speech dataset used for modeling was derived from UT-Scope (Ikeno et al., 2007) stressed speech corpus. Modification transformation was then calculated based on differences from neutral speaking conditions, and applied to the input speech to generate resulting Lombard synthetic speech. Subjective listening evaluation was performed with five

cochlear implant users to demonstrate the effectiveness of the proposed speech modification algorithm in adverse listening environments.

6.1.3 Significance

In contrast to the traditional methods for noise suppression, the proposed approach was inspired by the natural human speech production physiology. Leveraging vocal tract and excitation constraints were expected to produce a new class of speech enhancement strategies, which is more suited for hearing impaired individuals. We developed a model-based speech enhancement method which integrate Lombard effect signal processing for better speech perception. This method can be implemented for both speech enhancement and front-end processing for cochlear implant users. Algorithmic advancement proposed in this chapter offers a unique opportunity to improve the listening and decoding experience of cochlear implant users in the presence of noise.

6.2 Methods

6.2.1 Algorithm Development

Framework to develop algorithm

The purpose here was to investigate the prospect of modifying the Lombard speaking style of neutral style to enhance speech intelligibility. A previous proposed framework based on Source Generator theory was employed to represent production of speech under Lombard effect. Source Generator theory was originally developed by Hansen (Hansen and Cairns, 1995; Hansen, 1994), and successfully employed as a means for stress perturbed token (*e.g.,* loud and Lombard speech) in a recognition/synthesis application (Bou-Ghazale and Hansen, 1996; Hansen and Cairns, 1995). In this theory, it was assumed that speech under a given condition can be characterized by the statistical variations in speech parameters which is spoken in a given condition. For example, the production of a neutral word can be represented

as the fluid movement in speech production space with some degree of neutral variation. Under stress conditions, however, the resulting path is different from that of neutral due to the physical variations. Consequently, the proposed theory presumed that speech production under given condition (*e.g.,* Lombard speech) can be transformed by mapping the neutral speech parameters.

Based on the proposed framework above, we formulated statistical models to characterize speech parameters variations under Lombard effect. This model was directly applied to neutral speech to modify the speaking style. The modification areas considered here were: (1) amplification, (2) spectral contour, and (3) overall sentence duration. It is noted that fundamental frequency was not included in the intelligibility enhancement algorithm since the characteristics of cochlear implant devices. It has been known that reduced temporal fine structure provided by CIs results in poorer performance for pitch perception (Moore, 2008). In this study, we modeled the parameter variations across a number of speakers in this study, not variations particular to an individual speaker. This allowed us to develop a general method of speech modification which could be applied to any new input speaker. Each of these three modification for input neutral speech as well as dataset used for parameter

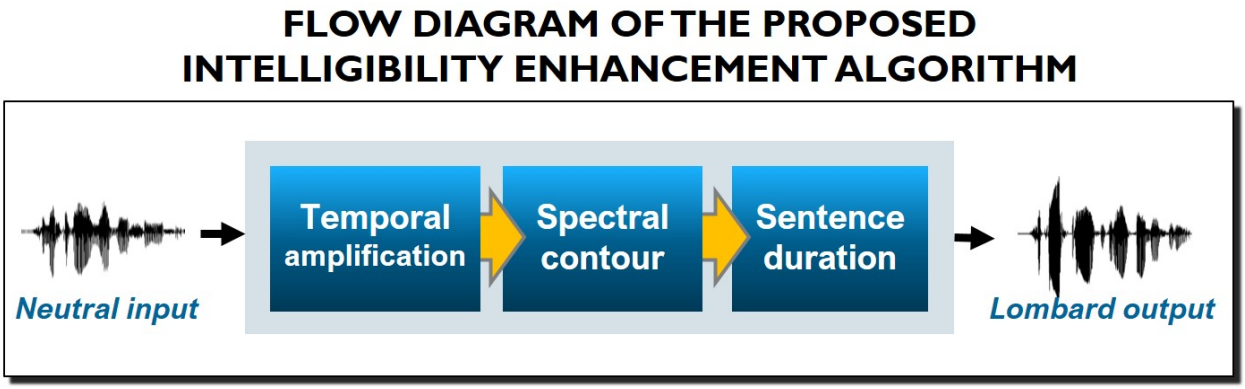


Figure 6.1. Plots presenting the overall description of the proposed speech enhancement algorithm. The proposed algorithm controls the acoustics features of input neutral speech to present Lombard speaking style output. The modification areas considered here were: (1) amplification, (2) spectral contour, and (3) overall sentence duration.

modeling were described in turn as follows. The proposed system for speech modification is shown in Figure 6.1.

Dataset for parameter modeling

In order to develop an effective Lombard perturbation modeling, a prior knowledge of how speech features vary under noise must be obtained. The analysis and modification conducted in this work were based on a previously established Lombard database, called UT-Scope (Speech under Cognitive and Physical Stress and Emotion) (Ikeno et al., 2007). In the corpus, 59 normal hearing speakers participated and produced 20 sentences chosen TIMIT corpus (Garofolo et al., 1993) under large crowd noise at 80 dB SPL. The same 20 sentences were repeated under quiet condition. The data collection was performed in a sound recording booth using closed-talk microphone. The noise sounds were presented open-air headphone worn by the speakers. This way, a nose-free Lombard effect speech data sequences were recorded. A total of 2360 tokens (20 sentences 2 speaking conditions 59 speakers) was used for modeling parametric variations under noise.

Entropy-based temporal amplification

The first part of speech modification employed intelligibility dependent strategy for amplifying signal power in time-domain. Conventional automatic gain control (AGC) and dynamic range compression (DRC) in general increased the level of the speech for both high and low important acoustic cues (Skinner et al., 1997; Donaldson and Allen, 2003; Zeng et al., 2002; Spahr et al., 2007). The proposed algorithm, however, provided to change only with the speech that are assumed to be important for speech intelligibility. For example, large amplification was applied for the case of high intelligibility while no amplification was applied for low intelligibility segments. This approach was motivated by previous studies proposed

Hansen (Hansen, 1988) and House *et al.* (House et al., 1965). They demonstrated that consonant energy increased at the expense of vowel energy under Lombard effect in an effort to increase speech intelligibility. Kewley-Port (Kewley-Port et al., 2007) also suggested that spectral change within sentences, for example, vowel-consonant boundaries, predict speech intelligibility better than individual phoneme classes (*e.g.,* vowel and consonant).

Motivated by the findings above, we employed a cochlear-scaled entropy estimation (Stilp and Kluender, 2010) to identify the information bearing segments from the input speech. The cochlear-scaled entropy can capture spectral change, particularly at identifying transition

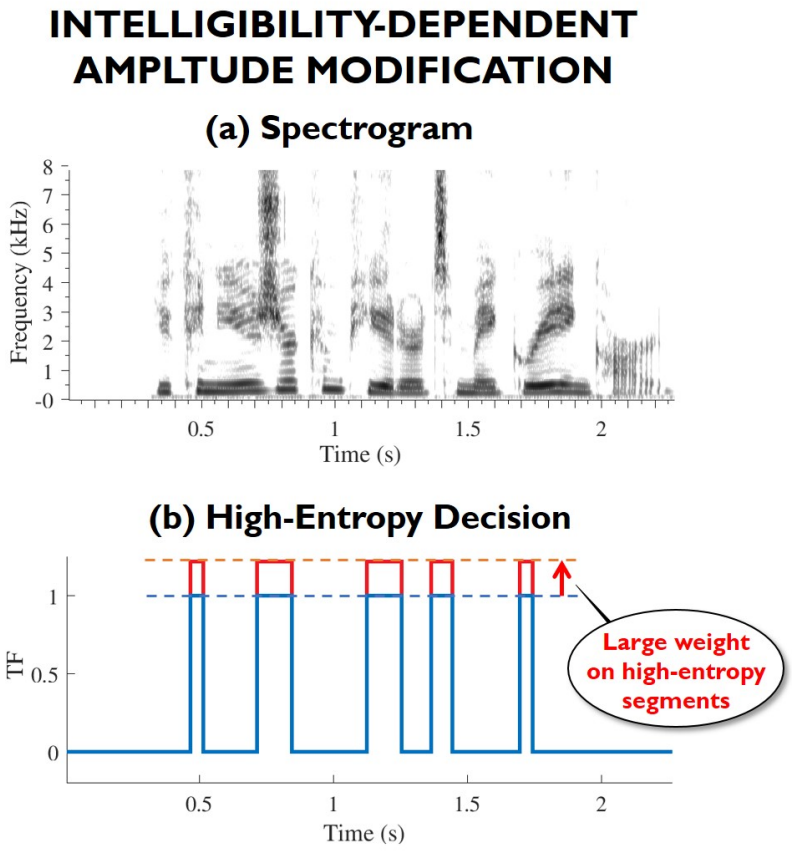


Figure 6.2. Plots demonstrating the entropy-based temporal modification: (a) spectrogram of the input sentence, and (b) high- and low- entropy decision output. A cochlear-scaled entropy (Stilp and Kluender, 2010) was used to estimate the high intelligibility segments (*e.g.,* consonants, vowel-consonant boundaries). Large weight was placed on the high-entropy segments, while no weight was applied for the low-entropy segments.

within vowel, consonant, and vowel-consonant boundaries. To compute the entropy, the input signal was first divided into 16 msec segments and bandpass filtered into 16 bands using Gammatone filterbank (Patterson et al., 1987). Euclidean distance between adjacent segments was calculated across the 16 filter outputs. Cumulative Euclidean distance with five successive segments (80 msec in duration) was taken as the output measures of cochlear-scaled entropy. The segmentation was then classified into two levels, *i.e.,* low- and high-entropy according to a threshold. The threshold was determined based on a proportion coefficient p, such that p percent of all entropy within the utterance/sentence are smaller than the threshold. In this study, the value of proportion coefficient p was set to 0.6, suggesting that 60% of the entropy values within the utterance were smaller than the threshold.

The classified segments were next used to compute the power ratio between the high-entropy and low-entropy segments. The average signal power for each segment was estimated, then, the ratio of high-to-low entropy power for each sentence was calculated. The metric for measuring the signal power was PRAAT software (Boersma, 2002). The power ratios obtained here were averaged across the whole utterances for each speaking condition. An amplitude scaling factor was calculated as the ratio of the Lombard and neutral speaking conditions. The high cochlear-scaled entropy was identified for the input neutral speech, and multiplied by the amplitude scaling ratio. Such amplification resulted in placing a large weight on the high-entropy segments, while no weight was applied for the low-entropy segments. Note that due to the resulting speech quality, the amplification of the signal was limited by 50 %. Figure 6.2 demonstrates the entropy-based amplitude modification.

Spectral contour transformation

Following the time-domain approach, spectral-domain modification strategy for the input speech was employed. In this domain, a time-invariant spectral shaping filter was proposed to increase the power in the high-frequency regions of the input speech signal. This proposed

filter was estimated by calculating the frequency difference between overall spectral contours for neutral and Lombard speech style, and, in general has a high-pass characteristics. It has been known in general that high frequency region of the spectrum is more robust in speech intelligibility against noise than low-frequency region (Summers et al., 1988; Lu and Cooke, 2008). Spectral balance of speech also has been affected by the higher vocal effort, resulting in relatively greater intensity in the higher frequency band of the spectrum (Junqua, 1992; Garnier et al., 2010). Note that proposed frequency-domain approach maintained the overall signal power before and after processing. We only redistributed speech energy over frequency using a time-invariant filter.

Here we represent the general steps needed for spectral modification of input speech. The first step was to estimate spectral contours for the two speaking styles. A second order estimate of the spectrum was used to capture the general spectral variation under stress condition in a frame-by-frame basis. The duration of the analysis window was set to 32 msec

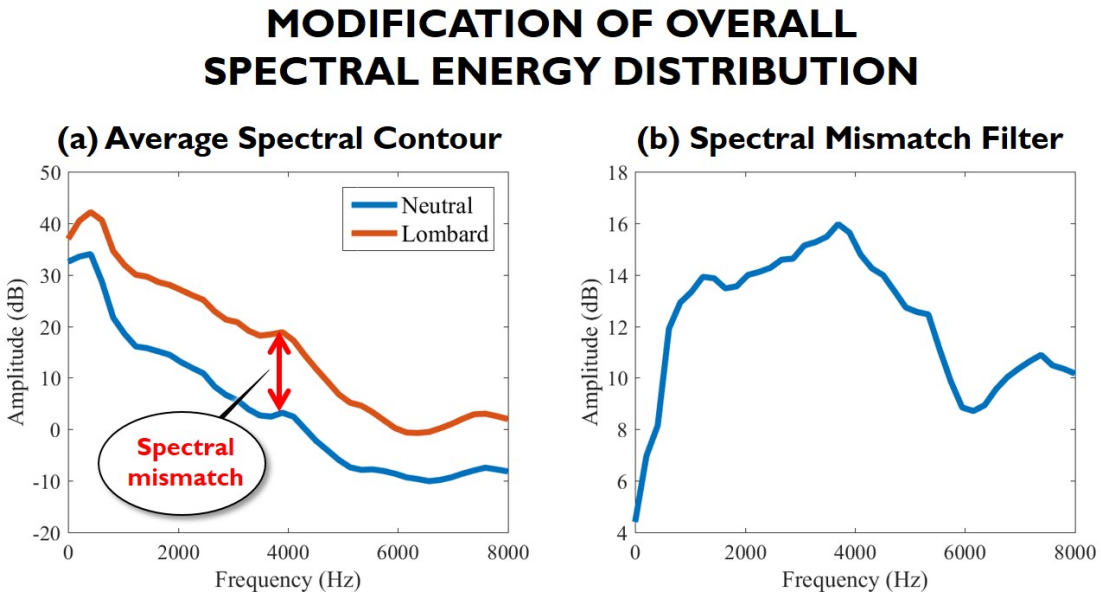


Figure 6.3. Plots showing the spectral contour transformation: (a) aveerge spectral contour for neutral and Lombard speech, and (b) a spectral mismatch filter estimated based on the difference between the two average spectral contours in frequency-domain. The proposed filter was then used to increase the mid- and high-frequency power of input speech.

with a 16 msec skip rate. After the spectral contour estimation, average spectral contour was then computed across the whole sentences of each speaking condition. The spectral mismatch of the two speech styles was then calculated based on the difference between the two average spectral contours in frequency domain. The spectral mismatch filter was then applied to modify overall spectral energy distribution for the input neutral speech. The resulting output signal was root mean squared normalized to match that of the unmodified input neutral signal. Figure 6.3 shows the spectral contour transformation using spectral mismatch filter.

Uniform time stretching

In the last area of modification, uniform time-stretching was employed to further intelligibility gains by listeners in noise. The time-stretching approach resulted in the repetition of speech frames, which could lead listeners have more chances at hearing the speech signal under noise (Bou-Ghazale and Hansen, 1996; Godoy and Stylianou, 2013). The proposed time-scale modification strategy began with modeling duration variations for the neutral and Lombard input speech. These variations were computed as the ratio of Lombard-to-neutral sentence duration. This ratio was then averaged across the whole sentences. The duration of the input neutral sentence was computed and multiplied by the duration scaling ratio. To account for the duration variation, the overlap rate of adjacent speech frames was modified using TD-PSOLA technique (Moulines and Charpentier, 1990). A fixed 32 msec frame size was used with a 50 % overlap between adjacent frames. Note that in this domain, all scaling values greater than 2 were hard limited to the value of 2. This constraint was imposed by the speech quality limitation of the uniform time stretching approach. Figure 6.4 illustrates the uniform time-stretching for lengthening overall sentence duration.

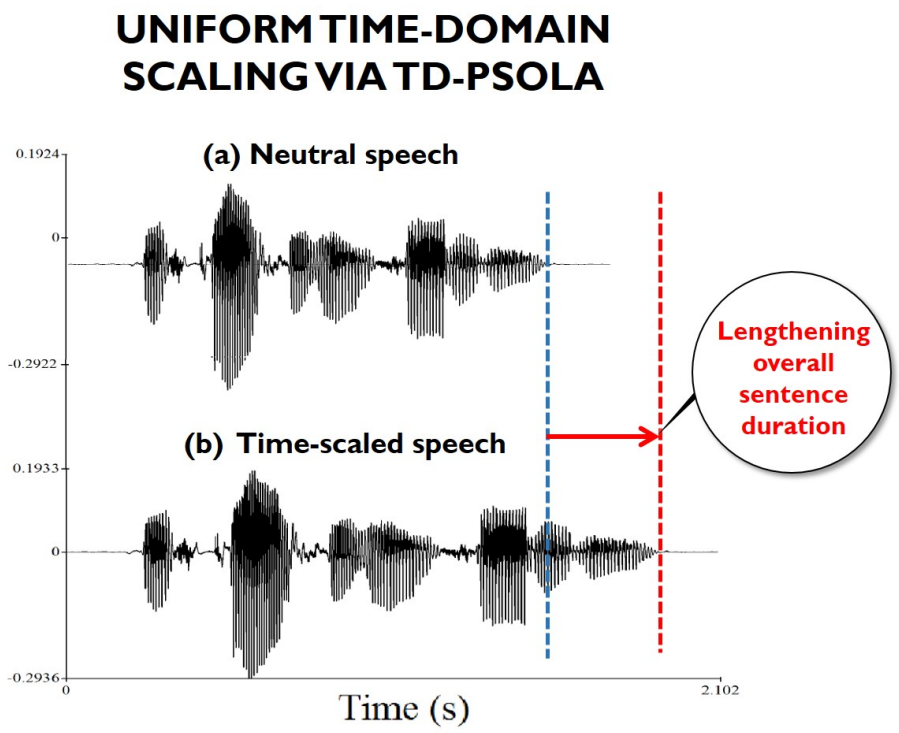


Figure 6.4. Plots illustrating the uniform time-stretching: waveforms for (a) the input neutral sentence, and (b) the output time-scaled sentence. Duration variations for the neutral and Lombard speech were calculated and multiplied to the duration of the input neutral sentence. TD-PSOLA technique (Moulines and Charpentier, 1990) was employed to account for the duration variation.

6.2.2 Evaluation

The final step was to perform a human intelligibility test. The proposed algorithm was evaluated using a formal subjective listening test with cochlear implant listeners. Five post-lingually deaf users of cochlear implant (mean age: 68 years) participated. The cochlear implant user was native speakers of American English and fitted their devices for at least ten months. Detailed biographical information of the cochlear implant participants is presented in Table 6.1

The neutral speech data used in the test was a subset of Lombard speech corpus described in Section 5.2.1. The proposed algorithm was used to modify the speaking style of the input

Table 6.1. Characteristic information of cochlear implant subjects who participated in the perceptual evaluation of Lombard effect.

Listener	Gender	Age (yrs.)	Years of hearing loss	Years implanted	Etiology of hearing loss	Implant ear	Coding strategy
Listener 1	Male	64	6	2	Noise	Bilateral	ACE
Listener 2	Male	78	38	1	Noise	Bilateral	SPEAK
Listener 3	Male	70	17	7	Hereditary	Left only	ACE
Listener 4	Female	69	13	7	Hereditary	Bilateral	ACE
Listener 5	Female	60	30	5	Hereditary	Bilateral	ACE

neutral speech to generate Lombard synthetic speech. The Lombard sentences were mixed with the a maskers at two signal-to-noise ratios. The noise type used here was large-crowd noise. The signal-to-noise ratio levels were 10 dB and 15 dB. Each conditions were comprised of 20 AzBio sentences read by 1 male and 1 female speakers (10 sentences for each speaker). For comparative purpose, neutral and natural Lombard sentences with and without the effect of noise were presented as well.

The listening test was conducted in anechoic sound proof booth. The cochlear implant subjects listened to the stimuli monaurally presented via loud speaker. The speech presentation level was set to 60 dB SPL. The noise level was varied to produce different signal-to-noise, while speech presentation level remained unchanged. Cochlear implant users allowed to listen to each test file only once. Recognition scores were then calculated based on the number of words correctly identified. All clean and noisy conditions were mixed together and presented in random order. None of the sentences were repeated over the entire listening test. Statistical test was performed if any significance of differences between neutral and Lombard modified speech. The statistic for measuring the significance was a repeated-measures

Figure 6.5. Proposed subjective listening test with cochlear implant patients to assess the performance of the proposed algorithm. Data collection was performed in an anechoic sound proof booth. A set-up for data acquisition using loud speaker is demonstrated.

ANOVA. In this study, a difference in means between two groups was considered significant if the significance level fell below 5.0 % ($p < 0.05$). The subjective listening evaluation with cochlear implant patients is illustrated in Figure 6.5.

6.3 Results and Discussion

Figure 6.6 and Table 6.2 show the results for the proposed algorithm and the unprocessed neutral reference in terms of percentage correctly understood words as a function of SNR. Intelligibility scores for naturally produced Lombard speech in noise at 90 dB SPL are also given to for comparison. Improvement in intelligibility was found with Lombard processed speech, particularly in noisy environments. For example, increase of up to approximately + 32 % was achieved for large-crowd noise. The average intelligibility score obtained in quiet condition was 75.3 %, and this intelligibility score dropped down to 47.6 % and 32.0 % in SNR = 15 dB and 10 dB. The decreased scores were, however, improved by up to + 32 % and + 22 % in SNR = 15 dB and 10 dB by employing Lombard effect characteristics via the

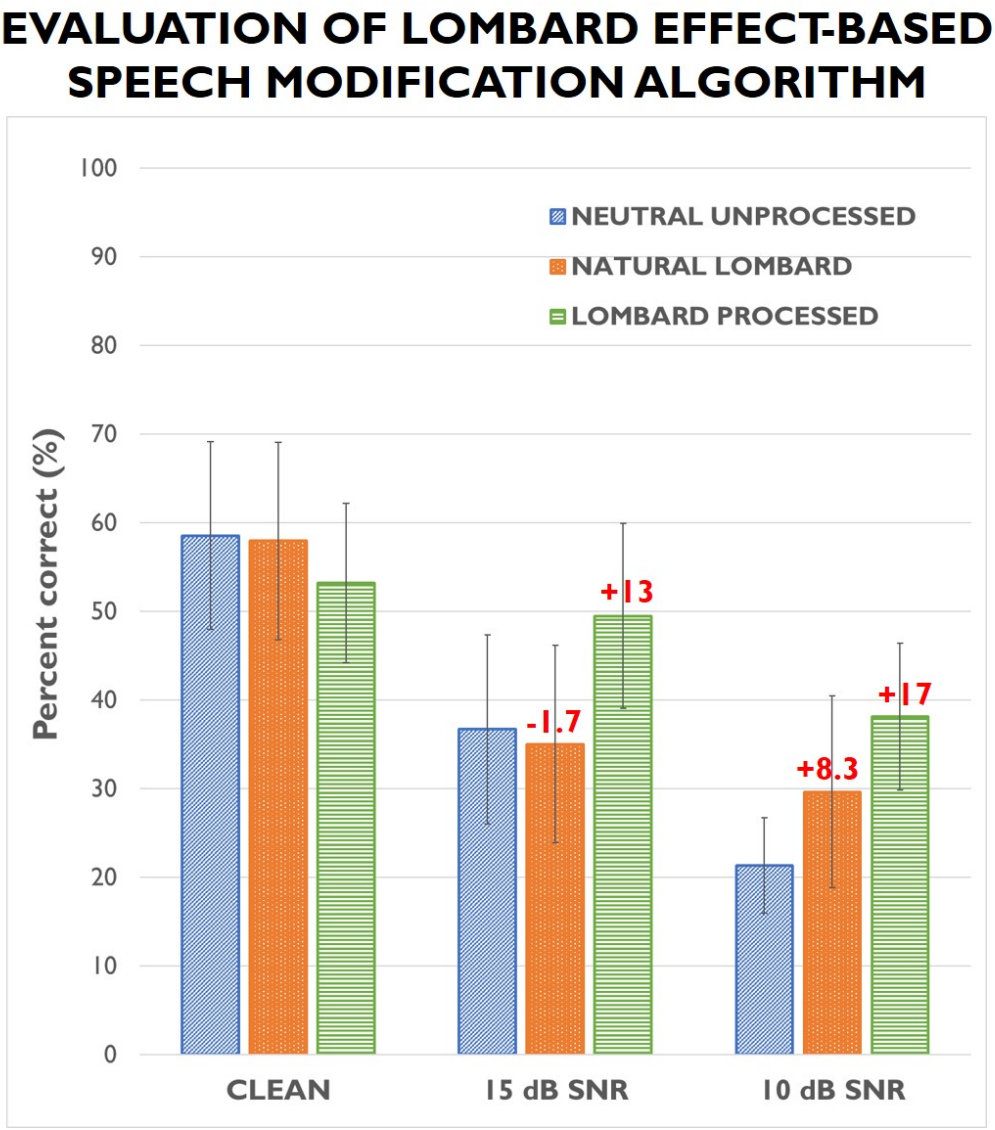


Figure 6.6. Word intelligibility scores for unprocessed neutral, natural Lombard, and processed Lombard speech with five cochlear implant users as a function of signal-to-noise ratio. The Lombard processed speech was generated by modifying the neutral speech via the proposed intelligibility-enhancing algorithm. The neutral unprocessed and natural Lombard speech were obtained by normal-hearing subjects while speaking in quiet and in large-crowd noise at 90 dB SPL respectively.

proposed algorithm. It was found that the proposed method does not increase intelligibility when speech is presented in quiet condition.

Table 6.2. Word recognition scores (%) of five cochlear implant users.

Condition	Unprocessed neutral	Natural Lombard	Lombard processed
Quiet	58.5	57.9	53.2
15 dB SNR	36.7	35.0	49.5
10 dB SNR	21.3	29.6	38.1

Larger increases were achieved for the Lombard processed speech compared to the natural Lombard speech. Intelligibility advantages of the natural Lombard conditions were + 4.2 % and + 11 % in SNR = 15 dB and 10 dB, while listener improved the scores by + 32 % and + 22 % for the Lombard synthetic speech in the same SNR conditions. The largest benefit was measured with the high SNR where intelligibility improved by +32 %. For 15 dB SNR, Lombard processed speech reached up to 79.1 % of word accuracy. This gain is substantial and similar to the levels in quiet environment (*e.g.,* 75.3 % for unprocessed neutral and 79.1 % for natural Lombard).

The improvement in intelligibility was attributed to the modification of speech parameters based on the Lombard effect characteristics via the proposed algorithm. It has been established that speech understanding increased when high-frequency power are amplified at a cost of low frequencies (Niederjohn and Grotelueschen, 1976; Schepker et al., 2013; Jokinen et al., 2016). Spectral change within sentence, for example, vowel-consonant boundaries, has been found to predict speech intelligibility better than individual phoneme classes (*e.g.,*vowel and consonant) (Kewley-Port et al., 2007; Stilp and Kluender, 2010). It has been demonstrated that consonant energy increased at the expense of vowel energy increased speech intelligibility (House et al., 1965; Hansen, 1988) Duration has been considered as one component that is important to speech intelligibility (Bradlow et al., 2003; Hazan and Markham, 2004; Uchanski et al., 1996; Krause and Braida, 2004).

In order to visually assess the effectiveness of the proposed modification scheme, we used example stimulus output patterns (electrodograms). Electrodograms of sentences under neutral and Lombard processed conditions are presented in Figure 6.7. A 22-band Advanced Combination Encoder (ACE) sound processing strategy was used to compute the electrodograms. Figure 6.7(b) and 6.7(d) show that Lombard characteristics imparted on neutral speech in quiet and noise enhanced the segmental characteristics of speech, thus making them more robust against background noise. The enhanced stimulus profile, par-

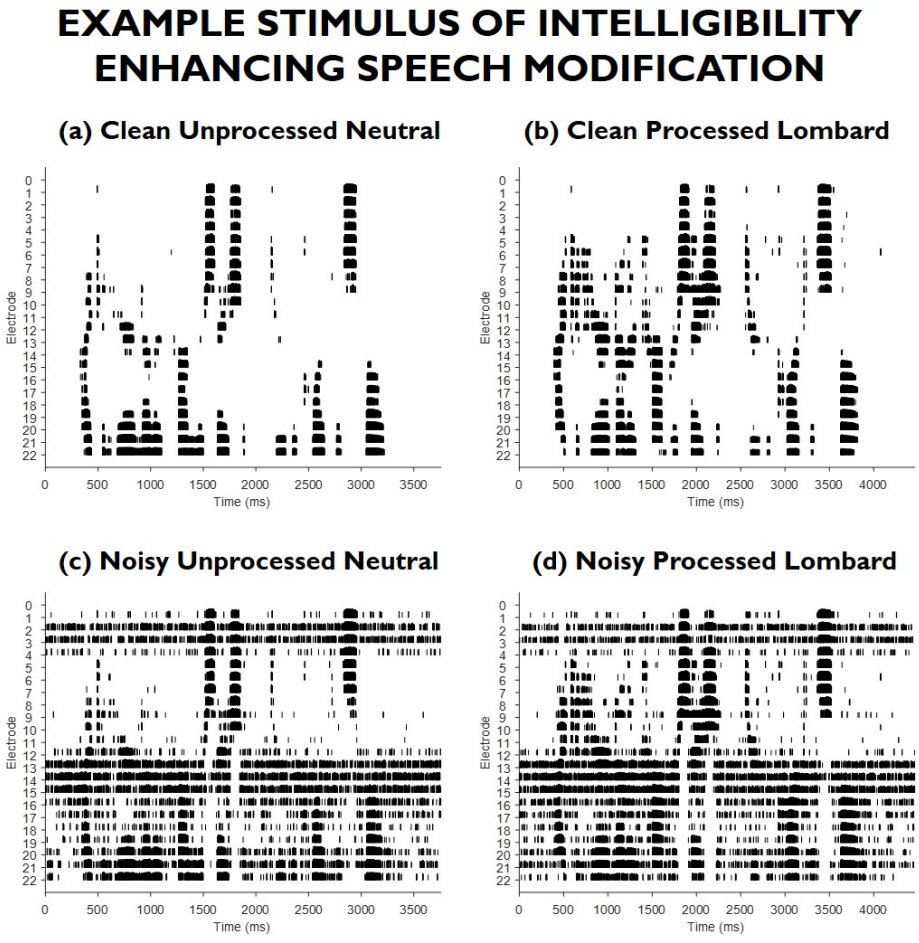


Figure 6.7. Stimulus output patterns (electrodograms) of the sentence "Basketball can be an entertaining sport" from a dataset developed in Section 5.2.1 : (a) original neutral speech reference, (b) Lombard processed speech via the proposed algorithm, (c) neutral speech mixed with speech-shaped noise at 10 dB SNR, and (d) Lombard processed speech mixed with speech-shaped noise at 10 dB SNR.

ticularly in the mid- and high-frequency regions of the electrodogram, suggests improved representation of speech features in Lombard conditions. This was more evident in noisy sentences (Figure 6.7(c) and 6.7(d)). Noise distorts the lower frequency channels more than higher frequency channels, which lead to consonant and formant signal tend to be more emphasized. The results here were highly encouraging a potential of the Lombard based speech modification scheme with perceptual benefit in cochlear implant users under noisy environments.

The data provided here was highly encouraging and serves as a proof of concept for the proposed Lombard-based front-end processing paradigms. For example, synthetically generated Lombard speech introduced Lombard perturbations in neutral speech signal prior to cochlear implant sound coding, which enhances speech characteristics, thereby improving the intelligibility. In addition, further improvement might also be obtained if the proposed modification scheme combined with existing noise suppression algorithm, *e.g.,* single-channel or environmental customization noise reductions (Hazrati and Loizou, 2012b; Wang et al., 2015). Specific knowledge provided here will offer a potential of the Lombard based speech modification approach with perceptual benefit in cochlear implant users in noisy environments.

6.4 Summary: Chapter 6

A new speech modification criterion based on the Lombard effect was proposed in this chapter. This chapter investigated how cochlear implant users perceive the artificially modified speech in challenging listening environments. A previously proposed framework based on Source Generator theory (Hansen, 1994; Hansen and Cairns, 1995) was employed to perturb neutral speech based on Lombard effect modification. The modification areas considered in the proposed algorithm were voice intensity, spectral contour, and overall sentence duration. The results with five cochlear implant listeners indicated improvement in intelligibility when

providing neutral speech which was modified based on Lombard effect properties. Larger increases were achieved for the Lombard processed speech compared to the natural Lombard speech. This improvement was attributed to the modification of speech parameters based on the Lombard effect characteristics via the proposed algorithm. The results provided potential of the Lombard effect based speech modification algorithm with perceptual benefits to cochlear implant users in noisy environments. Algorithmic advancements which modify neutral speech offers a unique opportunity to improve the listening/decoding experiences of cochlear implant users.

CHAPTER 7

CONCLUSION

In this dissertation, we have focused on the acoustic and perceptual role of naturalistic listening environments and its influences on speech production. Lombard effect investigated here was based on the speech of cochlear implant users who are post lingually deafened adults. In addition, the influence of Lombard effect on speech perception of cochlear implant users have been also examined. Finally, a new speech modification criterion based on the Lombard speech characteristics has been proposed in this dissertation.

The role of speech production under noise and its influence on intelligibility has been poorly understood in the field of cochlear implants. Lombard speech research presented in this dissertation may shed light on the underlying mechanisms of speech production in noisy communication settings. The specific knowledge provided in this dissertation provide a crucial step toward better understanding of the nature of Lombard speech and toward applying this knowledge in cochlear implant and other assistive listening devices. The contributions of the dissertation is described next in greater detail.

7.1 Dissertation Contributions

The first contribution of this dissertation was that it developed a new speech corpus that can be used to investigate the speech production in naturalistic environments. Mobile personal audio recordings from continuous single-session audio streams were used to observe this effect over an individual's daily life. Prior advancements in this domain include the "Prof-Life-Log" longitudinal study at UT-Dallas (Sangwan et al., 2012; Ali et al., 2013). In the data collection, speakers produced read and conversational speech over various UT-Dallas on-campus environments (*e.g.,* office and cafeteria). Following the controlled locations, additional audio recordings were collected in uncontrolled environments (outside campus, *e.g.,*

home and restaurants). A set of acoustic and orthographic transcription labels were assigned by a human transcriber. Phoneme-level transcription labels were assigned automatically by forced alignment procedures. A total of 100 hours of personal audio recordings were collected from 6 cochlear implant and 18 normal hearing participants. This collection of speech production provides a unique and unprecedented opportunity to explore real-world listening and speech in diverse environments for CI-to-NH communications.

The second dissertation contribution was that it analyzed speech production of cochlear implant adults with respect to environmental changes. The analysis conducted in this work were based on mobile personal audio recordings collected over various realistic environments on university campus. The parametric variations in vowel, consonant and individual phoneme production were investigated as a function of varying environments. Data from these analyses indicated that Lombard effect was found in speech of post-lingually deafened cochlear implant adults. Speakers demonstrated increased vocal effort, including F0 and speech power, as well as altered glottal spectral slope, and phoneme duration in response to challenging noisy environments. Segmental articulatory movements, for example, F1 for specific phonemes such as /a/, /æ/, /i/, and /u/, also appeared to play an important part in speech production under noise. The significance of the results is that the Lombard effect could potentially help cochlear implant users to ensure/maintain intelligible communication by compensating for the reduced signal-to-noise ratio. Lombard speech research in this research may shed light on the underlying mechanisms of speech production of cochlear implant users.

The third contribution was that it examined the perceptual analysis of Lombard effect by post-lingually deafened cochlear implant users. We analyzed acoustic properties of Lombard and neutral speech from two normal hearing speakers. A speech corpus that is intended for the perceptual experiments of Lombard speech was developed for this analysis. Subjective intelligibility was measured with cochlear implant users across signal-to-noise ratios. Data

from the perceptual analysis suggested Lombard speech had a higher intelligibility than neutral speech for electric hearing in both quiet and noisy environments. An advantage of + 7.6 $\sim$ + 13.2 percentage points in intelligibility for Lombard speech over neutral speech. Larger improvement in intelligibility was found in challenging listening environments. Speech with higher vocal effort were more intelligible than the speech produced with lower vocal effort. These results suggested that the modification of speech production parameters obtained under noise contribute to the Lombard speech intelligibility. The findings presented in this study further our understanding of the speech perception of cochlear implant users.

The fourth contribution of this dissertation was that it proposed a Lombard effect-based speech enhancement algorithm for cochlear implant users as a future work/direction. The proposed algorithm controlled the acoustic parameters of neutral speech to present Lombard speaking style to improve intelligibility in noisy environments. Acoustic variations for neutral and Lombard conditions were modeled and used to generate Lombard synthetic speech. The modification areas considered in the proposed algorithm were voice intensity, overall spectral contour, and sentence duration. The analysis and modeling conducted here was based on a previously established framework, Source Generator theory (Hansen, 1994; Hansen and Cairns, 1995). Subjective listening evaluation was performed with five cochlear implant users to demonstrate the effectiveness of the proposed speech modification algorithm. The results from perceptual analysis provided potential of the Lombard effect based speech modification algorithm with perceptual benefits to cochlear implant users in noisy environments. The specific knowledge provided in this study can apply practical implications for developing speech enhancement algorithms for cochlear implant users.

7.2 Future Work

While many significant advances have been made in the dissertation on Lombard effect, numerous other goals still remain to be achieved. The present study focused on the speech

production of cochlear implant users in varying environment types. However, it does not address the nature and the extent of Lombard effect as compared to the normal hearing listeners. As a part of our future work, we suggest repeating the same data collection with normal hearing individuals in the same environments to establish a one-to-one comparison of CI-NH pair and NH-NH pair.

While present study focused on the role of auditory feedback provided in the context of cochlear implant systems, there has been no study of cochlear implant signal processing parameters that may play a role in auditory feedback. We feel further discussion on other cochlear implant sound processing factors, such as automatic gain control (AGC), adaptive dynamic range optimization (ADRO), frequency band sampling or virtual channels, represent a wider range of issues, which are beyond the current scope of this study. In theory, a supplementary study could explore these various factors within the context of the Lombard effect in future.

A new speech modification criterion based on the Lombard speech characteristics was proposed in this research. The overarching hypothesis of this research was that additional benefit can be obtained when providing neutral speech which was modified based on Lombard effect properties. We expect further improvement might be obtained if the proposed modification scheme combined in conjunction with existing speech enhancement paradigms such as single- (Loizou et al., 2005; Yang and Fu, 2005; Hu and Loizou, 2007) or multi-channel (Hersbach et al., 2013; Wouters and Berghe, 2001; Doclo et al., 2015) or environmental customization (Hu and Loizou, 2010; Goehring et al., 2015; Hazrati et al., 2014) noise reduction algorithms. In addition, while this study focused on Lombard effect speech, our proposed scheme is applicable to a variety of stress and emotional speaking styles, such as clear or loud speech.

The algorithm formulated in Section 6.2.1 will be assessed with cochlear implant users. Testing will be carried out in laboratory environment with original neutral speech which

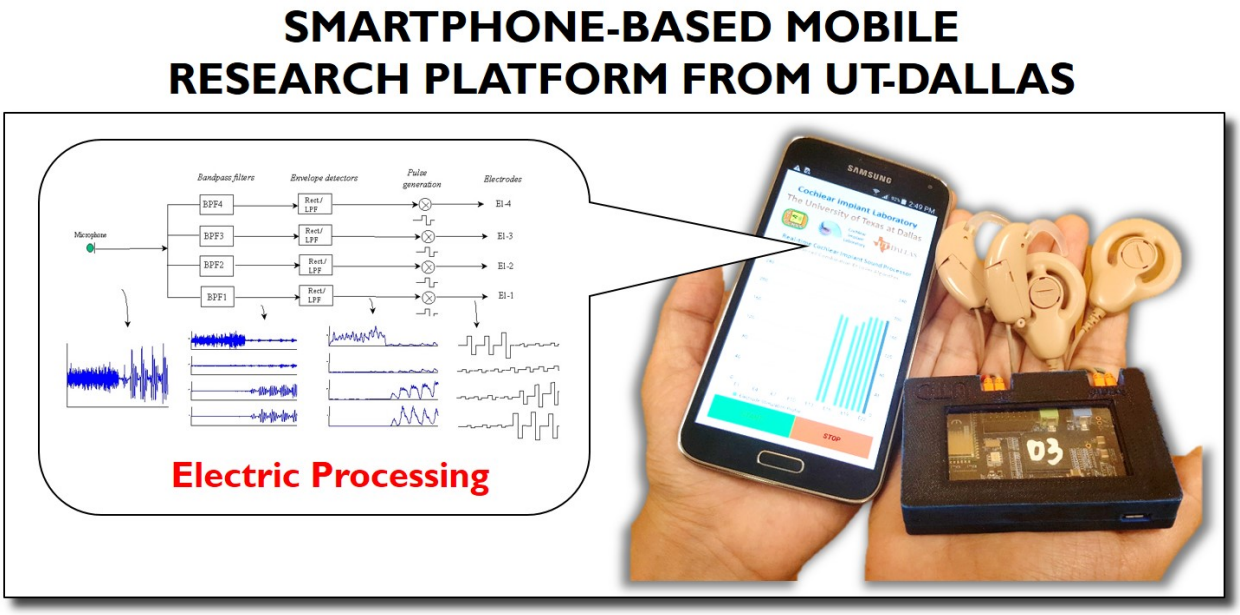


Figure 7.1. Android-based mobile research platform for cochlear implants. This platform offers unique high-performance computing capabilities as well as quick evaluation of approaches in everyday listening environments (Ali et al., 2013; Hong et al., 2015).

are modified by the proposed algorithm. Various types of noise types and signal-to-noise ratios will be considered. At least six cochlear implant users who are postlingually deaf will be recruited and tested. Standard speech intelligibility test will be conducted in a sound proof booth. Speech intelligibility will be measured as a percentage correct score evaluated using AzBio speech corpus (Spahr et al., 2012). The results of neutral speech and Lombard processed speech will be compared in various environments.

In addition to conventional assessment paradigm, algorithm will be implemented in real-time in the portable unit of CCI-Mobile platform (Ali et al., 2013; Hong et al., 2015) for field evaluation. CCI-Mobile platform is a research interface built on consumer electronics (*e.g.*, smart phone, tablet PC, etc.) and has potential to provide researcher with state-of-the art processing capabilities of signal processing for cochlear implants (See Figure 7.1). During field testing, the participants will be asked to wear the platform in naturalistic environments and provide their feedback. Custom apps will be used that integrate rating system which will

enable the users to record/submit their feedback for the proposed algorithm on the go using the mobile platform. Real-time implementation and testing in naturalistic environments will enable true assessment of the proposed technique to quantify changes in performance levels.

APPENDIX

SLIDES FOR ORAL EXAMINATION

Lombard Effect in Speech Production by Cochlear Implant Users: Analysis, Assessment and Implications

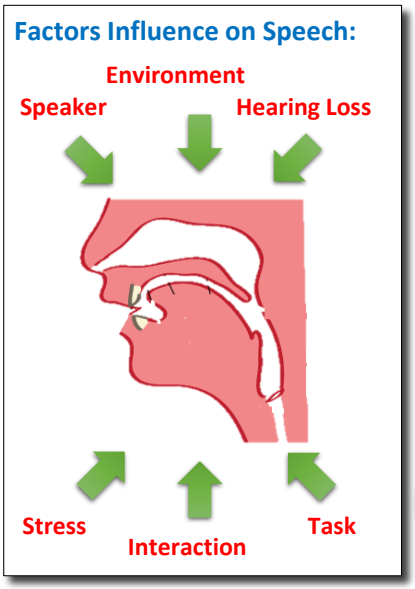
Ph.D THESIS DEFENSE

Jaewook Lee
March 20, 2017

Center for Robust Speech Systems -
Cochlear Implant Laboratory (CRSS-CIL)
Department of Electrical Engineering
The University of Texas at Dallas

Email: jaewook@utdallas.edu

External Factors which Influence on Speech Production



- Changes in speech production based on *auditory feedback* are an important research domain.
- Factors influence on speech:
  - **Speaker:** *age, gender, nativeness, emotion*
  - **Environment:** *noise, reverberation, distance*
  - **Conversation:** *read, spontaneous*
  - **Task:** *stress, workload, inter-personal interaction*

This study focused on *"Acoustic noisy environment"*

Lombard Effect

- **Lombard effect** - involuntary response a speaker experiences in the *presence of noise* (Lombard, 1911).
- Cause increased *vocal effort* including:
  - Vocal intensity
  - Pitch period structure
  - Formant characteristics
  - Glottal spectral slope
  - Speech rate, etc.

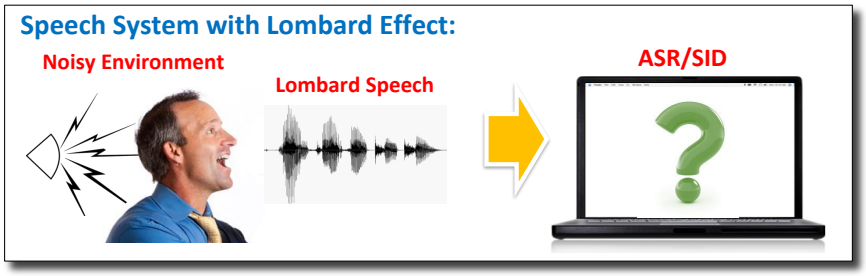
(Lane & Tranel, 1971; Hansen, 1988; Garnier *et al*., 2010; Junqua, 1992)

- Helps to maintain *intelligible communication* in challenging listening environments.

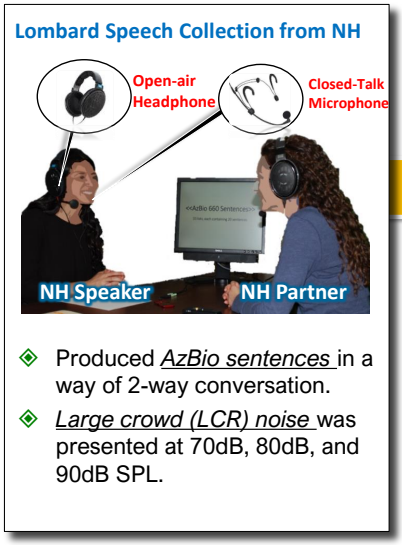
Lombard Effect

- Also cause to degrade *automatic speech system (ASR/SID)* performance when system is modeled with neutral speech.
- Several signal processing technique have been proposed to *compensate for LE* in speech to improve the performance of the speech systems (Hansen and Varadarajan, 2009; Hansen, 1994).

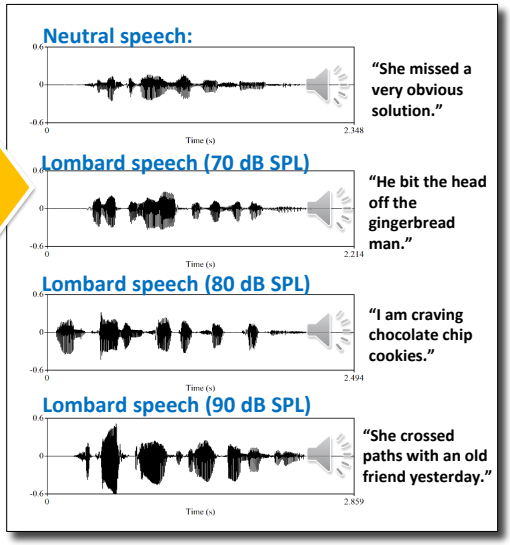
Well studied in *normal hearing (NH)* listeners and *speech systems*, but not in the field of *cochlear implants (CI).*



Example "Lombard" Speech

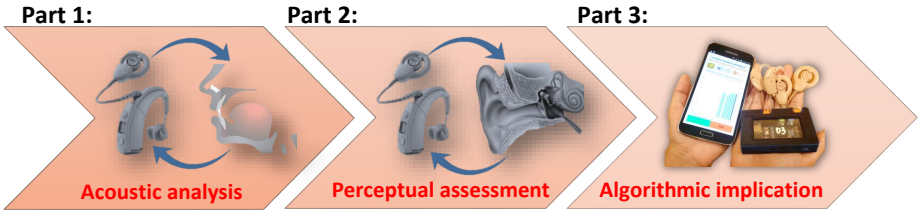


- Produced *AzBio sentences* in a way of 2-way conversation.
- *Large crowd (LCR) noise* was presented at 70dB, 80dB, and 90dB SPL.



Thesis Outline

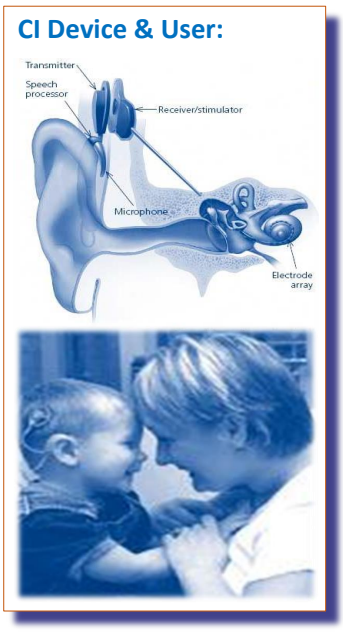
- **Background:**
  - Fundamentals of cochlear implants (CI)
- **Proposed research:**
  - **Part 1***:* *Acoustic* ***analysis*** of Lombard effect by CI users
  - **Part 2**: *Perceptual* ***assessment*** of Lombard speech by CI users
  - **Part 3**: *Algorithmic* ***implication*** of Lombard effect for CI users
- **Thesis contributions**



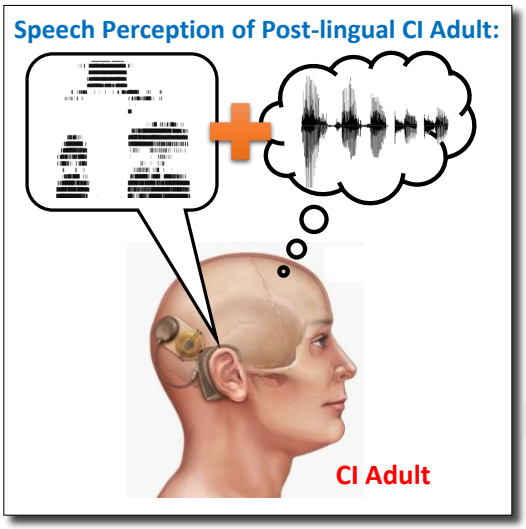
Cochlear Implants

- **Cochlear implant (CI)** - electronic device that is surgically implanted in *the inner ear*.
- Directly stimulates the basilar membrane to provide *partial sense of sound* (Loizou, 1999; Zeng *et al*.,2008).
- Children and adults who are *deaf* or *severely hard-of-hearing* can be fitted for this device.
- The field of cochlear implants has experienced a considerable growth over the past few decades (*324,000 users worldwide* (NIDCD, 2011)).

This study focused on *post-lingually deaf CI users*.



Cochlear Implants – Post-lingual Deaf Users

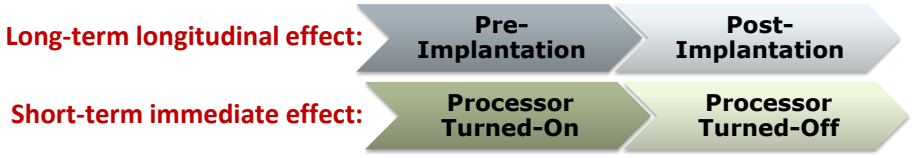


- **Post-lingual deaf -** adults who lost hearing after the age of 18.
- Able to learn *how to associate* the signal provided by an implant with the sound they remember.
- Allows to *identify speech* without lip-reading or sign language.

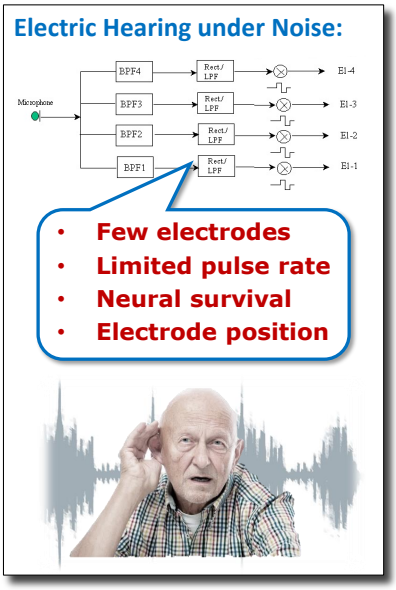
Only limited studies have been performed in the area of *speech production of CI adults*.

Cochlear Implants – Speech Production of CI Adults

- **Long-term longitudinal study:**
  - Restored auditory feedback after CI is a crucial factor for *increased adult speech perception* (Bilger *et al*.,1977).
  - *Improved quality of speech production* when compared to corresponding quality before single channel implantation (Hochmair-Desoyer *et al*., 1981).
- **Short-term immediate study:**
  - Rapid change in *formant frequencies* when turning speech processor either on or off in short-time (Svirsky and Tobey, 1991).
  - Immediate response for many speech parameters (*F0, vowel duration, etc.*) to short-time constraints (within few seconds or less) (Svirsky *et al*., 1992).



Cochlear Implants – Performance in Noisy Environments



- CI listener perform well in quiet environments (>95% WRR); but their *performance degrades significantly by background noise* (limited temporal/spectral resolution).
- Proposed numerous *speech enhancement algorithms*:
  - Single- and multi-channel (Dolco, 2005; Loizou, 2005)
  - Channel selection (Kim, 2010; Hazrati, 2013)
  - Environmental optimization (Hu, 2010; Goehring, 2015)

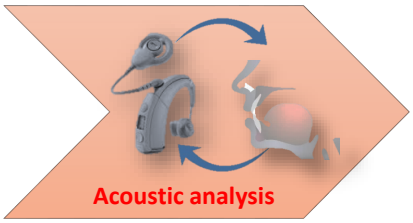
Mostly focused on *noise suppression strategies*.

Thesis Contributions

- **Contribution #1** – Development of a speech corpus to investigate speech production in naturalistic environments.
- **Contribution #2 -** Acoustic analysis of speech production by CI users with respect to noisy environmental changes.
- **Contribution #3 -** Perceptual analysis of Lombard effect by CI users in noisy environments.
- **Contribution #4 -** Development of a Lombard effect-based speech enhancement algorithm for CI users.

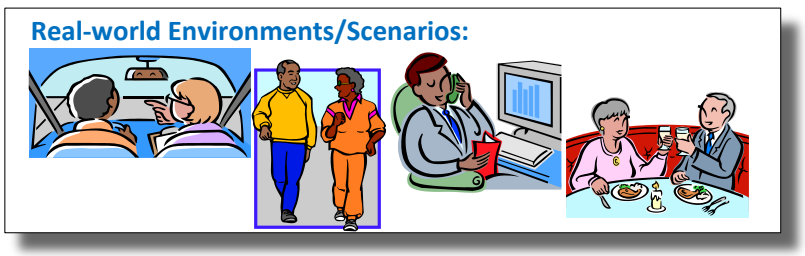
Part 1

Effect of environmental noise on speech production of CI users: a naturalistic study

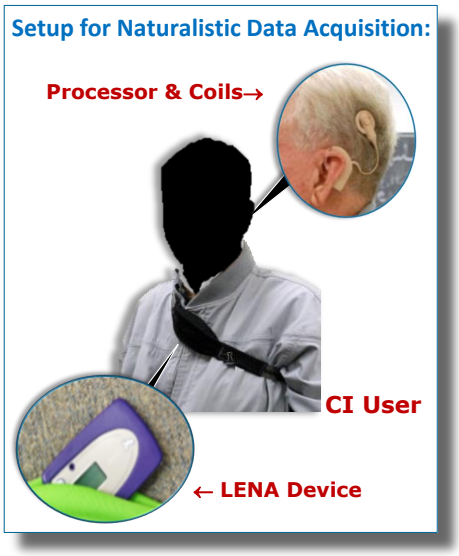


Objectives

- **Problem statement:**
  - Not all of the Lombard research goals have been achieved to date.
  - It is still unknown if CI users employ ***Lombard effect*** in noisy conditions.
- **Goal 1** - Analysis of *speech production of CI users* with respect to diverse environment conditions.
- **Goal 2** - Investigate the effect of auditory feedback on speech production in *everyday naturalistic environments*.



UTD-CI-LENA Corpus Development



- **Personal audio recording -** *A long single-session* audio stream (8-16 hrs.).
- Collected over an individual's *daily environments/scenarios*.
- **LENA unit** – A portable digital audio recorder capable of up to *16 hrs. of continuous audio* (LENA Foundation, 2014).

Prior advancement in this domain include the "***Prof-Life-Log***" *longitudinal study* at UT-Dallas (Ziaei *et al*., 2012; Ziaei *et al*., 2013)

Subjects

- Six *post-lingual deaf CI users* (mean age: 65 yrs.) participated and produced read and spontaneous speech.
- The same number of *NH speakers* (mean age: 37 yrs.) were participated as a conversational partner.
- IEEE sentences and a list of general topics to converse were given to participants (*e.g., sport, news, food, etc.).*

Biographical Information for CI Participants:

Speaker	Gender	Age (yrs)	Years of hearing loss	Years implanted	Etiology of hearing loss	Implant ear	Sound coding strategy
SPK1	Female	61	56	11	Hereditary	Bilateral	ACE
SPK2	Female	52	48	5	Hereditary	Bilateral	ACE
SPK3	Female	61	14	4	Hereditary	Bilateral	ACE
SPK4	Male	67	12	6	Hereditary	Left only	ACE
SPK5	Male	81	55	9	Hereditary	Bilateral	ACE
SPK6	Male	71	18	4	Unknown	Bilateral	ACE

Controlled On-Campus Environments

Six Locations on College Campus:

1. Office/Lab

Stationary (PC Fan)
1-5 People
1/10 Stationarity

2. Hallway

Impulsive (foot, door)
5-10 People
3/10 Stationarity

3. Lobby

Reverb + Impulsive
10-50 People
5/10 Stationarity

4. Outside

Babble + Wind
30-50 People
5/10 Stationarity

5. Cafeteria

Babble + Competing Talker
50-200 People
9/10 Stationarity

6. Gameroom

Babble + Music + Impulse
35-75 People
9/10 Stationarity

- Selected 6 *naturalistic environments* on UT-Dallas college campus.
- Noise conditions (*type, mixture, levels*) vary greatly across the conditions.

Note! - *Office environment* will be used as the quiet baseline (≤ 45 dB SPL).

Analysis of Noise & Listening Environments

- **Long-term Averaged Spectra (LTAS)**
- **SNR with Neutral speech (SNRN):**
  - Energy ratio of *neutral speech* in quiet baseline to *noise energy* in each environment, assumed to be without Lombard effect.

$$SNRN = 10\log_{10}\left(\frac{E_{Neutral}}{E_{noise}}\right) \quad (1)$$

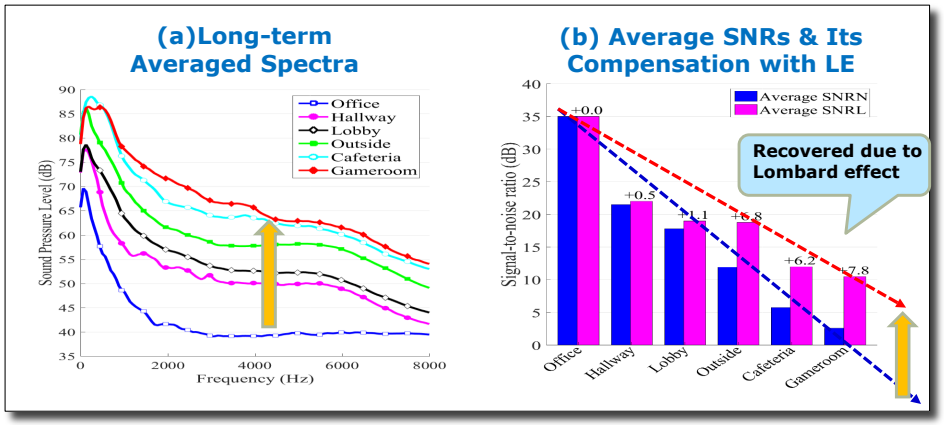
Fixed for every conditions

- **SNR with Lombard speech (SNRL):**
  - Energy ratio between *Lombard speech* in each environment and *corresponding background noise*, assumed to be with Lombard effect.

$$SNRL = 10\log_{10}\left(\frac{E_{Lombard}}{E_{noise}}\right) \quad (2)$$

Speech and background noise *boundary detections* were used to separate speech from noise.

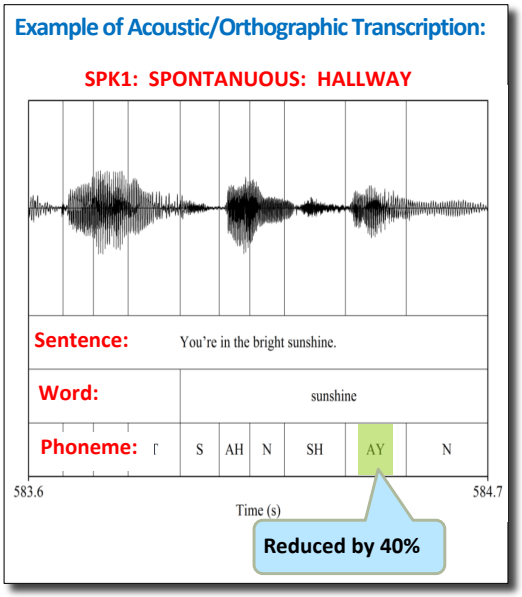
Results – Analysis of Noise & Listening Environments



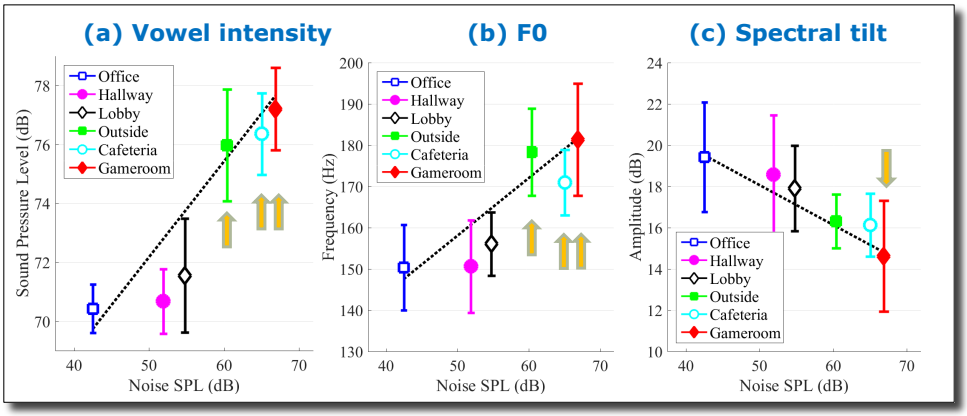
- Average spectral energy *gradually increased*
- SNRN also *decreased* from 27 dB down to -3dB (Gameroom)
- SNRN *recovered due to LE* by up to +11.3 dB (Gameroom).

Analysis of Speech

- Measured *speech production parameters* via PRAAT (Boersma, 2002):
  - **Vocal intensity (I)**
  - **Fundamental frequency (F0)**
  - **Overall spectral tilt (ST)**
  - **C-V Intensity Ratio (CVIR)**
  - **C-V Duration Ratio (CVDR)**
  - **First formant frequency (F1)**
  - **Second formant frequency (F2)**
- *Phoneme-level transcription* labels were automatically assigned by forced alignment tool (Yuan, 2008).
- The beginning and ending of each phoneme were reduced by 20% to eliminate any *transitional effect.*

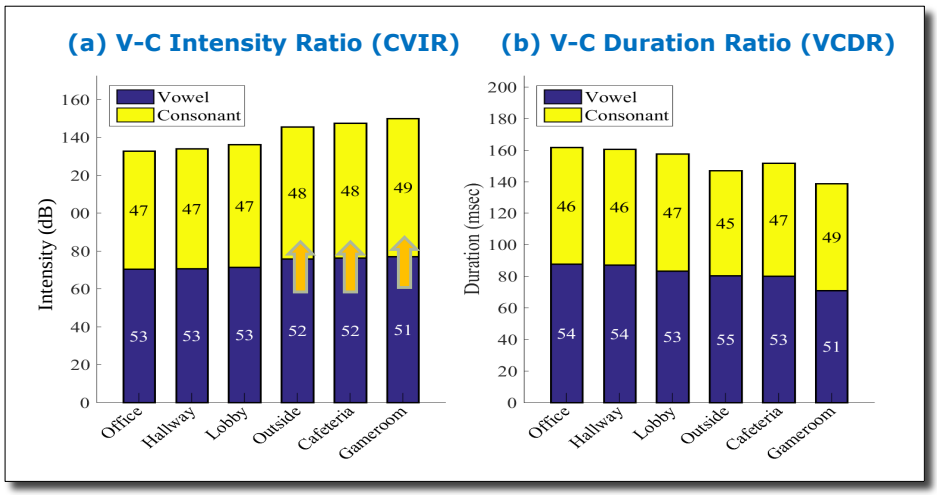


Results – Vowel Production



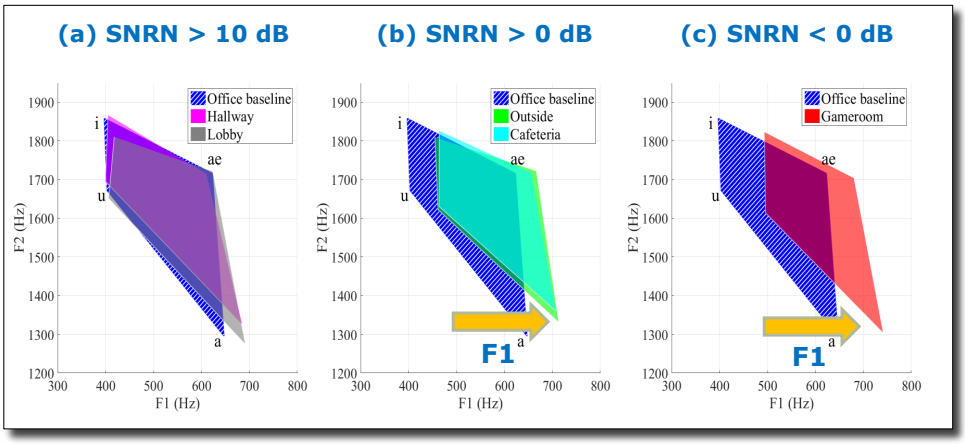
- I, F0 *increased significantly* (Out, Cafe, Game)
- ST *decreased significantly* (Game) - Higher spectral energy emphasized

Results – V-C Ratios



- Consonant intensity *significantly increased* (Out, Cafe, Game).
- Known that consonant is more important to intelligibility.
- *No significant shift* in duration between classes.

Results – Vowel Space



- Significant changes in F1 (/a/,/ae/,/i/,/u/) (<10 dB SNR)
- Only little changes in F2 (/a/,/ae/,/i/,/u/)

Results – Summary of Lombard Parameter Analysis

- Grouped into two areas: *low* (SNRN >10dB) and *high* (SNRN ≤10dB) noise conditions.
- Provide summary of important *Lombard effect parameters* by fixing two conditions.
- Employed Analysis of Variance (*ANOVA*) for significance test.
  - *I, F0, CVIR* ($F[1,40] > 52, p < 0.001$)
  - *ST, DUR, F1* ($F[1,40] > 9, p < 0.05$)
  - *CVDR, F2* ($F[1,40] < 2, p > 0.05$)

In line with the result of NH studies; indicating LE exist.

(*Lane, 1971; Jansen, 1988; Junqua, 1992; Garnier, 2010*).

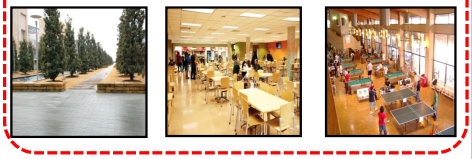
Multiple Naturalistic Environments:

Low noise group (> 10 dB SNR):

1. Office/Lab
2. Hallway
3. Lobby

High noise group (≤ 10 dB SNR):

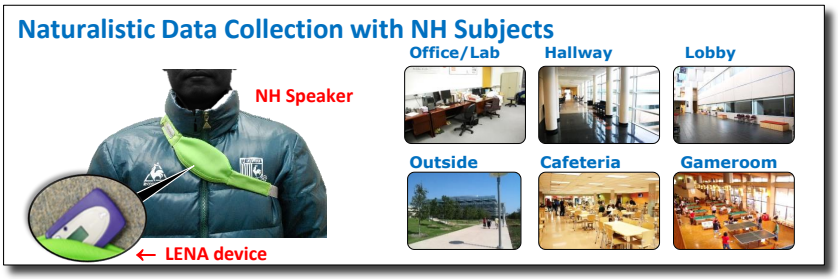
4. Outside
5. Cafeteria
6. Gameroom



Pairwise Comparison of CI vs NH

Still unknown how CI speech is different from those of NH in noisy conditions.

- Repeated the *same data collection* with six NH lage-mates (mean age: 57 yrs) at the same locations on UT-Dallas campus.
- Measured the *same speech parameters* used in the CI analysis.
- Performed *ANOVA test* to determine statistical difference between two speaker groups.



Results – Pairwise Comparison of CI vs NH

PAIRWISE COMPARISON OF CI VS NH

Acou. Param.		ANOVA
Vowel	Intensity	
	F0	*↑
	Spec. tilt	
	Duration	

Phon. Param.		ANOVA
F1	/a/	**↑
	/æ/	
	/i/	
	/u/	

Phon. Param.		AVOVA
F2	/a/	
	/æ/	
	/i/	**↑
	/u/	

Significance levels: *<0.001, **<0.01, and *<0.05.**

- Most CI parameters have *similar pattern* changes with NH results.
- Some CI parameters (F0, F1 for /a/, F2 for /i/) are *significantly different* from NH.
- These differences may be due to *partial restoration* of auditory feedback by CI device (*e.g.,* poor in pitch and formant perception)

Summary & Discussion

- ***Lombard effect*** has been found in the speech production of CI users who are post-lingually deaf adults.
- Auditory feedback interacts with both *suprasegmental* (I, F0, ST) and *segmental properties* (CVIR, CVDR, F1, F2) in naturalistic context.
- Demonstrated acoustic and phonetic feature patterns in Lombard speech which are *similar with NH listeners*.
- These consistency mainly due to the *presence of auditory feedback* provided by a CI device, allowing more nearly typical pattern of speech in response to noise.

Research Opportunity – Part 1

UTD-CI-LENA Corpus:

Controlled (24 hrs):

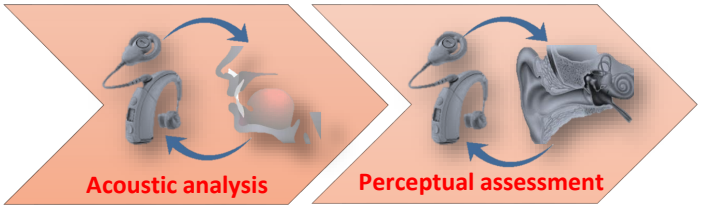
Take Home (75 hrs):

- A total of 24 hours of *personal audio recordings* were collected from 24 speakers (6 CI and 18 NH) users while participated in college campus.
- Additional 75 hours of *naturalistic audio* were collected from everyday situation (e.g., home, restaurant, store, etc.).
- A set of *acoustic and orthographic transcription labels* were assigned by a human annotator.
- Potential application areas for CI users:
  - Language acquisition/development
  - Speech-related disorder
  - User/environmental customized coding algorithm

(Gilkerson and Richards, 2008)

Part 2

Influences of Lombard effect on speech intelligibility in CI users



Objectives

- **Problem Statement:**
  - No study has been performed to if LE influences *speech perception performance* for CI users.
  - There are very limited data concerning how CI users respond to the *different speaking styles* (*e.g.,* whisper, soft, loud, etc.)
- **Goal 1 -** Examine the influence of *Lombard effect* on speech perception of post-lingually deaf CI users.
- **Goal 2 -** Investigate how *perception performance differs* from speech produced in various noisy environments.

Perceptual Difference of Speaking Styles:

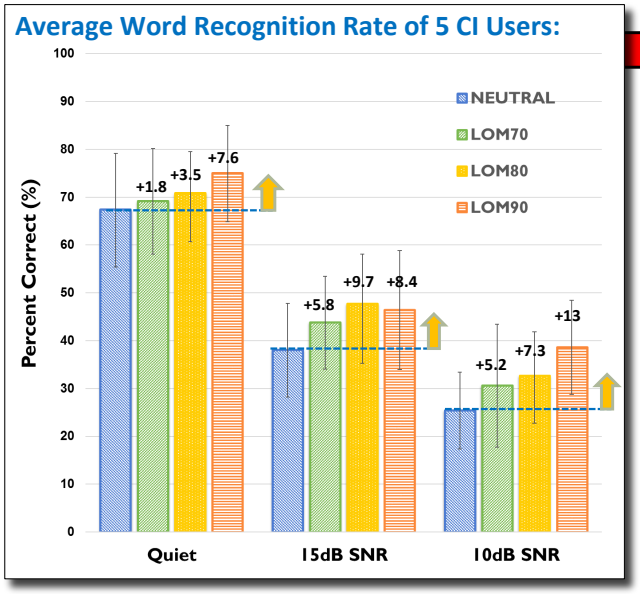
Corpus Formulation with NH Speakers

- Four *NH speakers* participated to read AzBio sentences (Spahr, 2012) in sound recording booth.
- *Speech partners* were seated 1m in front to speakers and listen to the speech.
- *Large crowd (LCR) noise* were presented via open-air headphone at 70dB, 80dB and 90 dB SPL.

Subjective Evaluation with CI Listeners

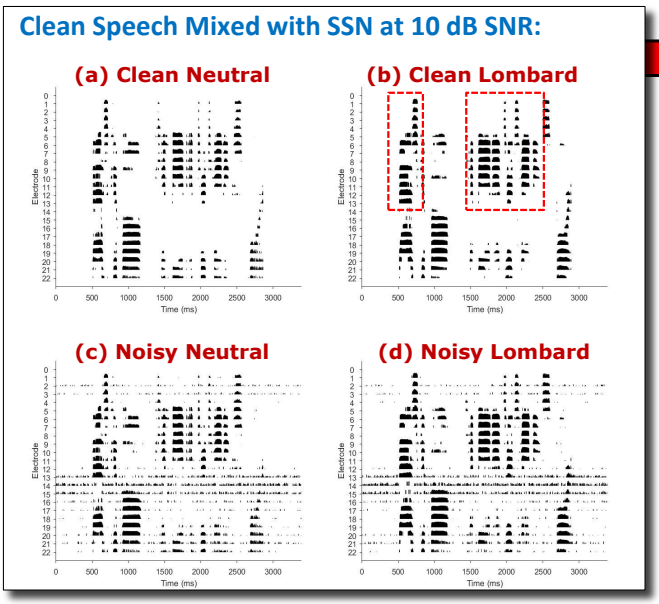
- Five *post-lingual CI users (mean age: 64 yrs.)* participated.
- Original clean stimuli were *corrupted by large crowd noise* at 10 dB and 15 dB SNR.
- CI subjects listened to the stimuli presented via *loud speakers*.
- *Recognition score* was calculated based on the number of words identified.

Results – Intelligibility of Lombard Speech



- *Improvement in intelligibility* when Lombard speech was presented.
- Larger improvement in *challenging listening environment* (10 dB vs 15 dB SNR)
- *Higher vocal effort speech* (Lom 90) are more intelligible than lower vocal effort speech (Lom 70).

Results – Example Stimuli Output Pattern



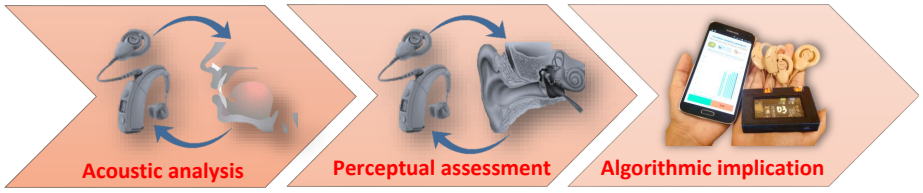
- LE speech provided increase in electrical activity on *high frequency channels* (1-16).
- Cause *consonant and formant signal* tend to be more enhanced.
- More evident in noisy sentences (noise *distorts the lower frequency* channels more).

Summary & Discussion

- *Acoustic and perceptual characteristics* of speech under LE were analyzed.
- Lombard speech had a *higher intelligibility* than neutral speech for electric hearing in both quiet and noisy environments.
- The improvement was larger in *challenging listening environments*.
- The modification of speech *using LE* might contribute to the higher intelligibility.

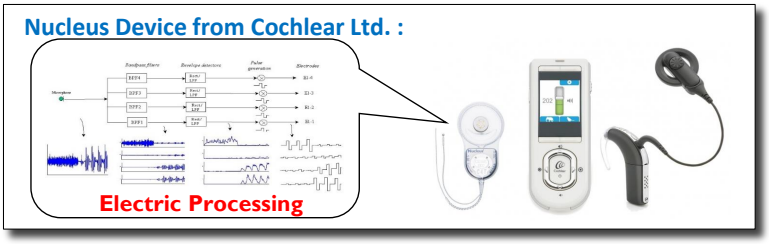
Part 3

Development of an intelligibility enhancement algorithm based on Lombard effect properties



Objectives

- **Problem Statement:**
  - Conventional signal processing for CI have focused on *noise suppression strategies* to improve performance in noisy conditions (not perceptually motivated).
  - It would be desirable to have an algorithmic modification of neutral speech based on *speech production constraints* (*e.g.,* LE, Clear).
- **Goal 1 –** Develop a novel *intelligibility enhancing speech modification algorithm* based on LE properties.
- **Goal 2 –** Examine how CI users perceive the *artificially modified Lombard speech* in noisy environments.

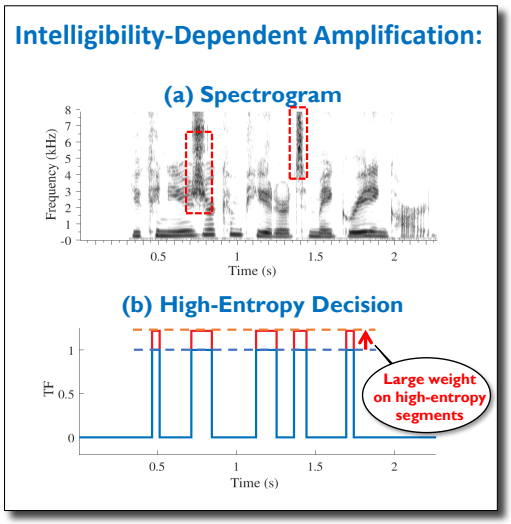


Proposed Algorithm – System Description

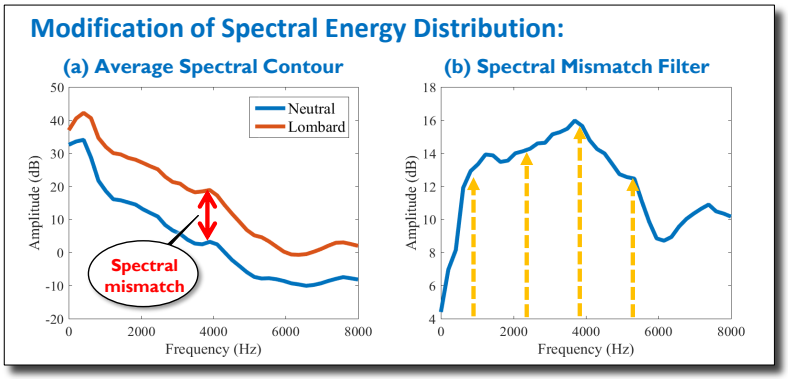
- Controlled the acoustic features of *input neutral speech* to present *Lombard speaking-style output*.
- The *modification areas* considered here were:
  1. ***Temporal amplification***
  2. ***Overall spectral contour***
  3. ***Sentence duration***
- Modeled *acoustic variation* for neutral and Lombard speech.
- Dataset for parameter modeling was derived from *UT-Scope stressed speech corpus* (Ikeno, 2007).

Feature 1 - Entropy-Based Temporal Amplification

- Aim to increase amplitude for high intelligibility segments (*e.g., consonants, V-C boundaries*).
- *A cochlear-scaled entropy* (Stilp and Kluender, 2010) was used to estimate the information bearing segments.
- Place a large weight on *high-entropy segments*.

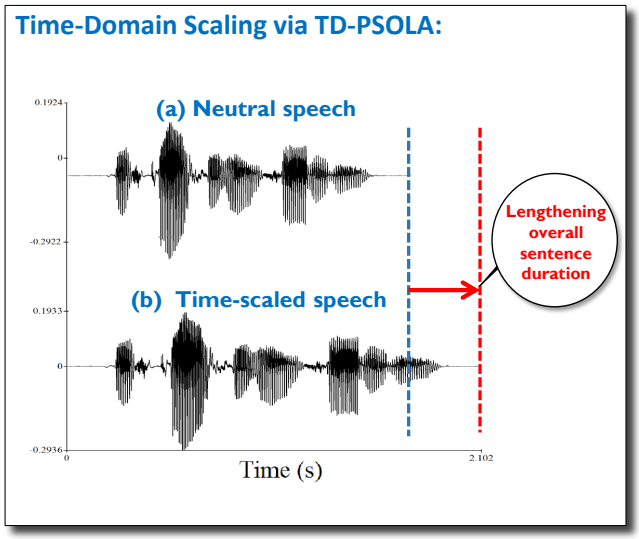


Feature 2 – Spectral Contour Transformation



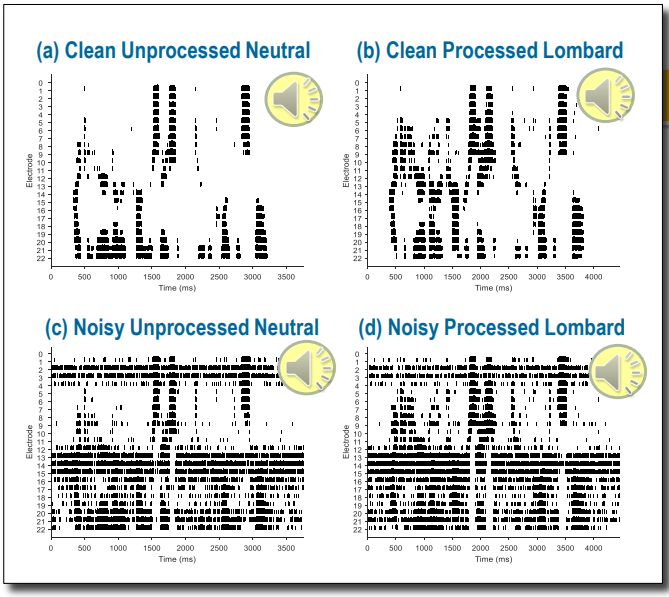
- Designed to increase *high frequency structure*; important for *channel selection* by CI platform.
- A time-invariant filter was estimated by computing *spectral difference* between neutral and Lombard speech.
- Used to adjust *overall spectral balance* of the input signal.

Feature 3 – Uniform Time Stretching



- Lengthening speech duration allowed listeners have *more chances at hearing*; improve CI decoding opportunity
- *TD-PSOLA technique* (Moulines and Charpentier, 1990) was employed to account for the duration variation.

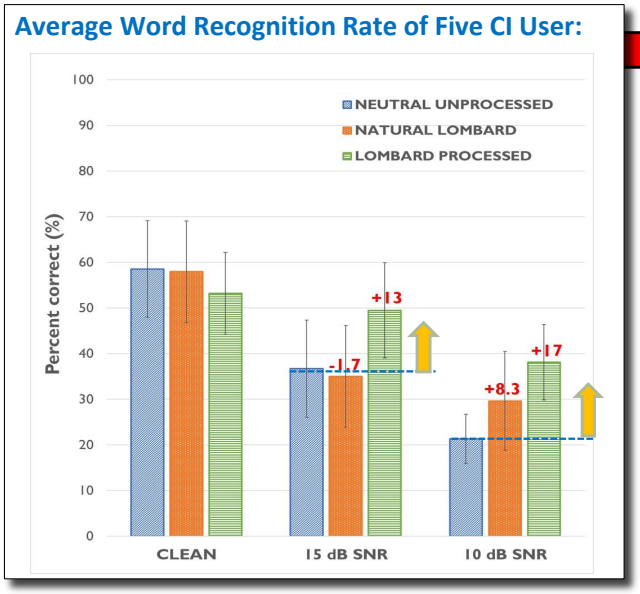
Example "Artificial Lombard" Sentences



Evaluated by CI users:

- Five CI subjects (mean age of 68) listened to the *original clean stimuli* corrupted by large crowd noise.
- *Natural Lombard sentence* were also presented for comparison.

Results – Subjective Evaluation with CI Users



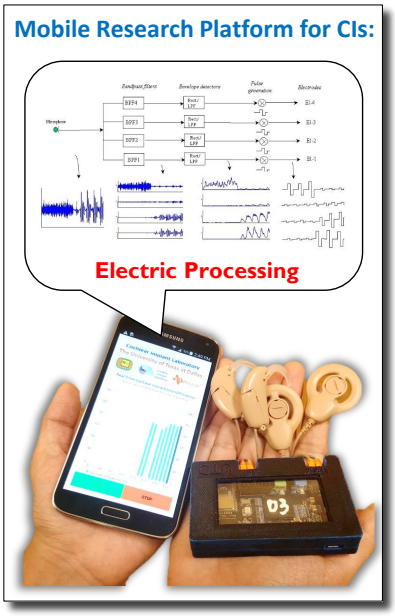
- Increases in intelligibility was found for Lombard processed speech, *particularly in noisy environments*.
- Larger increases were achieved for *Lombard processed speech* when compared to Natural Lombard speech.
- Largest benefit, *± 17%*, was measured in 10 dB SNR.

Summary & Discussion

- A new speech modification criterion using *Lombard effect characteristics* was proposed.
- *Improvement in intelligibility* was found with Lombard processed speech in noisy environments.
- The improvement in intelligibility was attributed to the modification of speech *via the proposed algorithm*.
- The result provided potential of the Lombard effect based *speech enhancement algorithm* for CI users.

Research Opportunity – Part 2, 3

- In contrast to the traditional methods for noise suppression, the proposed approach *motivated by natural human speech physiology*.
- This paradigm is more natural, thus, *desirable for hearing impaired individuals*.
- Algorithmic advancement proposed here offers a unique opportunity to *improve the listening/decoding experience* of CI users.



Thesis Contributions

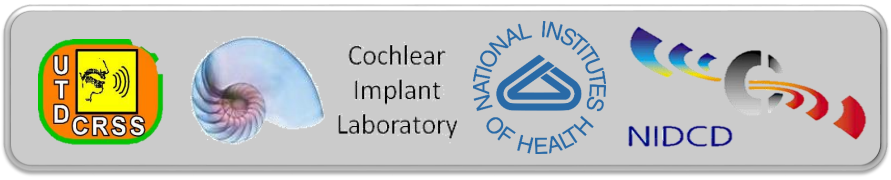
- **Contribution #1** – Development of a speech corpus to investigate speech production in naturalistic environments.
- **Contribution #2 -** Acoustic analysis of speech production by CI users with respect to noisy environmental changes.
- **Contribution #3 -** Perceptual analysis of Lombard effect by CI users in noisy environments.
- **Contribution #4 -** Development of a Lombard effect-based speech enhancement algorithm for CI users.

Publications

- **Journal papers**
  - **Jaewook Lee**, Hussnain Ali, Ali Ziaei, Emily A. Tobey, John H.L. Hansen, “Acoustic-phonetic analysis of speech communication with cochlear implant users under noisy Lombard environments: a naturalistic study,” *Journal of the Acoustical Society of America* (Accepted).
  - Oldooz Hazrati, **Jaewook Lee**, Philipos C. Loizou, "Blind binary masking for reverberation suppression in cochlear implants," *Journal of the Acoustical Society of America, Vol. 133 (3), pp. 1607~1614, March, 2013.*
- **Conference papers**
  - **Jaewook Lee**, Hussnain Ali, John H.L. Hansen, “Intelligibility enhancement of neutral speech based on Lombard effect modification with application to cochlear implant users," *in Proc. Annual Midwinter Meeting of Association for Research in Otolaryngology (ARO 2017)*, Baltimore, MD, February, 2017
  - **Jaewook Lee**, Hussnain Ali, John H.L. Hansen, “The Lombard reflex and its influences on speech perception in adult cochlear implant users,” in Proc. CI 2016, Toronto, Canada, May, 2016.
  - **Jaewook Lee**, Hussnain Ali, Ali Ziaei, Emily A. Tobey, John H.L. Hansen, "Impact analysis of naturalistic environmental noise type on speech production for cochlear implant users versus normal hearing listeners," *in Proc. CIAP 2015*, Lake Tahoe, CA, USA, July, 2015.
  - **Jaewook Lee**, Hussnain Ali, Ali Ziaei, John H.L. Hansen, “Analysis of speech and language communication for cochlear implant users in noisy Lombard conditions,” *in Proc. IEEE ICASSP 2015*, Brisbane, Australia, April, 2015.
  - **Jaewook Lee**, Hussnain Ali, Ali Ziaei, John H.L. Hansen, “Lombard effect based speech analysis across noisy environments for voice communications with cochlear implant subjects,” *in Proc. 168th Meeting of Acoustical Society of America*, Indianapolis, USA, October, 2014.

Acknowledgments

- COMMITTEE
  - Dr. John H. L. Hansen, Chair
  - Dr. Peter F. Assmann
  - Dr. P. K. Rajasekaran
  - Dr. Carlos Busso
- CRSS-CIL Members
- CI & NH Subjects who participated in the studies
- NIH for their project support



REFERENCES

Ali, H., A. P. Lobo, and P. C. Loizou (2013). Design and evaluation of a personal digital assistant-based research platform for cochlear implants. *IEEE transactions on biomedical engineering 60*(11), 3060–3073.

Bilger, R. C., F. O. Black, and N. T. Hopkinson (1977). Research plan for evaluating subjects presently fitted with implanted auditory prostheses. *Annals of Otology, Rhinology & Laryngology. Supplement 86*, 21–24.

Blamey, P., P. Arndt, F. Bergeron, G. Bredberg, J. Brimacombe, G. Facer, J. Larky, B. Lindström, J. Nedzelski, A. Peterson, et al. (1996). Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants. *Audiology and Neurotology 1*(5), 293–306.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot International 5*, 341–345.

Bořil, H. and J. H. Hansen (2010). Unsupervised equalization of lombard effect for speech recognition in noisy adverse environments. *IEEE Transactions on Audio, Speech, and Language Processing 18*, 1379–1393.

Bou-Ghazale, S. E. and J. H. Hansen (1996). Generating stressed speech from neutral speech using a modified celp vocoder. *Speech Communication 20*(1), 93–110.

Bou-Ghazale, S. E. and J. H. L. Hansen (2000). A comparative study of traditional and newly proposed features for recognition of speech under stress. *IEEE Transactions on Speech and Audio Processing 8*, 429–442.

Bradlow, A. R., N. Kraus, and E. Hayes (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research 46*(1), 80–97.

Cooke, M., C. Mayo, and C. Valentini-Botinhao (2013). Intelligibility-enhancing speech modifications: the hurricane challenge. In *Proceedings of Annual Conference of the International Speech Communication Association (INTERSPEECH)*. Lyon, France, 3572–3576.

Denes, P. B. and E. Pinson (1993). *The speech chain*. Macmillan. 246 pgs.

Doclo, S., W. Kellermann, S. Makino, and S. E. Nordholm (2015). Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones. *IEEE Signal Processing Magazine 32*(2), 18–30.

Donaldson, G. S. and S. L. Allen (2003). Effects of presentation level on phoneme and sentence recognition in quiet by cochlear implant listeners. *Ear and Hearing 24*(5), 392–405.

Dorman, M. and A. Spahr (2006). *Speech perception by adults with multichannel cochlear implants in Cochlear Implants (2nd edition)*. Thieme Medical Publishers, New York, NY. 193-204.

Dorman, M. F., M. T. Hannley, K. Dankowski, L. Smith, and G. McCandless (1989). Word recognition by 50 patients fitted with the symbion multichannel cochlear implant. *Ear and Hearing 10*(1), 44–49.

Dowell, R. C., D. J. Mecklenburg, and G. M. Clark (1986). Speech recognition for 40 patients receiving multichannel cochlear implants. *Archives of Otolaryngology–Head & Neck Surgery 112*(10), 1054–1059.

Doyle, J. H., J. B. DOYLE, and F. M. TURNBULL (1964). Electrical stimulation of eighth cranial nerve. *Archives of Otolaryngology 80*(4), 388–391.

Dreher, J. J. and J. O'Neill (1957). Effects of ambient noise on speaker intelligibility for words and phrases. *The Journal of the Acoustical Society of America 29*(12), 1320–1323.

Eddington, D. K. (1980). Speech discrimination in deaf subjects with cochlear implants. *The Journal of the Acoustical Society of America 68*(3), 885–891.

Ferguson, S. H. and D. Kewley-Port (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America 112*(1), 259–271.

Friesen, L. M., R. V. Shannon, D. Baskent, and X. Wang (2001). Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America 110*(2), 1150–1163.

Fu, Q.-J., R. V. Shannon, and X. Wang (1998). Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing. *The Journal of the Acoustical Society of America 104*(6), 3586–3596.

Garnier, M., N. Henrich, and D. Dubois (2010). Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language, and Hearing Research 53*, 588–608.

Garofolo, J. S., L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue (1993). Timit acoustic-phonetic continuous speech corpus. *Linguistic data consortium, Philadelphia 33*.

Gilkerson, J. and J. A. Richards (2008). The LENA foundation natural language study. Technical report, LENA Research Foundation, Boulder, CO.

Godoy, E., M. Koutsogiannaki, and Y. Stylianou (2014). Approaching speech intelligibility enhancement with inspiration from lombard and clear speaking styles. *Computer Speech & Language 28*(2), 629–647.

Godoy, E. and Y. Stylianou (2013). Increasing speech intelligibility via spectral shaping with frequency warping and dynamic range compression plus transient enhancement. In *Proceedings of Annual Conference of the International Speech Communication Association (INTERSPEECH)*. Lyon, France, 3572–3576.

Goehring, T., F. Bolner, J. J. Monaghan, B. van Dijk, J. Wouters, M. Moonen, and S. Bleeck (2015). Neural network based speech enhancement applied to cochlear implant coding strategies. *The Journal of the Acoustical Society of America 138*(3), 1833–1833.

Google Inc (2014). Google Glass. *Web page, retrieved [Sept. 2014] from https://www.google.com/glass/start/*.

Hanley, T. and M. Steer (1949). Effect of level of distracting noise upon speaking rate, duration and intensity. *Journal of Speech and Hearing Disorders 14*(4), 363–368.

Hansen, J. H. and D. A. Cairns (1995). Icarus: Source generator based real-time recognition of speech in noisy stressful and lombard effect environments. *Speech Communication 16*(4), 391–422.

Hansen, J. H. L. (1988). *Analysis and compensation of stressed and noisy speech with application to robust automatic recognition*. Ph. D. thesis, Georgia Institute of Technology, Atlanta, GA, USA. 428 pgs.

Hansen, J. H. L. (1994, 10/01). Morphological constrained feature enhancement with adaptive cepstral compensation (MCE-ACC) for speech recognition in noise and Lombard effect. *IEEE Transactions on Speech and Audio Processing 2*, 598–614.

Hansen, J. H. L. (1996). Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition. *Speech Communication 20*, 151–173.

Hansen, J. H. L. and L. M. Arslan (1995). Robust feature-estimation and objective quality assessment for noisy speech recognition using the credit card corpus. *IEEE Transactions on Speech and Audio Processing 3*, 169–184.

Hansen, J. H. L. and V. Varadarajan (2009). Analysis and compensation of Lombard speech across noise type and levels with application to in-set/out-of-set speaker recognition. *IEEE Transactions on Audio, Speech, and Language Processing 17*, 366–378.

Hanson, H. M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *The Journal of the Acoustical Society of America 101*, 466–481.

Hart, B. and T. R. Risley (1995). *Meaningful differences in the everyday experience of young American children.* Paul H Brookes Publishing, Baltimore, MD. 268 pgs.

Hazan, V. and D. Markham (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *The Journal of the Acoustical Society of America 116*(5), 3108–3118.

Hazrati, O., J. Lee, and P. C. Loizou (2013). Blind binary masking for reverberation suppression in cochlear implants. *The Journal of the Acoustical Society of America 133*(3), 1607–1614.

Hazrati, O. and P. C. Loizou (2012a). The combined effects of reverberation and noise on speech intelligibility by cochlear implant listeners. *International journal of audiology 51*(6), 437–443.

Hazrati, O. and P. C. Loizou (2012b). Tackling the combined effects of reverberation and masking noise using ideal channel selection. *Journal of Speech, Language, and Hearing Research 55*(2), 500–510.

Hazrati, O., S. O. Sadjadi, and J. H. Hansen (2014). Robust and efficient environment detection for adaptive speech enhancement in cochlear implants. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Florence, Italy, 900–904.

Healy, E. W., S. E. Yoho, Y. Wang, and D. Wang (2013). An algorithm to improve speech recognition in noise for hearing-impaired listeners. *The Journal of the Acoustical Society of America 134*(4), 3029–3038.

Hersbach, A. A., D. B. Grayden, J. B. Fallon, and H. J. McDermott (2013). A beamformer post-filter for cochlear implant noise reductiona). *The Journal of the Acoustical Society of America 133*(4), 2412–2420.

Hochberg, I., A. Boothroyd, M. Weiss, and S. Hellman (1992). Effects of noise and noise suppression on speech perception by cochlear implant users. *Ear and Hearing 13*(4), 263–271.

Hochmair-Desoyer, I. J., E. S. Hochmair, K. Burian, and R. E. Fischer (1981). Four years of experience with cochlear prostheses. *Medical progress through technology 8*, 107–119.

Hong, F., H. Ali, J. H. Hansen, and E. Tobey (2015). Android-based research platform for cochlear implants. In *Proceedings of Conference on Implantable Auditory Prostheses (CIAP)*. Lake Tahoe, CA, pp127.

House, A. S., C. E. Williams, M. H. Hecker, and K. D. Kryter (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set. *The Journal of the Acoustical Society of America 37*, 158–166.

Hu, Y. and P. C. Loizou (2007). A comparative intelligibility study of single-microphone noise reduction algorithms. *The Journal of the Acoustical Society of America 122*(3), 1777–1786.

Hu, Y. and P. C. Loizou (2010). Environment-specific noise suppression for improved speech intelligibility by cochlear implant users. *The Journal of the Acoustical Society of America 127*(6), 3689–3695.

Huber, J. E. and B. Chandrasekaran (2006). Effects of increasing sound pressure level on lip and jaw movement parameters and consistency in young adults. *Journal of Speech, Language, and Hearing Research 49*, 1368–1379.

Ikeno, A., V. Varadarajan, S. Patil, and J. H. Hansen (2007). Ut-scope: Speech under lombard effect and cognitive stress. In *Proceedings of IEEE Aerospace Conference*. Big Sky, MT, 1–7.

Iseli, M., Y.-L. Shue, and A. Alwan (2007). Age, sex, and vowel dependencies of acoustic measures related to the voice sourcea). *The Journal of the Acoustical Society of America 121*, 2283–2295.

Jokinen, E., U. Remes, and P. Alku (2016). The use of read versus conversational lombard speech in spectral tilt modeling for intelligibility enhancement in near-end noise conditions. In *Proceedings of Annual Conference of the International Speech Communication Association (INTERSPEECH)*. San Francisco, CA, 2771–2775.

Junqua, J.-C. (1992). The Lombard reflex and its role on human listeners and automatic speech recognizers. *The Journal of the Acoustical Society of America 93*, 510–524.

Junqua, J.-C. (1996, 11). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. *Speech Communication 20*, 13–22.

Kewley-Port, D., T. Z. Burkle, and J. H. Lee (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listenersa). *The Journal of the Acoustical Society of America 122*(4), 2365–2375.

Kim, G. and P. C. Loizou (2010). Improving speech intelligibility in noise using environment-optimized algorithms. *IEEE transactions on audio, speech, and language processing 18*(8), 2080–2090.

Kirk, K. and B. Edgerton (1983). The effects of cochlear implant use on voice parameters. *Otolaryngologic Clinics of North America 16*, 281–292.

Krause, J. C. and L. D. Braida (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America 115*, 362–378.

Krishnamurthy, N. and J. H. L. Hansen (2009). Babble noise: modeling, analysis, and applications. *IEEE Transactions on Speech and Audio Processing 17*, 1394–1407.

Lane, H., M. L. Matthies, F. H. Guenther, M. Denny, J. S. Perkell, E. Stockmann, M. Tiede, J. Vick, and M. Zandipour (2007). Effects of short-and long-term changes in auditory feedback on vowel and sibilant contrasts. *Journal of Speech, Language, and Hearing Research 50*, 913–927.

Lane, H. and B. Tranel (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech, Language, and Hearing Research 14*, 677–709.

Lane, H. and J. W. Webster (1991). Speech deterioration in postlingually deafened adults. *The Journal of the Acoustical Society of America 89*(2), 859–866.

Leder, S. B., J. B. Spitzer, and J. C. Kirchner (1987). Speaking fundamental frequency of postlingually profoundly deaf adult men. *Annals of Otology, Rhinology and Laryngology 96*, 322–324.

LENA Foundation (2014). LENA research foundation. *Web page, retrieved [Sept. 2014] from www.lenafoundation.org/*.

Lindblom, B. (1990). *Explaining phonetic variation: A sketch of the H & H theory*. Springer. 403–439.

Loizou, P. C. (1998). Mimicking the human ear. *IEEE signal processing magazine 15*(5), 101–130.

Loizou, P. C. (1999). Introduction to cochlear implants. *IEEE Transactions on Engineering in Medicine and Biology Society 18*, 32–42.

Loizou, P. C. (2013). *Speech enhancement: theory and practice*. CRC press, Boca Raton, FL. 69–93.

Loizou, P. C., A. Lobo, and Y. Hu (2005). Subspace algorithms for noise reduction in cochlear implants. *The Journal of the Acoustical Society of America 118*(5), 2791–2793.

Lombard, E. (1911). Le signe de lelevation de la voix. *Annales des Maladies de l'Oreille, du Larynx, du Nez et du Pharynx 37*, 101–119.

Lu, Y. and M. Cooke (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America 124*, 3261–3275.

Lu, Y. and M. Cooke (2009). Speech production modifications produced in the presence of low-pass and high-pass filtered noise. *The Journal of the Acoustical Society of America 126*, 1495–1499.

Matthies, M. L., M. Svirsky, J. Perkell, and H. Lane (1996). Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant. *Journal of Speech, Language, and Hearing Research 39*, 936–946.

Moore, B. C. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology 9*(4), 399–406.

Moulines, E. and F. Charpentier (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech communication 9*(5-6), 453–467.

Müller, J., F. Schon, and J. Helms (2002). Speech understanding in quiet and noise in bilateral users of the med-el combi 40/40+ cochlear implant system. *Ear and hearing 23*(3), 198–206.

NIDCD (2012). Cochlear implants. *Web page, retrived [Sept. 2014] from https://www.nidcd.nih.gov/health/hearing/pages/coch.aspx*.

Niederjohn, R. and J. Grotelueschen (1976). The enhancement of speech intelligibility in high noise levels by high-pass filtering followed by rapid amplitude compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing 24*(4), 277–282.

Oller, D. K., P. Niyogi, S. Gray, J. A. Richards, J. Gilkerson, D. Xu, U. Yapanel, and S. F. Warren (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. In *Proceedings of National Academy of Sciences of the United States of America (PNAS)*. USA, **107**, 13354-13359.

Patterson, R., I. Nimmo-Smith, J. Holdsworth, and P. Rice (1987). An efficient auditory filterbank based on the gammatone function. In *Proceedings of A Meeting of the IOC Speech Group on Auditory Modelling at RSRE*, Volume 2.

Paxton, A., K. Rodriguez, and R. Dale (2015). Psyglass: Capitalizing on google glass for naturalistic data collection. *Behavior research methods 47*(3), 608–619.

Payton, K. L., R. M. Uchanski, and L. D. Braida (1994). Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *The Journal of the Acoustical Society of America 95*, 1581–1592.

Perkell, J. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics 25*, 382–407.

Perkell, J., M. Matthies, H. Lane, F. Guenther, R. Wilhelms-Tricarico, J. Wozniak, and P. Guiod (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication 22*, 227–250.

Picheny, M. A., N. I. Durlach, and L. D. Braida (1986). Speaking clearly for the hard of hearing iiacoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research 29*(4), 434–446.

Pickett, J. M. (1956). Effects of vocal force on the intelligibility of speech sounds. *The journal of the acoustical society of america 28*(5), 902–905.

Pisoni, D. B., R. H. Bernacki, H. C. Nusbaum, and M. Yuchtman (1985). Some acoustic-phonetic correlates of speech produced in noise. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Tampa, FL, 1581-1584.

Pittman, A. L. and T. L. Wiley (2001). Recognition of speech produced in noise. *Journal of Speech, Language, and Hearing Research 44*(3), 487–496.

Rostolland, D. and C. Parant (1973). Distorsion and intelligibility of shouted voice. In *Proceedings of Symposium Speech Intelligibility, Liège*, pp. 293–304.

Sangwan, A., A. Ziaei, and J. H. L. Hansen (2012). ProfLifeLog: Environmental analysis and keyword recognition for naturalistic daily audio streams. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Kyoto, Japan, 4941-4944.

Schepker, H. F., J. Rennies, and S. Doclo (2013). Improving speech intelligibility in noise by sii-dependent preprocessing using frequency-dependent amplification and dynamic range compression. In *Proceedings of Annual Conference of the International Speech Communication Association (INTERSPEECH)*. Lyon, France, 3577–3581.

Simmons, F. B. (1966). Electrical stimulation of the auditory nerve in man. *Archives of otolaryngology 84*(1), 2–54.

Skinner, M. W., L. K. Holden, T. A. Holden, M. E. Demorest, and M. S. Fourakis (1997). Speech recognition at simulated soft, conversational, and raised-to-loud vocal efforts by adults with cochlear implants. *The Journal of the Acoustical Society of America 101*(6), 3766–3782.

Skinner, M. W., L. K. Holden, T. A. Holden, R. C. Dowell, P. M. Seligman, J. A. Brimacombe, and A. L. Beiter (1991). Performance of postlinguistically deaf adults with the wearable speech processor (wsp iii) and mini speech processor (msp) of the nucleus multi-electrode cochlear implant. *Ear and Hearing 12*(1), 3–22.

Sodersten, M., S. Ternstrom, and M. Bohman (2005). Loud speech in realistic environmental noise: phonetogram data, perceptual voice quality, subjective ratings, and gender differences in healthy speakers. *Journal of Voice 19*, 29–46.

Spahr, A. J., M. F. Dorman, L. M. Litvak, S. Van Wie, R. H. Gifford, P. C. Loizou, L. M. Loiselle, T. Oakes, and S. Cook (2012). Development and validation of the azbio sentence lists. *Ear and hearing 33*(1), 112.

Spahr, A. J., M. F. Dorman, and L. H. Loiselle (2007). Performance of patients using different cochlear implant systems: effects of input dynamic range. *Ear and Hearing 28*(2), 260–275.

Stilp, C. E. and K. R. Kluender (2010). Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility. *Proceedings of the National Academy of Sciences (PNAS) 107*(27), 12387–12392.

Summers, W. V., D. B. Pisoni, R. H. Bernacki, R. I. Pedlow, and M. A. Stokes (1988, 09/01). Effects of noise on speech production: acoustic and perceptual analyses. *The Journal of the Acoustical Society of America 84*, 917–928.

Svirsky, M. A., H. Lane, J. S. Perkell, and J. Wozniak (1992). Effects of short-term auditory deprivation on speech production in adult cochlear implant users. *The Journal of the Acoustical Society of America 92*, 1284–1300.

Svirsky, M. A., A. M. Robbins, K. I. Kirk, D. B. Pisoni, and R. T. Miyamoto (2000). Language development in profoundly deaf children with cochlear implants. *Psychological science 11*, 153–158.

Svirsky, M. A. and E. A. Tobey (1991). Effect of different types of auditory stimulation on vowel formant frequencies in multichannel cochlear implant users. *The Journal of the Acoustical Society of America 89*, 2895–2904.

Tong, Y., G. Clark, P. Seligman, and J. Patrick (1980). Speech processing for a multiple-electrode cochlear implant hearing prosthesis. *The Journal of the Acoustical Society of America 68*, 1897–1899.

Tye-Murray, N., B. Gantz, F. Kuk, and R. Tyler (1988). Word recognition performance of patients using three different cochlear implant designs. In *Proceedings of International Conference of Cochlear Implants*.

Tyler, R. S. (1988). Open-set word recognition with the 3m/vienna single-channel cochlear implant. *Archives of Otolaryngology–Head & Neck Surgery 114*(10), 1123–1126.

Tyler, R. S., P. Abbas, N. Tye-Murray, B. J. Gantz, J. F. Knutson, B. F. Mccabe, C. Lansing, C. Brown, G. Woodworth, J. Hinrichs, et al. (1988). Evaluation of five different cochlear implant designs: audiologic assessment and predictors of performance. *The Laryngoscope 98*(10), 1100–1106.

Uchanski, R. M., S. S. Choi, L. D. Braida, C. M. Reed, and N. I. Durlach (1996). Speaking clearly for the hard of hearing ivfurther studies of the role of speaking rate. *Journal of Speech, Language, and Hearing Research 39*(3), 494–509.

Vandali, A. E., L. A. Whitford, K. L. Plant, G. M. Clark, et al. (2000). Speech perception as a function of electrical stimulation rate: using the nucleus 24 cochlear implant system. *Ear and Hear. 21*, 608–624.

Vick, J. C., H. Lane, J. S. Perkell, M. L. Matthies, J. Gould, and M. Zandipour (2001). Covariation of cochlear implant users' perception and production of vowel contrasts and their identification by listeners with normal hearing. *Journal of Speech, Language, and Hearing Research 44*, 1257–1267.

Wang, D., J. H. Hansen, and E. Tobey (2015). Speech enhancement based on glimpse detection to improve the speech intelligibility for cochlear implant recipient. In *Proceedings of Conference on Implantable Auditory Prostheses (CIAP)*. Lake Tahoe, CA, pp127.

Warren, S. F., J. Gilkerson, J. A. Richards, D. K. Oller, D. Xu, U. Yapanel, and S. Gray (2010). What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of Autism and Developmental Disorders 40*, 555–569.

Weiss, M. R. et al. (1993). Effects of noise and noise reduction processing on the operation of the nucleus-22 cochlear implant processor. *Journal of rehabilitation research and development 30*, 117–117.

Wilson, B. S., C. C. Finley, D. T. Lawson, R. D. Wolford, D. K. Eddington, and W. M. Rabinowitz (1991). Better speech recognition with cochlear implants. *Nature 352*, 236–238.

Wouters, J. and J. V. Berghe (2001). Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system. *Ear and hearing 22*(5), 420–430.

Xu, D., U. Yapanel, S. Gray, J. Gilkerson, J. Richards, and J. H. L. Hansen (2008). Signal processing for young child speech language development. In *Proceedings of the Workshop on Child, Computer and Interaction (WOCCI)*. Chania, Greece, 20-25.

Yang, L.-P. and Q.-J. Fu (2005). Spectral subtraction-based speech enhancement for cochlear implant patients in background noise. *The Journal of the Acoustical Society of America 117*(3), 1001–1004.

Yu, C., J. H. L. Hansen, and D. W. Oard (2014). Houston, we have a solution': A case study of the analysis of astronaut speech during NASA apollo 11 for long-term speaker modeling. In *Proceedings of Annual Conference of the International Speech Communication Association (INTERSPEECH)*. Singapore, 945-948.

Yuan, J. and M. Liberman (2008). Speaker identification on the SCOTUS corpus. In *In Proceedings of Acoustics*. Paris, France, 5687-5690.

Zeng, F.-G., G. Grant, J. Niparko, J. Galvin, R. Shannon, J. Opie, and P. Segel (2002). Speech dynamic range and its effect on cochlear implant performance. *The Journal of the Acoustical Society of America 111*(1), 377–386.

Zeng, F. G., S. Rebscher, W. Harrison, X. Sun, and H. Feng (2008). Cochlear implants: system design, integration, and evaluation. *IEEE Reviews in Biomedical Engineering 1*, 115–142.

Ziaei, A., A. Sangwan, and J. H. L. Hansen (2012). Prof-Life-Log: Audio environment detection for naturalistic audio streams. In *Proceedings of Annual Conference of the International Speech Communication Association (INTERSPEECH)*. Portland,OR, 2514-2517.

Ziaei, A., A. Sangwan, and J. H. L. Hansen (2013). Prof-Life-Log: Personal interaction analysis for naturalistic audio streams. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Vancouver, Canada, 7770-7774.

Zorila, T.-C., V. Kandia, and Y. Stylianou (2012). Speech-in-noise intelligibility improvement based on spectral shaping and dynamic range compression. In *Proceedings of Annual Conference of the International Speech Communication Association (INTERSPEECH)*. Portland, OR, 634-637.

BIOGRAPHICAL SKETCH

Jaewook Lee is currently pursuing a Ph.D degree in Electrical Engineering from The University of Texas at Dallas (UTD). He completed Bachelor of Engineering and Master of Engineering degrees in Electrical Engineering from Dong Seo University in 2002 and 2005 respectively. After completing his Master's degree, he served as a lecturer of the Information and Communication department at Huree University. Since June 2011, he has been a Research Assistant with the Center for Robust Speech Systems - Cochlear Implant Lab (CRSS-CIL) at UTD under the supervision of Professor John H. L. Hansen. His research interests include digital signal processing, speech signal processing, biomedical signal processing, and cochlear implants.

CURRICULUM VITAE

Jaewook Lee

March 1, 2016

Contact Information:

Department of Computer Science
The University of Texas at Dallas
800 W. Campbell Rd.
Richardson, TX 75080-3021, U.S.A.

Voice: (469) 236-2529
Email: jaewook@utdallas.edu
Permanent Email: jaewooklee76@gmail.com

Educational History:

B.E., Electrical Engineering, Dong Seo University, Pusan, S. Korea, 2002
M.E., Electrical Engineering, Dong Seo University, Pusan, S. Korea, 2005
Ph.D., Electrical Engineering, The University of Texas at Dallas, Richardson, TX, present

Professional Experience:

Research Assistant, The University of Texas at Dallas, June 2011– present
Teaching Assistant, The University of Texas at Dallas, June 2011– May 2013
Lecturer, Huree University, Ulaanbaatar, Mongolia, January 2005 – June 2007

Publications: Journal Articles

1. **Jaewook Lee**, Hussnain Ali, Ali Ziaei, Emily A. Tobey, John H.L. Hansen, "The Lombard effect in speech production by cochlear implant users in noisy environments: a naturalistic study, *The Journal of the Acoustical Society of America* (accepted)

2. Oldooz Hazrati, **Jaewook Lee** and Philipos C. Loizou, "Blind binary masking for reverberation suppression in cochlear implants," *The Journal of the Acoustical Society of America*, Vol. 133 (3), pp. 1607 1614, March, 2013.

Publications: Conference Proceedings

1. **Jaewook Lee**, Hussnain Ali, John H.L. Hansen, "Intelligibility enhancement of neutral speech based on Lombard effect modification with application to cochlear implant users," *in Proc. Annual Midwinter Meeting of Association for Research in Otolaryngology (ARO 2017)*, Baltimore, MD, February, 2017

2. **Jaewook Lee**, Hussnain Ali, John H.L. Hansen, "The Lombard reflex and its influence on speech perception in adult cochlear implant users," *in Proc. International Conference on Cochlear Implants (CI 2016)* , Toronto, Canada, May, 2016.

3. Juliana Saba, **Jaewook Lee**, Hussnain Ali, Son Ta, Tuan Nguyen, John H.L. Hansen, "Impulse suppression algorithm development of a compatible program for cochlear implant users," *in Proc. Meeting of the Acoustical Society of America (ASA 2016)*, Salt Lake City, UT, April, 2016

4. **Jaewook Lee**, Hussnain Ali, Ali Ziaei, Emily A. Tobey, John H.L. Hansen, "Impact analysis of naturalistic environmental noise type on speech production for cochlear implant users versus normal hearing listeners," *in Proc. Conference on Implantable Auditory Prostheses (CIAP 2015)*, Lake Tahoe, CA, July, 2015.

5. **Jaewook Lee**, Hussnain Ali, Ali Ziaei, John H.L. Hansen, "Analysis of speech and language communication for cochlear implant users in noisy Lombard conditions, *in Proc. International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP 2015)*, Brisbane, Australia, April, 2015.

6. **Jaewook Lee**, Hussnain Ali, Ali Ziaei, John H.L. Hansen, "Lombard effect based speech analysis across noisy environments for voice communications with cochlear implant subjects, *in Proc. Meeting of the Acoustical Society of America (ASA 2014)*, Indianapolis, October, 2014.

7. Oldooz Hazrati, **Jaewook Lee** and Philipos C. Loizou, "Binary mask estimation for improved speech intelligibility in reverberant environments," *in Proc. Annual Conference of the International Speech Communication Association (INTERSPEECH 2012)*, Portland, OR, September, 2012.

8. Oldooz Hazrati, **Jaewook Lee** and Philipos C. Loizou, "The contribution of vowel-consonant boundaries to speech recognition in reverberation by cochlear implant users," *in Proc. Annual Midwinter Meeting of Association for Research in Otolaryngology (ARO 2012)*, San Diego, CA, February, 2012